UNIT I

1. What is classification?
a) when the output variable is a **category,** such as “red” or “blue” or “disease” and “no disease”.
b) when the output variable is a **real value**, such as “dollars” or “weight”.

Ans: Solution A

2. What is regression?
a) When the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.
b) When the output variable is a real value, such as “dollars” or “weight”.

Ans: Solution B

3. What is supervised learning?
a) All data is unlabelled and the algorithms learn to inherent structure from the input data
b) All data is labelled and the algorithms learn to predict the output from the input data
c) It is a framework for learning where an agent interacts with an environment and receives a reward for each interaction
d) Some data is labelled but most of it is unlabelled and a mixture of supervised and unsupervised techniques can be used.

Ans: Solution B

4. What is Unsupervised learning?
a) All data is unlabelled and the algorithms learn to inherent structure from the input data
b) All data is labelled and the algorithms learn to predict the output from the input data
c) It is a framework for learning where an agent interacts with an environment and receives a reward for each interaction
d) Some data is labelled but most of it is unlabelled and a mixture of supervised and unsupervised techniques can be used.

Ans: Solution A

5. What is Semi-Supervised learning?
a) All data is unlabelled and the algorithms learn to inherent structure from the input data
b) All data is labelled and the algorithms learn to predict the output from the input data
c) It is a framework for learning where an agent interacts with an environment and receives a reward for each interaction
d) Some data is labelled but most of it is unlabelled and a mixture of supervised and unsupervised techniques can be used.

Ans: Solution D

6. What is Reinforcement learning?

a) All data is unlabelled and the algorithms learn to inherent structure from the input data
b) All data is labelled and the algorithms learn to predict the output from the input data
c) It is a framework for learning where an agent interacts with an environment and receives a reward for each interaction
d) Some data is labelled but most of it is unlabelled and a mixture of supervised and unsupervised techniques can be used.

Ans: Solution C

7. Sentiment Analysis is an example of:

Regression,

Classification

Clustering

Reinforcement Learning

Options:

A. 1 Only

B. 1 and 2

C. 1 and 3

D. 1, 2 and 4

Ans : Solution D

8. The process of forming general concept definitions from examples of concepts to be learned.

a) Deduction
b) abduction
c) induction
d) conjunction

Ans : Solution C

9. Computers are best at learning
   a) facts.
   b) concepts.
   c) procedures.
   d) principles.

Ans : Solution A

10. Data used to build a data mining model.
    a) validation data
    b) training data
    c) test data
    d) hidden data

Ans : Solution B

11. Supervised learning and unsupervised clustering both require at least one
    a) hidden attribute.
    b) output attribute.
    c) input attribute.
    d) categorical attribute.

Ans : Solution A

12. Supervised learning differs from unsupervised clustering in that supervised learning requires
    a) at least one input attribute.
    b) input attributes to be categorical.
    c) at least one output attribute.
    d) output attributes to be categorical.

Ans : Solution B

13. A regression model in which more than one independent variable is used to predict the dependent variable is called
    a) a simple linear regression model
    b) a multiple regression models
    c) an independent model
    d) none of the above

Ans : Solution C

14. A term used to describe the case when the independent variables in a multiple regression model are correlated is
    a) Regression
    b) correlation
    c) multicollinearity
    d) none of the above

Ans : Solution C

15. A multiple regression model has the form: y = 2 + 3x1 + 4x2. As x1 increases by 1 unit (holding x2 constant), y will

a) increase by 3 units
b) decrease by 3 units
c) increase by 4 units
d) decrease by 4 units

Ans : Solution C

16. A multiple regression model has

a) only one independent variable
b) more than one dependent variable
c) more than one independent variable
d) none of the above

Ans : Solution B

17. A measure of goodness of fit for the estimated regression equation is the

a) multiple coefficient of determination
b) mean square due to error
c) mean square due to regression
d) none of the above

Ans : Solution C

18. The adjusted multiple coefficient of determination accounts for

a) the number of dependent variables in the model
b) the number of independent variables in the model
c) unusually large predictors
d) none of the above

Ans : Solution D

19. The multiple coefficient of determination is computed by

a) dividing SSR by SST
b) dividing SST by SSR
c) dividing SST by SSE
d) none of the above

Ans : Solution C

20. For a multiple regression model, SST = 200 and SSE = 50. The multiple coefficient of determination is

a) 0.25

b) 4.00
c) 0.75
d) none of the above

Ans : Solution B

21. A nearest neighbor approach is best used
    a) with large-sized datasets.
    b) when irrelevant attributes have been removed from the data.
    c) when a generalized model of the data is desirable.
    d) when an explanation of what has been found is of primary importance.

Ans : Solution B

22. Another name for an output attribute.
    a) predictive variable
    b) independent variable
    c) estimated variable
    d) dependent variable

Ans : Solution B

23. Classification problems are distinguished from estimation problems in that
    a) classification problems require the output attribute to be numeric.
    b) classification problems require the output attribute to be categorical.
    c) classification problems do not allow an output attribute.
    d) classification problems are designed to predict future outcome.

Ans : Solution C

24. Which statement is true about prediction problems?
    a) The output attribute must be categorical.
    b) The output attribute must be numeric.
    c) The resultant model is designed to determine future outcomes.
    d) The resultant model is designed to classify current behavior.

Ans : Solution D

25. Which statement about outliers is true?
    a) Outliers should be identified and removed from a dataset.
    b) Outliers should be part of the training dataset but should not be present in the test data.
    c) Outliers should be part of the test dataset but should not be present in the training data.
    d) The nature of the problem determines how outliers are used.

Ans : Solution D

26. Which statement is true about neural network and linear regression models?
    a) Both models require input attributes to be numeric.
    b) Both models require numeric attributes to range between 0 and 1.
    c) The output of both models is a categorical attribute value.
    d) Both techniques build models whose output is determined by a linear sum of weighted input attribute values.

Ans : Solution A

27. Which of the following is a common use of unsupervised clustering?
    a) detect outliers
    b) determine a best set of input attributes for supervised learning
    c) evaluate the likely performance of a supervised learner model
    d) determine if meaningful relationships can be found in a dataset

Ans : Solution A

28. The average positive difference between computed and desired outcome values.
    a) root mean squared error
    b) mean squared error
    c) mean absolute error
    d) mean positive error

Ans : Solution D

29. Selecting data so as to assure that each class is properly represented in both the training and test set.
    a) cross validation
    b) stratification
    c) verification
    d) bootstrapping

Ans : Solution B

30. The standard error is defined as the square root of this computation.
    a) The sample variance divided by the total number of sample instances.
    b) The population variance divided by the total number of sample instances.
    c) The sample variance divided by the sample mean.
    d) The population variance divided by the sample mean.

Ans : Solution A

31. Data used to optimize the parameter settings of a supervised learner model.

a) Training
b) Test
c) Verification
d) Validation

Ans : Solution D

32. Bootstrapping allows us to

a) choose the same training instance several times.
b) choose the same test set instance several times.
c) build models with alternative subsets of the training data several times.
d) test a model with alternative subsets of the test data several times.

Ans : Solution A

33. The correlation between the number of years an employee has worked for a company and the salary of the employee is 0.75. What can be said about employee salary and years worked?

a) There is no relationship between salary and years worked.
b) Individuals that have worked for the company the longest have higher salaries.
c) Individuals that have worked for the company the longest have lower salaries.
d) The majority of employees have been with the company a long time.
e) The majority of employees have been with the company a short period of time.

Ans : Solution B

34. The correlation coefficient for two real-valued attributes is –0.85. What does this value tell you?

a) The attributes are not linearly related.
b) As the value of one attribute increases the value of the second attribute also increases.
c) As the value of one attribute decreases the value of the second attribute increases.
d) The attributes show a curvilinear relationship.

Ans : Solution C

35. The average squared difference between classifier predicted output and actual output.

a) mean squared error
b) root mean squared error
c) mean absolute error
d) mean relative error

Ans : Solution A

36. Simple regression assumes a ___________ relationship between the input attribute and output attribute.

a) Linear

b) Quadratic
c) reciprocal
d) inverse

Ans : Solution A

37. Regression trees are often used to model ________ data.
    a) Linear
    b) Nonlinear
    c) Categorical
    d) Symmetrical

Ans : Solution B

38. The leaf nodes of a model tree are
    a) averages of numeric output attribute values.
    b) nonlinear regression equations.
    c) linear regression equations.
    d) sums of numeric output attribute values.

Ans : Solution C

39. Logistic regression is a _________ regression technique that is used to model data having a ______outcome.
    a) linear, numeric
    b) linear, binary
    c) nonlinear, numeric
    d) nonlinear, binary

Ans : Solution D

40. This technique associates a conditional probability value with each data instance.
    a) linear regression
    b) logistic regression
    c) simple regression
    d) multiple linear regression

Ans : Solution B

41. This supervised learning technique can process both numeric and categorical input attributes.
    a) linear regression
    b) Bayes classifier
    c) logistic regression
    d) backpropagation learning

Ans : Solution A

42. With Bayes classifier, missing data items are
    a) treated as equal compares.
    b) treated as unequal compares.
    c) replaced with a default value.
    d) ignored.

Ans : Solution B

43. This clustering algorithm merges and splits nodes to help modify nonoptimal partitions.
    a) agglomerative clustering
    b) expectation maximization
    c) conceptual clustering
    d) K-Means clustering

Ans : Solution D

44. This clustering algorithm initially assumes that each data instance represents a single cluster.

a) agglomerative clustering
b) conceptual clustering
c) K-Means clustering
d) expectation maximization

Ans : Solution C

45. This unsupervised clustering algorithm terminates when mean values computed for the current iteration of the algorithm are identical to the computed mean values for the previous iteration.
    a) agglomerative clustering
    b) conceptual clustering
    c) K-Means clustering
    d) expectation maximization

Ans : Solution C

46. Machine learning techniques differ from statistical techniques in that machine learning methods
    a) typically assume an underlying distribution for the data.
    b) are better able to deal with missing and noisy data.
    c) are not able to explain their behavior.
    d) have trouble with large-sized datasets.

Ans : Solution B

UNIT –II

1.True- False: Over fitting is more likely when you have huge amount of data to train?

A) TRUE

B) FALSE

Ans **Solution: (B)**

With a small training dataset, it's easier to find a hypothesis to fit the training data exactly i.e. over fitting.

2.What is pca.components_ in Sklearn?

Set of all eigen vectors for the projection space

Matrix of principal components

Result of the multiplication matrix

None of the above options

Ans A

3.**Which of the following techniques would perform better for reducing dimensions of a data set?**

A. Removing columns which have too many missing values

B. Removing columns which have high variance in data

C. Removing columns with dissimilar data trends

D. None of these

Ans Solution: **(A)**

If a columns have too many missing values, (say 99%) then we can remove such columns.

4.**It is not necessary to have a target variable for applying dimensionality reduction algorithms.**

A. TRUE

B. FALSE

Ans Solution: **(A)**

LDA is an example of supervised dimensionality reduction algorithm.

5. **PCA can be used for projecting and visualizing data in lower dimensions.**

A. TRUE

B. FALSE

Ans Solution: **(A)**

Sometimes it is very useful to plot the data in lower dimensions. We can take the first 2 principal components and then visualize the data using scatter plot.

6. The most popularly used dimensionality reduction algorithm is Principal Component Analysis (PCA). Which of the following is/are true about PCA?

PCA is an unsupervised method

It searches for the directions that data have the largest variance
Maximum number of principal components <= number of features
All principal components are orthogonal to each other
A. 1 and 2
B. 1 and 3
C. 2 and 3
D. All of the above

Ans D

7. PCA works better if there is?
A linear structure in the data
If the data lies on a curved surface and not on a flat surface
If variables are scaled in the same unit
A. 1 and 2
B. 2 and 3
C. 1 and 3
D. 1 ,2 and 3
Ans Solution: **(C)**

8. What happens when you get features in lower dimensions using PCA?
The features will still have interpretability
The features will lose interpretability
The features must carry all information present in data
The features may not carry all information present in data
A. 1 and 3
B. 1 and 4
C. 2 and 3
D. 2 and 4
Ans Solution: **(D)**
When you get the features in lower dimensions then you will lose some information of data most of the times and you won't be able to interpret the lower dimension data.

9. Which of the following option(s) is / are true?
You need to initialize parameters in PCA
You don't need to initialize parameters in PCA
PCA can be trapped into local minima problem
PCA can't be trapped into local minima problem
A. 1 and 3
B. 1 and 4
C. 2 and 3
D. 2 and 4

Ans Solution: **(D)**

PCA is a deterministic algorithm which doesn't have parameters to initialize and it doesn't have local minima problem like most of the machine learning algorithms has.

10. **What is of the following statement is true about t-SNE in comparison to PCA?**

A. When the data is huge (in size), t-SNE may fail to produce better results.

B. T-NSE always produces better result regardless of the size of the data

C. PCA always performs better than t-SNE for smaller size data.

D. None of these

Ans Solution: **(A)**

Option A is correct

11. **[True or False] PCA can be used for projecting and visualizing data in lower dimensions.**

A. TRUE

B. FALSE

Solution: (A)

Sometimes it is very useful to plot the data in lower dimensions. We can take the first 2 principal components and then visualize the data using scatter plot.

12. **A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college.**

1) Which of the following statement is true in following case?

A) Feature F1 is an example of nominal variable.

B) Feature F1 is an example of ordinal variable.

C) It doesn't belong to any of the above category.

D) Both of these

Solution: (B)

Ordinal variables are the variables which has some order in their categories. For example, grade A should be consider as high grade than grade B.

13. **Which of the following is an example of a deterministic algorithm?**

A) PCA

B) K-Means

C) None of the above

Solution: (A)

A deterministic algorithm is that in which output does not change on different runs. PCA would give the same result if we run again, but not k-means.

UNIT –III

1. Which of the following methods do we use to best fit the data in Logistic Regression?

A) Least Square Error

B) Maximum Likelihood

C) Jaccard distance

D) Both A and B

Ans Solution: B

2. Choose which of the following options is true regarding One-Vs-All method in Logistic Regression.

A) We need to fit n models in n-class classification problem

B) We need to fit n-1 models to classify into n classes

C) We need to fit only 1 model to classify into n classes

D) None of these

Ans **Solution: A**

3. Suppose, You applied a Logistic Regression model on a given data and got a training accuracy X and testing accuracy Y. Now, you want to add a few new features in the same data. Select the option(s) which is/are correct in such a case.

Note: Consider remaining parameters are same.

A) Training accuracy increases

B) Training accuracy increases or remains the same

C) Testing accuracy decreases

D) Testing accuracy increases or remains the same

Ans Solution: A and D

Adding more features to model will increase the training accuracy because model has to consider more data to fit the logistic regression. But testing accuracy increases if feature is found to be significant

4. Which of the following algorithms do we use for Variable Selection?

A) LASSO

B) Ridge

C) Both

D) None of these

Ans Solution: A

In case of lasso we apply a absolute penality, after increasing the penality in lasso some of the coefficient of variables may become zero

5. Which of the following statement is true about outliers in Linear regression?

A) Linear regression is sensitive to outliers
B) Linear regression is not sensitive to outliers
C) Can't say
D) None of these
Ans Solution: (A)
The slope of the regression line will change due to outliers in most of the cases. So Linear Regression is sensitive to outliers.

6. Which of the following methods do we use to find the best fit line for data in Linear Regression?
A) Least Square Error
B) Maximum Likelihood
C) Logarithmic Loss
D) Both A and B
Ans Solution: (A)
In linear regression, we try to minimize the least square errors of the model to identify the line of best fit.

7. Which of the following is true about Residuals?
A) Lower is better
B) Higher is better
C) A or B depend on the situation
D) None of these
Ans Solution: (A)
Residuals refer to the error values of the model. Therefore lower residuals are desired.

8. Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and you found that there is a relationship between them. Which of the following conclusion do you make about this situation?

A) Since the there is a relationship means our model is not good
B) Since the there is a relationship means our model is good
C) Can't say
D) None of these
Ans Solution: (A)
There should not be any relationship between predicted values and residuals. If there exists any relationship between them, it means that the model has not perfectly captured the information in the data.

9. Suppose you have fitted a complex regression model on a dataset. Now, you are using Ridge regression with penalty x.

Choose the option which describes bias in best manner.

A) In case of very large x; bias is low

B) In case of very large x; bias is high

C) We can’t say about bias

D) None of these

Ans Solution: (B)

If the penalty is very large it means model is less complex, therefore the bias would be high.

10. Which of the following option is true?

A) Linear Regression errors values has to be normally distributed but in case of Logistic Regression it is not the case

B) Logistic Regression errors values has to be normally distributed but in case of Linear Regression it is not the case

C) Both Linear Regression and Logistic Regression error values have to be normally distributed

D) Both Linear Regression and Logistic Regression error values have not to be normally distributed

Ans Solution: A

11. Suppose you have trained a logistic regression classifier and it outputs a new example x with a prediction ho(x) = 0.2. This means

Our estimate for P(y=1 | x)

Our estimate for P(y=0 | x)

Our estimate for P(y=1 | x)

Our estimate for P(y=0 | x)

Ans Solution: B

12. **True-False: Linear Regression is a supervised machine learning algorithm.**

A) TRUE

B) FALSE

Solution: (A)

Yes, Linear regression is a supervised learning algorithm because it uses true labels for training. Supervised learning algorithm should have input variable (x) and an output variable (Y) for each example.

13. **True-False: Linear Regression is mainly used for Regression.**

A) TRUE

B) FALSE

Solution: (A)

Linear Regression has dependent variables that have continuous values.

14. True-False: It is possible to design a Linear regression algorithm using a neural network?

A) TRUE
B) FALSE

Solution: (A)

True. A Neural network can be used as a universal approximator, so it can definitely implement a linear regression algorithm.

15. **Which of the following methods do we use to find the best fit line for data in Linear Regression?**
A) Least Square Error
B) Maximum Likelihood
C) Logarithmic Loss
D) Both A and B
Solution: (A)
In linear regression, we try to minimize the least square errors of the model to identify the line of best fit.

16. **Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?**
A) AUC-ROC
B) Accuracy
C) Logloss
D) Mean-Squared-Error
Solution: (D)
Since linear regression gives output as continuous values, so in such case we use mean squared error metric to evaluate the model performance. Remaining options are use in case of a classification problem.

17. **True-False: Lasso Regularization can be used for variable selection in Linear Regression.**
A) TRUE
B) FALSE
Solution: (A)
True, In case of lasso regression we apply absolute penalty which makes some of the coefficients zero.

18. **Which of the following is true about Residuals ?**
A) Lower is better
B) Higher is better

C) A or B depend on the situation

D) None of these

Solution: (A)

Residuals refer to the error values of the model. Therefore lower residuals are desired.

19. Suppose that we have N independent variables (X1,X2… Xn) and dependent variable is Y. Now Imagine that you are applying linear regression by fitting the best fit line using least square error on this data.

You found that correlation coefficient for one of it's variable(Say X1) with Y is -0.95.

Which of the following is true for X1?

A) Relation between the X1 and Y is weak

B) Relation between the X1 and Y is strong

C) Relation between the X1 and Y is neutral

D) Correlation can't judge the relationship

Solution: (B)

The absolute value of the correlation coefficient denotes the strength of the relationship. Since absolute correlation is very high it means that the relationship is strong between X1 and Y.

20. **Looking at above two characteristics, which of the following option is the correct for Pearson correlation between V1 and V2?**

If you are given the two variables V1 and V2 and they are following below two characteristics.

1. If V1 increases then V2 also increases
2. If V1 decreases then V2 behavior is unknown

A) Pearson correlation will be close to 1

B) Pearson correlation will be close to -1

C) Pearson correlation will be close to 0

D) None of these

Solution: (D)

We cannot comment on the correlation coefficient by using only statement 1. We need to consider the both of these two statements. Consider V1 as x and V2 as |x|. The correlation coefficient would not be close to 1 in such a case.

21. **Suppose Pearson correlation between V1 and V2 is zero. In such case, is it right to conclude that V1 and V2 do not have any relation between them?**

A) TRUE

B) FALSE

Solution: (B)

Pearson correlation coefficient between 2 variables might be zero even when they have a relationship between them. If the correlation coefficient is zero, it just means that that they don't move together. We can take examples like y=|x| or y=x^2.

22. **True- False: Overfitting is more likely when you have huge amount of data to train?**

A) TRUE

B) FALSE

Solution: (B)

With a small training dataset, it's easier to find a hypothesis to fit the training data exactly i.e. overfitting.

23. **We can also compute the coefficient of linear regression with the help of an analytical method called "Normal Equation". Which of the following is/are true about Normal Equation?**

1. We don't have to choose the learning rate
2. It becomes slow when number of features is very large
3. Thers is no need to iterate

A) 1 and 2

B) 1 and 3

C) 2 and 3

D) 1,2 and 3

Solution: (D)

Instead of gradient descent, Normal Equation can also be used to find coefficients.

Question Context 24-26:

Suppose you have fitted a complex regression model on a dataset. Now, you are using Ridge regression with penality x.

24. Choose the option which describes bias in best manner.

A) In case of very large x; bias is low

B) In case of very large x; bias is high

C) We can't say about bias

D) None of these

Solution: (B)

If the penalty is very large it means model is less complex, therefore the bias would be high.

25. **What will happen when you apply very large penalty?**

A) Some of the coefficient will become absolute zero

B) Some of the coefficient will approach zero but not absolute zero

C) Both A and B depending on the situation

D) None of these

Solution: (B)

In lasso some of the coefficient value become zero, but in case of Ridge, the coefficients become close to zero but not zero.

26. **What will happen when you apply very large penalty in case of Lasso?**

A) Some of the coefficient will become zero

B) Some of the coefficient will be approaching to zero but not absolute zero

C) Both A and B depending on the situation

D) None of these

Solution: (A)

As already discussed, lasso applies absolute penalty, so some of the coefficients will become zero.

27. **Which of the following statement is true about outliers in Linear regression?**

A) Linear regression is sensitive to outliers

B) Linear regression is not sensitive to outliers

C) Can't say

D) None of these

Solution: (A)

The slope of the regression line will change due to outliers in most of the cases. So Linear Regression is sensitive to outliers.

28. **Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and you found that there is a relationship between them. Which of the following conclusion do you make about this situation?**

A) Since the there is a relationship means our model is not good

B) Since the there is a relationship means our model is good

C) Can't say

D) None of these

Solution: (A)

There should not be any relationship between predicted values and residuals. If there exists any relationship between them,it means that the model has not perfectly captured the information in the data.

Question Context 29-31:

Suppose that you have a dataset D1 and you design a linear regression model of degree 3 polynomial and you found that the training and testing error is "0" or in another terms it perfectly fits the data.

29. What will happen when you fit degree 4 polynomial in linear regression?

A) There are high chances that degree 4 polynomial will over fit the data

B) There are high chances that degree 4 polynomial will under fit the data

C) Can't say

D) None of these

Solution: (A)

Since is more degree 4 will be more complex(overfit the data) than the degree 3 model so it will again perfectly fit the data. In such case training error will be zero but test error may not be zero.

30. **What will happen when you fit degree 2 polynomial in linear regression?**
A) It is high chances that degree 2 polynomial will over fit the data
B) It is high chances that degree 2 polynomial will under fit the data
C) Can’t say
D) None of these
Solution: (B)
If a degree 3 polynomial fits the data perfectly, it’s highly likely that a simpler model(degree 2 polynomial) might under fit the data.

31. **In terms of bias and variance. Which of the following is true when you fit degree 2 polynomial?**

A) Bias will be high, variance will be high
B) Bias will be low, variance will be high
C) Bias will be high, variance will be low
D) Bias will be low, variance will be low
Solution: (C)
Since a degree 2 polynomial will be less complex as compared to degree 3, the bias will be high and variance will be low.

Question Context 32-33:
We have been given a dataset with n records in which we have input attribute as x and output attribute as y. Suppose we use a linear regression method to model this data. To test our linear regressor, we split the data in training set and test set randomly.
32. Now we increase the training set size gradually. As the training set size increases, what do you expect will happen with the mean training error?

A) Increase
B) Decrease
C) Remain constant
D) Can’t Say
Solution: (D)
Training error may increase or decrease depending on the values that are used to fit the model. If the values used to train contain more outliers gradually, then the error might just increase.

33. What do you expect will happen with bias and variance as you increase the size of training data?

A) Bias increases and Variance increases
B) Bias decreases and Variance increases
C) Bias decreases and Variance decreases

D) Bias increases and Variance decreases
E) Can't Say False
Solution: (D)
As we increase the size of the training data, the bias would increase while the variance would decrease.

Question Context 34:
Consider the following data where one input(X) and one output(Y) is given.

34. What would be the root mean square training error for this data if you run a Linear Regression model of the form (Y = A0+A1X)?

A) Less than 0
B) Greater than zero
C) Equal to 0
D) None of these
Solution: (C)
We can perfectly fit the line on the following data so mean error will be zero.

Question Context 35-36:
Suppose you have been given the following scenario for training and validation error for Linear Regression.

Scenario	Learning Rate	Number of iterations	Training Error	Validation Error
1	0.1	1000	100	110
2	0.2	600	90	105

3	0.3	400	110	110
4	0.4	300	120	130
5	0.4	250	130	150

35. Which of the following scenario would give you the right hyper parameter?

A) 1

B) 2

C) 3

D) 4

Solution: (B)

Option B would be the better option because it leads to less training as well as validation error.

36. Suppose you got the tuned hyper parameters from the previous question. Now, Imagine you want to add a variable in variable space such that this added feature is important. Which of the following thing would you observe in such case?

A) Training Error will decrease and Validation error will increase

B) Training Error will increase and Validation error will increase

C) Training Error will increase and Validation error will decrease

D) Training Error will decrease and Validation error will decrease

E) None of the above

Solution: (D)

If the added feature is important, the training and validation error would decrease.

Question Context 37-38:

Suppose, you got a situation where you find that your linear regression model is under fitting the data.

37. In such situation which of the following options would you consider?

1. I will add more variables
2. I will start introducing polynomial degree variables
3. I will remove some variables

A) 1 and 2

B) 2 and 3

C) 1 and 3

D) 1, 2 and 3

Solution: (A)

In case of under fitting, you need to induce more variables in variable space or you can add some polynomial degree variables to make the model more complex to be able to fir the data better.

38. Now situation is same as written in previous question(under fitting).Which of following regularization algorithm would you prefer?

A) L1
B) L2
C) Any
D) None of these
Solution: (D)
I won't use any regularization methods because regularization is used in case of overfitting.

39. **True-False: Is Logistic regression a supervised machine learning algorithm?**
A) TRUE
B) FALSE
Solution: A
True, Logistic regression is a supervised learning algorithm because it uses true labels for training. Supervised learning algorithm should have input variables (x) and an target variable (Y) when you train the model .

40. **True-False: Is Logistic regression mainly used for Regression?**
A) TRUE
B) FALSE
Solution: B
Logistic regression is a classification algorithm, don't confuse with the name regression.

41. **True-False: Is it possible to design a logistic regression algorithm using a Neural Network Algorithm?**
A) TRUE
B) FALSE
Solution: A
True, Neural network is a is a *universal* approximator so it can implement linear regression algorithm.

42. **True-False: Is it possible to apply a logistic regression algorithm on a 3-class Classification problem?**
A) TRUE
B) FALSE
Solution: A
Yes, we can apply logistic regression on 3 classification problem, We can use One Vs all method for 3 class classification in logistic regression.

43. **Which of the following methods do we use to best fit the data in Logistic Regression?**

A) Least Square Error

B) Maximum Likelihood

C) Jaccard distance

D) Both A and B

Solution: B

Logistic regression uses maximum likely hood estimate for training a logistic regression.

44. **Which of the following evaluation metrics can not be applied in case of logistic regression output to compare with target?**

A) AUC-ROC

B) Accuracy

C) Logloss

D) Mean-Squared-Error

Solution: D

Since, Logistic Regression is a classification algorithm so it's output can not be real time value so mean squared error can not use for evaluating it

45. **One of the very good methods to analyze the performance of Logistic Regression is AIC, which is similar to R-Squared in Linear Regression. Which of the following is true about AIC?**

A) We prefer a model with minimum AIC value

B) We prefer a model with maximum AIC value

C) Both but depend on the situation

D) None of these

Solution: A

We select the best model in logistic regression which can least AIC.

46. **[True-False] Standardisation of features is required before training a Logistic Regression.**

A) TRUE

B) FALSE

Solution: B

Standardization isn't required for logistic regression. The main goal of standardizing features is to help convergence of the technique used for optimization.

47. Which of the following algorithms do we use for Variable Selection?

A) LASSO

B) Ridge

C) Both

D) None of these

Solution: A

In case of lasso we apply a absolute penality, after increasing the penality in lasso some of the coefficient of variables may become zero.

Context: 48-49

Consider a following model for logistic regression: P (y =1|x, w)= g(w0 + w1x)
where g(z) is the logistic function.

In the above equation the P (y =1|x; w) , viewed as a function of x, that we can get by changing the parameters w.

48 What would be the range of p in such case?

A) (0, inf)
B) (-inf, 0)
C) (0, 1)
D) (-inf, inf)

Solution: C

For values of *x* in the range of real number from −∞ to +∞ Logistic function will give the output between (0,1)

49 In above question what do you think which function would make p between (0,1)?

A) logistic function
B) Log likelihood function
C) Mixture of both
D) None of them

Solution: A

Explanation is same as question number 10

50. **Suppose you have been given a fair coin and you want to find out the odds of getting heads. Which of the following option is true for such a case?**

A) odds will be 0
B) odds will be 0.5
C) odds will be 1
D) None of these

Solution: C

Odds are defined as the ratio of the probability of success and the probability of failure. So in case of fair coin probability of success is 1/2 and the probability of failure is 1/2 so odd would be 1

51. **The logit function(given as l(x)) is the log of odds function. What could be the range of logit function in the domain x=[0,1]?**

A) $(-\infty, \infty)$
B) (0,1)
C) $(0, \infty)$
D) $(-\infty, 0)$

Solution: A

For our purposes, the odds function has the advantage of transforming the probability function, which has values from 0 to 1, into an equivalent function with values between 0 and ∞. When we take the natural log of the odds function, we get a range of values from -∞ to ∞.

52. **Which of the following option is true?**

A) Linear Regression errors values has to be normally distributed but in case of Logistic Regression it is not the case
B) Logistic Regression errors values has to be normally distributed but in case of Linear Regression it is not the case
C) Both Linear Regression and Logistic Regression error values have to be normally distributed
D) Both Linear Regression and Logistic Regression error values have not to be normally distributed

Solution:A

53. **Which of the following is true regarding the logistic function for any value "x"?**

Note:
Logistic(x): is a logistic function of any number "x"

Logit(x): is a logit function of any number "x"

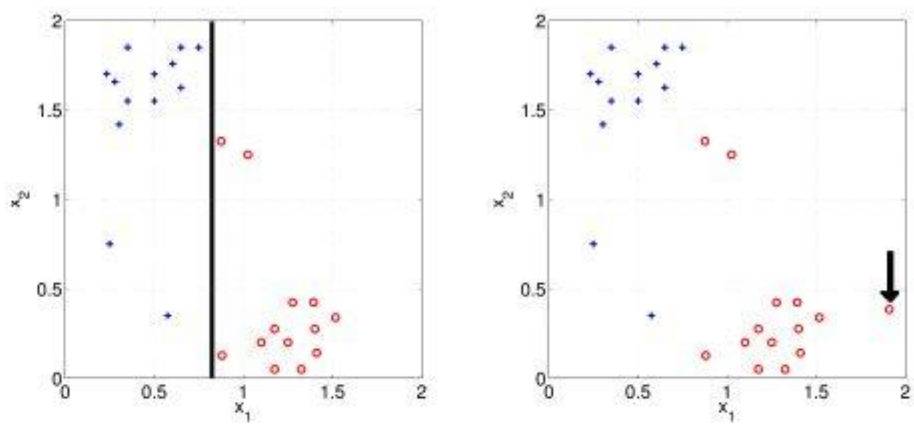
Logit_inv(x): is a inverse logit function of any number "x"

A) Logistic(x) = Logit(x)
B) Logistic(x) = Logit_inv(x)
C) Logit_inv(x) = Logit(x)
D) None of these

Solution: B

54. **How will the bias change on using high(infinite) regularisation?**

Suppose you have given the two scatter plot "a" and "b" for two classes(blue for positive and red for negative class). In scatter plot "a", you correctly classified all data points using logistic regression (black line is a decision boundary).



(a) (b)

A) Bias will be high
B) Bias will be low
C) Can’t say
D) None of these

Solution: A

Model will become very simple so bias will be very high.

55. **Suppose, You applied a Logistic Regression model on a given data and got a training accuracy X and testing accuracy Y. Now, you want to add a few new features in the same data. Select the option(s) which is/are correct in such a case.**

Note: Consider remaining parameters are same.

A) Training accuracy increases
B) Training accuracy increases or remains the same
C) Testing accuracy decreases
D) Testing accuracy increases or remains the same

Solution: A and D

Adding more features to model will increase the training accuracy because model has to consider more data to fit the logistic regression. But testing accuracy increases if feature is found to be significant

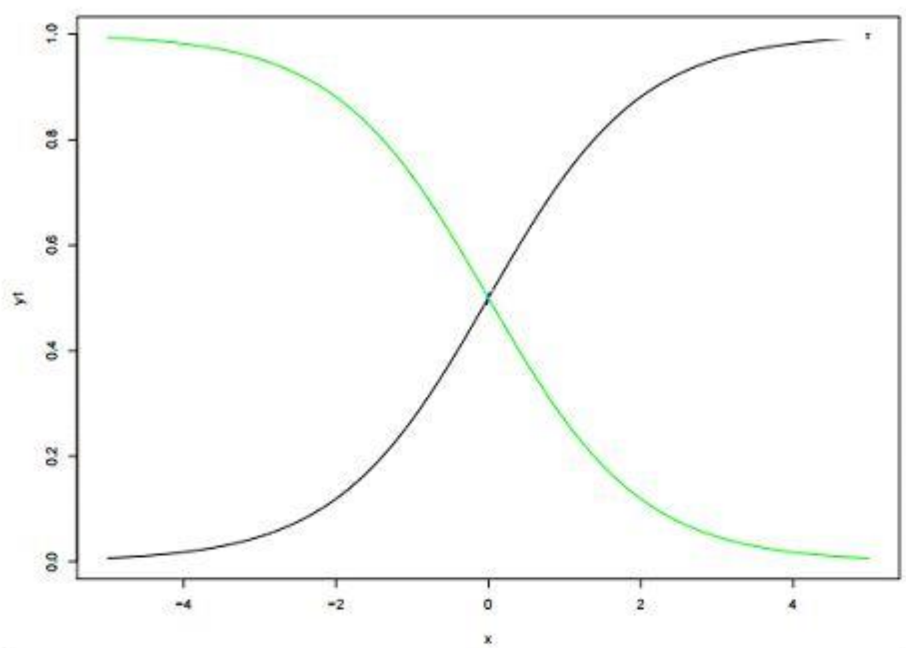
56. **Choose which of the following options is true regarding One-Vs-All method in Logistic Regression.**

A) We need to fit n models in n-class classification problem
B) We need to fit n-1 models to classify into n classes
C) We need to fit only 1 model to classify into n classes
D) None of these

Solution: A

If there are n classes, then n separate logistic regression has to fit, where the probability of each category is predicted over the rest of the categories combined.

57. **Below are two different logistic models with different values for β0 and β1.**



Which of the following statement(s) is true about β0 and β1 values of two logistics models (Green, Black)?

Note: consider Y = β0 + β1*X. Here, β0 is intercept and β1 is coefficient.

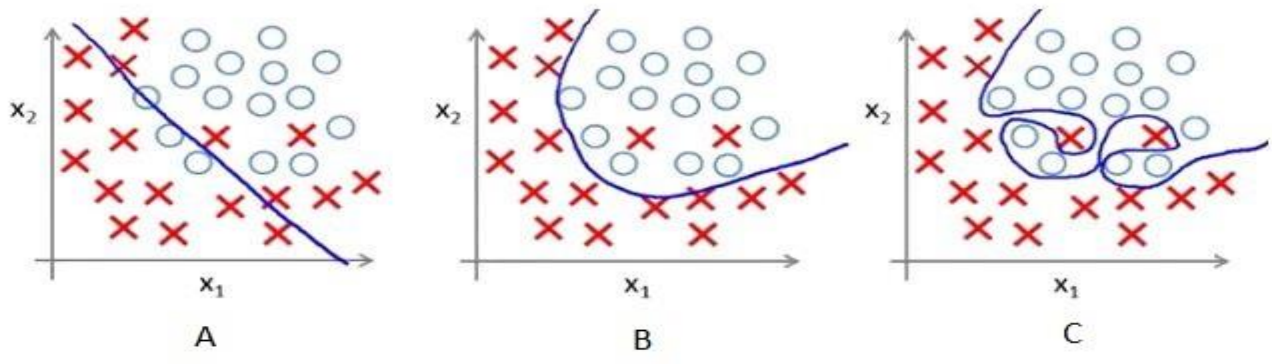
A) β1 for Green is greater than Black
B) β1 for Green is lower than Black
C) β1 for both models is same
D) Can't Say

Solution: B

β0 and β1: β0 = 0, β1 = 1 is in X1 color(black) and β0 = 0, β1 = −1 is in X4 color (green)

Context 58-60

Below are the three scatter plot(A,B,C left to right) and hand drawn decision boundaries for logistic regression.



58. Which of the following above figure shows that the decision boundary is overfitting the training data?

A) A
B) B
C) C
D)None of these

Solution: C

Since in figure 3, Decision boundary is not smooth that means it will over-fitting the data.

59. What do you conclude after seeing this visualization?

1. The training error in first plot is maximum as compare to second and third plot.
2. The best model for this regression problem is the last (third) plot because it has minimum training error (zero).
3. The second model is more robust than first and third because it will perform best on unseen data.
4. The third model is overfitting more as compare to first and second.
5. All will perform same because we have not seen the testing data.

A) 1 and 3
B) 1 and 3
C) 1, 3 and 4
D) 5

Solution: C

The trend in the graphs looks like a quadratic trend over independent variable X. A higher degree(Right graph) polynomial might have a very high accuracy on the train population but is expected to fail badly on test dataset. But if you see in left graph we will have training error maximum because it underfits the training data

60. Suppose, above decision boundaries were generated for the different value of regularization. Which of the above decision boundary shows the maximum regularization?

A) A
B) B
C) C
D) All have equal regularization

Solution: A

Since, more regularization means more penality means less complex decision boundry that shows in first figure A.

61. **What would do if you want to train logistic regression on same data that will take less time as well as give the comparatively similar accuracy(may not be same)?**

Suppose you are using a Logistic Regression model on a huge dataset. One of the problem you may face on such huge data is that Logistic regression will take very long time to train.

A) Decrease the learning rate and decrease the number of iteration
B) Decrease the learning rate and increase the number of iteration
C) Increase the learning rate and increase the number of iteration
D) Increase the learning rate and decrease the number of iteration

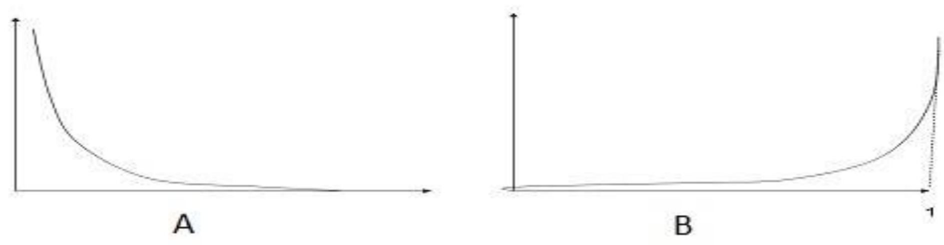
Solution: D

If you decrease the number of iteration while training it will take less time for surly but will not give the same accuracy for getting the similar accuracy but not exact you need to increase the learning rate.

62. **Which of the following image is showing the cost function for y =1.**

Following is the loss function in logistic regression(Y-axis loss function and x axis log probability) for two class classification problem.

Note: Y is the target class

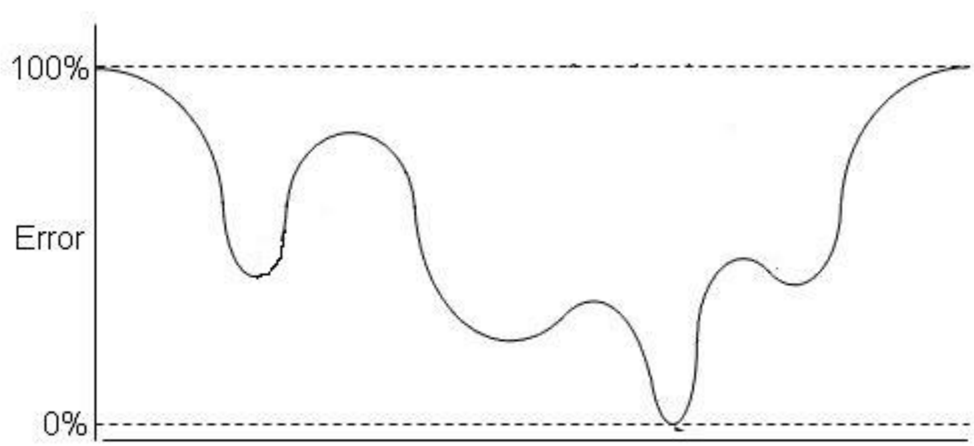


A) A
B) B
C) Both
D) None of these

Solution: A

A is the true answer as loss function decreases as the log probability increases

63. **Suppose, Following graph is a cost function for logistic regression.**



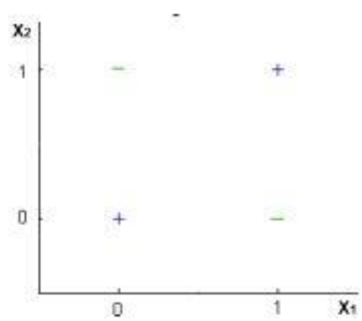
Now, How many local minimas are present in the graph?

A) 1
B) 2
C) 3
D) 4

Solution: C

There are three local minima present in the graph

64. **Can a Logistic Regression classifier do a perfect classification on the below data?**



Note: You can use only X1 and X2 variables where X1 and X2 can take only two binary values(0,1).

A) TRUE
B) FALSE
C) Can’t say
D) None of these

Solution: B

No, logistic regression only forms linear decision surface, but the examples in the figure are not linearly separable.

UNIT IV

1. The SVM's are less effective when:

A) The data is linearly separable
B) The data is clean and ready to use
C) The data is noisy and contains overlapping points

Ans Solution: C

When the data has noise and overlapping points, there is a problem in drawing a clear hyperplane without misclassifying.

2. The cost parameter in the SVM means:

A) The number of cross-validations to be made
B) The kernel to be used
C) The tradeoff between misclassification and simplicity of the model
D) None of the above

Ans Solution: C

The cost parameter decides how much an SVM should be allowed to "bend" with the data. For a low cost, you aim for a smooth decision surface and for a higher cost, you aim to classify more points correctly. It is also simply referred to as the cost of misclassification.

3. Which of the following are real world applications of the SVM?

A) Text and Hypertext Categorization
B) Image Classification
C) Clustering of News Articles
D) All of the above

Ans Solution: D

SVM's are highly versatile models that can be used for practically all real world problems ranging from regression to clustering and handwriting recognitions.

4. Which of the following is true about Naive Bayes ?

Assumes that all the features in a dataset are equally important

Assumes that all the features in a dataset are independent

Both A and B - answer

None of the above options

Ans Solution: C

5 What do you mean by generalization error in terms of the SVM?

A) How far the hyperplane is from the support vectors
B) How accurately the SVM can predict outcomes for unseen data
C) The threshold amount of error in an SVM

Ans Solution: B

Generalisation error in statistics is generally the out-of-sample error which is the measure of how accurately a model can predict values for previously unseen data.

6 The SVM's are less effective when:

A) The data is linearly separable
B) The data is clean and ready to use
C) The data is noisy and contains overlapping points

Ans Solution: C

When the data has noise and overlapping points, there is a problem in drawing a clear hyperplane without misclassifying.

7 What is/are true about kernel in SVM?

1. Kernel function map low dimensional data to high dimensional space
2. It's a similarity function

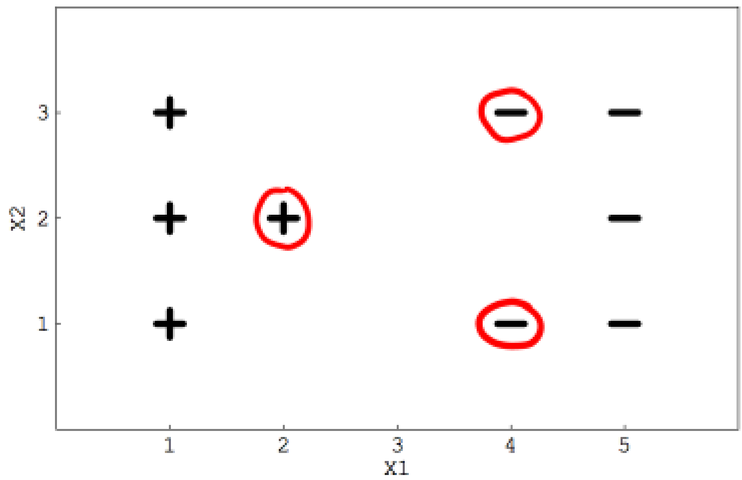
A) 1
B) 2
C) 1 and 2
D) None of these

Ans Solution: C

Both the given statements are correct.

Question Context:8– 9

Suppose you are using a Linear SVM classifier with 2 class classification problem. Now you have been given the following data in which some points are circled red that are representing support vectors.



8. If you remove the following any one red points from the data. Does the decision boundary will change?

A) Yes
B) No

Solution: A

These three examples are positioned such that removing any one of them introduces slack in the constraints. So the decision boundary would completely change.

9. [True or False] If you remove the non-red circled points from the data, the decision boundary will change?

A) True
B) False

Solution: B

On the other hand, rest of the points in the data won’t affect the decision boundary much.

10. What do you mean by generalization error in terms of the SVM?

A) How far the hyperplane is from the support vectors
B) How accurately the SVM can predict outcomes for unseen data
C) The threshold amount of error in an SVM

Solution: B

Generalization error in statistics is generally the out-of-sample error which is the measure of how accurately a model can predict values for previously unseen data.

11. When the C parameter is set to infinite, which of the following holds true?

A) The optimal hyperplane if exists, will be the one that completely separates the data
B) The soft-margin classifier will separate the data
C) None of the above

Solution: A

At such a high level of misclassification penalty, soft margin will not hold existence as there will be no room for error.

12. What do you mean by a hard margin?

A) The SVM allows very low error in classification
B) The SVM allows high amount of error in classification
C) None of the above

Solution: A

A hard margin means that an SVM is very rigid in classification and tries to work extremely well in the training set, causing overfitting.

13. The minimum time complexity for training an SVM is O(n2). According to this fact, what sizes of datasets are not best suited for SVM's?

A) Large datasets
B) Small datasets
C) Medium sized datasets
D) Size does not matter

Solution: A

Datasets which have a clear classification boundary will function best with SVM's.

14. The effectiveness of an SVM depends upon:

A) Selection of Kernel
B) Kernel Parameters
C) Soft Margin Parameter C
D) All of the above

Solution: D

The SVM effectiveness depends upon how you choose the basic 3 requirements mentioned above in such a way that it maximises your efficiency, reduces error and overfitting.

15. upport vectors are the data points that lie closest to the decision surface.

A) TRUE
B) FALSE

Solution: A

They are the points closest to the hyperplane and the hardest ones to classify. They also have a direct bearing on the location of the decision surface.

16. The SVM's are less effective when:

A) The data is linearly separable
B) The data is clean and ready to use
C) The data is noisy and contains overlapping points

Solution: C

When the data has noise and overlapping points, there is a problem in drawing a clear hyperplane without misclassifying.

17. Suppose you are using RBF kernel in SVM with high Gamma value. What does this signify?

A) The model would consider even far away points from hyperplane for modeling
B) The model would consider only the points close to the hyperplane for modeling
C) The model would not be affected by distance of points from hyperplane for modeling
D) None of the above

Solution: B

The gamma parameter in SVM tuning signifies the influence of points either near or far away from the hyperplane.

For a low gamma, the model will be too constrained and include all points of the training dataset, without really capturing the shape.

For a higher gamma, the model will capture the shape of the dataset well.

18. The cost parameter in the SVM means:

A) The number of cross-validations to be made
B) The kernel to be used
C) The tradeoff between misclassification and simplicity of the model
D) None of the above

Solution: C

The cost parameter decides how much an SVM should be allowed to “bend” with the data. For a low cost, you aim for a smooth decision surface and for a higher cost, you aim to classify more points correctly. It is also simply referred to as the cost of misclassification.

19. Suppose you are building a SVM model on data X. The data X can be error prone which means that you should not trust any specific data point too much. Now think that you want to build a SVM model which has quadratic kernel function of polynomial degree 2 that uses Slack variable C as one of it’s hyper parameter. Based upon that give the answer for following question.

What would happen when you use very large value of C(C->infinity)?

Note: For small C was also classifying all data points correctly

A) We can still classify data correctly for given setting of hyper parameter C
B) We can not classify data correctly for given setting of hyper parameter C
C) Can’t Say
D) None of these

Solution: A

For large values of C, the penalty for misclassifying points is very high, so the decision boundary will perfectly separate the data if possible.

20. What would happen when you use very small C (C~0)?

A) Misclassification would happen
B) Data will be correctly classified
C) Can’t say
D) None of these

Solution: A

The classifier can maximize the margin between most of the points, while misclassifying a few points, because the penalty is so low.

21. If I am using all features of my dataset and I achieve 100% accuracy on my training set, but ~70% on validation set, what should I look out for?

A) Underfitting
B) Nothing, the model is perfect
C) Overfitting

Solution: C

If we’re achieving 100% training accuracy very easily, we need to check to verify if we’re overfitting our data.

22. **Which of the following are real world applications of the SVM?**

A) Text and Hypertext Categorization
B) Image Classification
C) Clustering of News Articles
D) All of the above

Solution: D

SVM's are highly versatile models that can be used for practically all real world problems ranging from regression to clustering and handwriting recognitions.

Question Context: 23 – 25

Suppose you have trained an SVM with linear decision boundary after training SVM, you correctly infer that your SVM model is under fitting.

23. Which of the following option would you more likely to consider iterating SVM next time?

A) You want to increase your data points
B) You want to decrease your data points
C) You will try to calculate more variables
D) You will try to reduce the features

Solution: C

The best option here would be to create more features for the model.

24. Suppose you gave the correct answer in previous question. What do you think that is actually happening?

1. We are lowering the bias
2. We are lowering the variance
3. We are increasing the bias
4. We are increasing the variance

A) 1 and 2
B) 2 and 3
C) 1 and 4
D) 2 and 4

Solution: C

Better model will lower the bias and increase the variance

25. In above question suppose you want to change one of it's(SVM) hyperparameter so that effect would be same as previous questions i.e model will not under fit?

A) We will increase the parameter C
B) We will decrease the parameter C
C) Changing in C don't effect
D) None of these

Solution: A

Increasing C parameter would be the right thing to do here, as it will ensure regularized model

26. **We usually use feature normalization before using the Gaussian kernel in SVM. What is true about feature normalization?**

1. We do feature normalization so that new feature will dominate other
2. Some times, feature normalization is not feasible in case of categorical variables
3. Feature normalization always helps when we use Gaussian kernel in SVM

A) 1
B) 1 and 2
C) 1 and 3
D) 2 and 3

Solution: B

Statements one and two are correct.

Question Context: 27-29

Suppose you are dealing with 4 class classification problem and you want to train a SVM model on the data for that you are using One-vs-all method. Now answer the below questions?

27. How many times we need to train our SVM model in such case?

A) 1
B) 2
C) 3
D) 4

Solution: D

For a 4 class problem, you would have to train the SVM at least 4 times if you are using a one-vs-all method.

28. Suppose you have same distribution of classes in the data. Now, say for training 1 time in one vs all setting the SVM is taking 10 second. How many seconds would it require to train one-vs-all method end to end?

A) 20
B) 40
C) 60
D) 80

Solution: B

It would take 10×4 = 40 seconds

29 Suppose your problem has changed now. Now, data has only 2 classes. What would you think how many times we need to train SVM in such case?

A) 1
B) 2
C) 3
D) 4

Solution: A

Training the SVM only one time would give you appropriate results

Question context: 30 –31

Suppose you are using SVM with linear kernel of polynomial degree 2, Now think that you have applied this on data and found that it perfectly fit the data that means, Training and testing accuracy is 100%.

30. Now, think that you increase the complexity (or degree of polynomial of this kernel). What would you think will happen?

A) Increasing the complexity will over fit the data
B) Increasing the complexity will under fit the data
C) Nothing will happen since your model was already 100% accurate
D) None of these

Solution: A

Increasing the complexity of the data would make the algorithm overfit the data.

31. In the previous question after increasing the complexity you found that training accuracy was still 100%. According to you what is the reason behind that?

1. Since data is fixed and we are fitting more polynomial term or parameters so the algorithm starts memorizing everything in the data
2. Since data is fixed and SVM doesn't need to search in big hypothesis space

A) 1
B) 2
C) 1 and 2
D) None of these

Solution: C

Both the given statements are correct.

32. **What is/are true about kernel in SVM?**

1. Kernel function map low dimensional data to high dimensional space
2. It's a similarity function

A) 1
B) 2
C) 1 and 2
D) None of these

Solution: C

Both the given statements are correct.

UNIT V

1. **Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?**

a) Decision Tree
b) Regression
c) Classification
d) Random Forest

Ans D

2. Which of the following is a disadvantage of decision trees?

a) Factor analysis
b) Decision trees are robust to outliers
c) Decision trees are prone to be overfit

d) None of the above

Ans **C**

3. **Can decision trees be used for performing clustering?**

a. True
b. False

Ans **Solution: (A)**

Decision trees can also be used to for clusters in the data but clustering often generates natural clusters and is not dependent on any objective function.

4. **Which of the following algorithm is most sensitive to outliers?**

a. K-means clustering algorithm
b. K-medians clustering algorithm
c. K-modes clustering algorithm
d. K-medoids clustering algorithm

Ans Solution: (A)

5 **Sentiment Analysis is an example of:**

Regression

Classification

Clustering

Reinforcement Learning

Options:

a. 1 Only
b. 1 and 2
c. 1 and 3
d. 1, 2 and 4

Ans **D**

6 **Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:**

Capping and flouring of variables

Removal of outliers

Options:

a. 1 only
b. 2 only
c. 1 and 2
d. None of the above

Ans **A**

7 **Which of the following is/are true about bagging trees?**

1. In bagging trees, individual trees are independent of each other
2. Bagging is the method for improving the performance by aggregating the results of weak learners

A) 1
B) 2
C) 1 and 2
D) None of these

Ans **Solution: C**

Both options are true. In Bagging, each individual trees are independent of each other because they consider different subset of features and samples.

8. **Which of the following is/are true about boosting trees?**

1. In boosting trees, individual weak learners are independent of each other
2. It is the method for improving the performance by aggregating the results of weak learners

A) 1
B) 2
C) 1 and 2
D) None of **these**

Ans **Solution: B**

In boosting tree individual weak learners are not independent of each other because each tree correct the results of previous tree. Bagging and boosting both can be consider as improving the base learners results.

9. In Random forest you can generate hundreds of trees (say T1, T2 …..Tn) and then aggregate the results of these tree. Which of the following is true about individual (Tk) tree in Random Forest?

1. Individual tree is built on a subset of the features
2. Individual tree is built on all the features
3. Individual tree is built on a subset of observations
4. Individual tree is built on full set of observations

A) 1 and 3
B) 1 and 4
C) 2 and 3
D) 2 and 4

Ans Solution: A

Random forest is based on bagging concept, that consider faction of sample and faction of feature for building the individual trees.

10. Suppose you are using a bagging based algorithm say a RandomForest in model building. Which of the following can be true?

1. Number of tree should be as large as possible
2. You will have interpretability after using Random Forest

A) 1
B) 2
C) 1 and 2
D) None of these

Ans Solution: A

Since Random Forest aggregate the result of different weak learners, If It is possible we would want more number of trees in model building. Random Forest is a black box model you will lose interpretability after using it.

11. Which of the following is/are true about Random Forest and Gradient Boosting ensemble methods?

1. Both methods can be used for classification task
2. Random Forest is use for classification whereas Gradient Boosting is use for regression task
3. Random Forest is use for regression whereas Gradient Boosting is use for Classification task
4. Both methods can be used for regression task

A) 1
B) 2
C) 3
D) 4
E) 1 and 4

Solution: E

Both algorithms are design for classification as well as regression task.

12. In Random forest you can generate hundreds of trees (say T1, T2 …..Tn) and then aggregate the results of these tree. Which of the following is true about individual(Tk) tree in Random Forest?

1. Individual tree is built on a subset of the features
2. Individual tree is built on all the features
3. Individual tree is built on a subset of observations
4. Individual tree is built on full set of observations

A) 1 and 3
B) 1 and 4
C) 2 and 3
D) 2 and 4

Solution: A

Random forest is based on bagging concept, that consider faction of sample and faction of feature for building the individual trees.

13. **Which of the following algorithm doesn't uses learning Rate as of one of its hyperparameter?**

1. Gradient Boosting
2. Extra Trees
3. AdaBoost
4. Random Forest

A) 1 and 3
B) 1 and 4
C) 2 and 3
D) 2 and 4

Solution: D

Random Forest and Extra Trees don't have learning rate as a hyperparameter.

14. **Which of the following algorithm are not an example of ensemble learning algorithm?**

A) Random Forest
B) Adaboost
C) Extra Trees
D) Gradient Boosting
E) Decision Trees

Solution: E

Decision trees doesn't aggregate the results of multiple trees so it is not an ensemble algorithm.

15. **Suppose you are using a bagging based algorithm say a RandomForest in model building. Which of the following can be true?**

1. Number of tree should be as large as possible

2. You will have interpretability after using RandomForest

A) 1
B) 2
C) 1 and 2
D) None of these

Solution: A

Since Random Forest aggregate the result of different weak learners, If It is possible we would want more number of trees in model building. Random Forest is a black box model you will lose interpretability after using it.

16. **True-False: The bagging is suitable for high variance low bias models?**

A) TRUE
B) FALSE

Solution: A

The bagging is suitable for high variance low bias models or you can say for complex models.

17. **To apply bagging to regression trees which of the following is/are true in such case?**

1. We build the N regression with N bootstrap sample

2. We take the average the of N regression tree

3. Each tree has a high variance with low bias

A) 1 and 2
B) 2 and 3
C) 1 and 3
D) 1,2 and 3

Solution: D

All of the options are correct and self-explanatory

18. **How to select best hyper parameters in tree based models?**

A) Measure performance over training data
B) Measure performance over validation data
C) Both of these
D) None of these

Solution: B

We always consider the validation results to compare with the test result.

19. **In which of the following scenario a gain ratio is preferred over Information Gain?**

A) When a categorical variable has very large number of category
B) When a categorical variable has very small number of category
C) Number of categories is the not the reason
D) None of these

Solution: A

When high cardinality problems, gain ratio is preferred over Information Gain technique.

20. **Suppose you have given the following scenario for training and validation error for Gradient Boosting. Which of the following hyper parameter would you choose in such case?**

Scenario	**Depth**	**Training Error**	**Validation Error**
1	2	100	110
2	4	90	105
3	6	50	100
4	8	45	105

5	10	30	150

A) 1
B) 2
C) 3
D) 4

Solution: B

Scenario 2 and 4 has same validation accuracies but we would select 2 because depth is lower is better hyper parameter.

21. **Which of the following is/are not true about DBSCAN clustering algorithm:**

    1. **For data points to be in a cluster, they must be in a distance threshold to a core point**
    2. **It has strong assumptions for the distribution of data points in** dataspace
    3. **It has substantially high time complexity of order O(n^3)**
    4. **It does not require prior knowledge of the no. of desired clusters**
    5. **It is robust to outliers**

Options:

A. 1 only

B. 2 only

C. 4 only

D. 2 and 3

Solution: D

- DBSCAN can form a cluster of any arbitrary shape and does not have strong assumptions for the distribution of data points in the data space.
- DBSCAN has a low time complexity of order O (n log n) only.

22. Point out the correct statement.
a) The choice of an appropriate metric will influence the shape of the clusters
b) Hierarchical clustering is also called HCA
c) In general, the merges and splits are determined in a greedy manner
d) All of the mentioned

Answer: d

Explanation: Some elements may be close to one another according to one distance and farther away according to another.

23. Which of the following is required by K-means clustering?

a) defined distance metric

b) number of clusters

c) initial guess as to cluster centroids

d) all of the mentioned

Answer: d

Explanation: K-means clustering follows partitioning approach.

24. Point out the wrong statement.

a) k-means clustering is a method of vector quantization

b) k-means clustering aims to partition n observations into k clusters

c) k-nearest neighbor is same as k-means

d) none of the mentioned

Answer: c

Explanation: k-nearest neighbour has nothing to do with k-means.

25. Which of the following function is used for k-means clustering?

a) k-means

b) k-mean

c) heat map

d) none of the mentioned

Answer: a

Explanation: K-means requires a number of clusters.

26. K-means is not deterministic and it also consists of number of iterations.

a) True

b) False

Answer: a

Explanation: K-means clustering produces the final estimate of cluster centroids.

27.

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Choose the options that is incorrect regarding machine learning (ML) and artificial intelligence (AI)
((OPTION_A)) THIS IS MANDATORY OPTION	ML is an alternate way of programming intelligent machines.
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	ML and AI have very different goals
((OPTION_C)) This is optional	ML is a set of techniques that turns a dataset into a software.
((OPTION_D)) This is optional	AI is a software that can emulate the human mind
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Which of the following sentence is FALSE regarding regression
((OPTION_A)) THIS IS MANDATORY OPTION	It is used for prediction
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	It may be used for interpretation
((OPTION_C)) This is optional	It relates inputs to outputs.
((OPTION_D)) This is optional	It discovers causal relationships
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	D
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Grid search is
((OPTION_A)) THIS IS MANDATORY OPTION	Linear in D
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Exponential in D
((OPTION_C)) This is optional	Linear in N
((OPTION_D)) This is optional	Both B&C
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	D
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Find incorrect regarding Gradient of a continuous and differentiable function
((OPTION_A)) THIS IS MANDATORY OPTION	is zero at a minimum
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	is non-zero at a maximum
((OPTION_C)) This is optional	is zero at a saddle point
((OPTION_D)) This is optional	decreases as you get closer to the minimum
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Consider a linear-regression model with $N = 3$ and $D = 1$ with input-ouput pairs as follows: $y_1 = 22$, $x_1 = 1$, $y_2 = 3$, $x_2 = 1$, $y_3 = 3$, $x_3 = 2$. What is the gradient of mean-square error (MSE) with respect to β_1 when $\beta_0 = 0$ and $\beta_1 = 1$? Give your answer correct to two decimal digits.
((OPTION_A)) THIS IS MANDATORY OPTION	-1.66
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	2
((OPTION_C)) This is optional	3
((OPTION_D)) This is optional	4
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Let us say that we have computed the gradient of our cost function and stored it in a vector **g**. What is the cost of one gradient descent update given the gradient?
((OPTION_A)) THIS IS MANDATORY OPTION	$O(D)$
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	$O(N)$
((OPTION_C)) This is optional	$O(ND)$
((OPTION_D)) This is optional	$O(ND_2)$
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	You observe the following while fitting a linear regression to the data: As you increase the amount of training data, the test error decreases and the training error increases. The train error is quite low (almost what you expect it to), while the test error is much higher than the train error. What do you think is the main reason behind this behavior. Choose the most probable option
((OPTION_A)) THIS IS MANDATORY OPTION	High variance
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	High model bias
((OPTION_C)) This is optional	High estimation bias
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Adding more basis functions in a linear model... (pick the most probably option)
((OPTION_A)) THIS IS MANDATORY OPTION	Decreases model bias
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Decreases estimation bias
((OPTION_C)) This is optional	Decreases variance
((OPTION_D)) This is optional	Doesn’t affect bias and variance
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**The problem of finding hidden structure in unlabeled data is called**
((OPTION_A)) THIS IS MANDATORY OPTION	Supervised learning
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	UnSupervised learning
((OPTION_C)) This is optional	Reinforcement learning
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Task of inferring a model from labeled training data is called**
((OPTION_A)) THIS IS MANDATORY OPTION	Unsupervised learning
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	supervised learning
((OPTION_C)) This is optional	Reinforcement learning
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Some telecommunication company wants to segment their customers into distinct groups in order to send appropriate subscription offers, this is an example of**
((OPTION_A)) THIS IS MANDATORY OPTION	Supervised learning
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Data extraction
((OPTION_C)) This is optional	Serration
((OPTION_D)) This is optional	Unsupervised learning
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	D
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Self-organizing maps are an example of**
((OPTION_A)) THIS IS MANDATORY OPTION	Unsupervised learning
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Supervised learning
((OPTION_C)) This is optional	Reinforcement learning
((OPTION_D)) This is optional	Missing data imputation
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**You are given data about seismic activity in Japan, and you want to predict a magnitude of the next earthquake, this is in an example of**
((OPTION_A)) THIS IS MANDATORY OPTION	Supervised learning
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Unsupervised learning
((OPTION_C)) This is optional	Serration
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Assume you want to perform supervised learning and to predict number of newborns according to size of storks' population (http://www.brixtonhealth.com/storksBabies.pdf), it is an example of**
((OPTION_A)) THIS IS MANDATORY OPTION	Classification
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Regression
((OPTION_C)) This is optional	Clustering
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Discriminating between spam and ham e-mails is a classification task, true or false?**
((OPTION_A)) THIS IS MANDATORY OPTION	True
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	False
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**In the example of predicting number of babies based on storks' population size, number of babies is**
((OPTION_A)) THIS IS MANDATORY OPTION	Outcome
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Feature
((OPTION_C)) This is optional	Attribute
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**It may be better to avoid the metric of ROC curve as it can suffer from accuracy paradox.**
((OPTION_A)) THIS IS MANDATORY OPTION	True
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	False
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**which of the following is not involve in data mining**
((OPTION_A)) THIS IS MANDATORY OPTION	Knowledge extraction
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Data archaeology
((OPTION_C)) This is optional	Data exploration
((OPTION_D)) This is optional	Data transformation
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	D
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	The expected value or ________ of a random variable is the center of its distribution.
((OPTION_A)) THIS IS MANDATORY OPTION	Mode
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	median
((OPTION_C)) This is optional	mean
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Point out the correct statement.
((OPTION_A)) THIS IS MANDATORY OPTION	Some cumulative distribution function F is non-decreasing and right-continuous
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Every cumulative distribution function F is decreasing and right-continuous
((OPTION_C)) This is optional	Every cumulative distribution function F is increasing and left-continuous
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	D
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Which of the following of a random variable is a measure of spread
((OPTION_A)) THIS IS MANDATORY OPTION	variance
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	standard deviation
((OPTION_C)) This is optional	empirical mean
((OPTION_D)) This is optional	All above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	The square root of the variance is called the _________ deviation
((OPTION_A)) THIS IS MANDATORY OPTION	empirical
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	mean
((OPTION_C)) This is optional	continuous
((OPTION_D)) This is optional	standard
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	D
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	For continuous random variables, the CDF is the derivative of the PDF.
((OPTION_A)) THIS IS MANDATORY OPTION	True
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	False
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Cumulative distribution functions are used to specify the distribution of multivariate random variables.
((OPTION_A)) THIS IS MANDATORY OPTION	True
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	False
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Consider the results of a medical experiment that aims to predict whether someone is going to develop myopia based on some physical measurements and heredity. In this case, the input dataset consists of the person's medical characteristics and the target variable is binary: 1 for those who are likely to develop myopia and 0 for those who aren't. This can be best classified as
((OPTION_A)) THIS IS MANDATORY OPTION	Regression
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Desicion Tree
((OPTION_C)) This is optional	Clustering
((OPTION_D)) This is optional	Association Rule
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	The purpose of a machine learning model is to approximate an unknown function that associates input elements to output ones
((OPTION_A)) THIS IS MANDATORY OPTION	True
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	False
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Training set is normally a representation of a global distribution
((OPTION_A)) THIS IS MANDATORY OPTION	True
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	False
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	2
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	The model has an excessive capacity and it's not more able to generalize considering the original dynamics provided by the training set. This problem is called as
((OPTION_A)) THIS IS MANDATORY OPTION	Underfitting
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Overfitting
((OPTION_C)) This is optional	Both
((OPTION_D)) This is optional	None
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	It can associate almost perfectly all the known samples to the corresponding output values, but when an unknown input is presented, the corresponding prediction error can be very high, This problem is called as
((OPTION_A)) THIS IS MANDATORY OPTION	Underfitting
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Overfitting
((OPTION_C)) This is optional	Both
((OPTION_D)) This is optional	None
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	---------- may prove to be more difficult to discover as it could be initially considered the result of a perfect fitting
((OPTION_A)) THIS IS MANDATORY OPTION	Underfitting
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Overfitting
((OPTION_C)) This is optional	Both
((OPTION_D)) This is optional	None
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	when working with a supervised scenario, we define a non-negative error measure e_m which takes two arguments and allows us to compute a total error value over the whole dataset. Those two arguments are.
((OPTION_A)) THIS IS MANDATORY OPTION	expected and predicted output
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	calculated and predicted output
((OPTION_C)) This is optional	calculated and measured output
((OPTION_D)) This is optional	none
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Initial value represents a starting point over the surface of a n-variables function. A generic training algorithm has to find the global minimum or a point quite close to it (there's always a tolerance to avoid an excessive number of iterations and a consequent risk of overfitting). This measure is also called
((OPTION_A)) THIS IS MANDATORY OPTION	loss function
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	predicted output
((OPTION_C)) This is optional	measured output
((OPTION_D)) This is optional	mean square error
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	In 1984, the computer scientist L. Valiant proposed a mathematical approach to determine whether a problem is learnable by a computer. The name of this technique is
((OPTION_A)) THIS IS MANDATORY OPTION	Max likelihood
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Zero one loss error
((OPTION_C)) This is optional	Probably approximately correct
((OPTION_D)) This is optional	none
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	In particular, a concept is a subset of input patterns X which determine the same output element
((OPTION_A)) THIS IS MANDATORY OPTION	True
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	False
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	Therefore, learning a concept (parametrically) means minimizing the corresponding loss function restricted to a specific class, while learning all possible concepts (belonging to the same universe), means finding the minimum of a global loss function
((OPTION_A)) THIS IS MANDATORY OPTION	True
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	False
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	An exponential time could lead to computational explosions when the datasets are too large or the optimization starting point is very far from an acceptable minimum. Moreover, it's important to remember the so-called ....
((OPTION_A)) THIS IS MANDATORY OPTION	**curse of dimensionality**
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	**Hughes phenomenon**
((OPTION_C)) This is optional	Probably approximately correct
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	In many cases, in order to capture the full expressivity, it's necessary to have a very large dataset and without enough training data, the approximation can become problematic. This is called...
((OPTION_A)) THIS IS MANDATORY OPTION	**curse of dimensionality**
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	**Hughes phenomenon**
((OPTION_C)) This is optional	Probably approximately correct
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	$$P(h_{pi} \mid X) \propto P(X \mid h_{pi}) P(h_{pi})$$ First term is called as
((OPTION_A)) THIS IS MANDATORY OPTION	**posteriori**
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	**Apriori**
((OPTION_C)) This is optional	**likelihood**.
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	$P(h_{pi} \mid X) \propto P(X \mid h_{pi}) P(h_{pi})$ second term is called as
((OPTION_A)) THIS IS MANDATORY OPTION	**posteriori**
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	**Apriori**
((OPTION_C)) This is optional	**likelihood**.
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	**B**
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	$$P(h_{pi} \mid X) \propto P(X \mid h_{pi}) P(h_{pi})$$ Third term is called as
((OPTION_A)) THIS IS MANDATORY OPTION	**posteriori**
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	**Apriori**
((OPTION_C)) This is optional	**likelihood**.
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	We can create the object of abstract class
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which of the following step / assumption in regression modeling impacts the trade-off between under-fitting and over-fitting the most**
((OPTION_A)) THIS IS MANDATORY OPTION	The polynomial degree
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Whether we learn the weights by matrix inversion or gradient descent
((OPTION_C)) This is optional	The use of a constant-term
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Suppose you have the following data with one real-value input variable & one real-value output variable. What is leave-one out cross validation mean square error in case of linear regression (Y = bX+c)?** X(independent variable) | Y(dependent variable) 0 | 2 2 | 2 3 | 1
((OPTION_A)) THIS IS MANDATORY OPTION	10/27
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	20/27
((OPTION_C)) This is optional	50/27
((OPTION_D)) This is optional	49/27
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	D
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which of the following is/ are true about "Maximum Likelihood estimate (MLE)"?** 1. **MLE may not always exist** 2. **MLE always exists** 3. **If MLE exist, it (they) may not be unique** 4. **If MLE exist, it (they) must be unique**
((OPTION_A)) THIS IS MANDATORY OPTION	1and4
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	2 and3
((OPTION_C)) This is optional	1 and3
((OPTION_D)) This is optional	2 and4
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Let's say, a "Linear regression" model perfectly fits the training data (train error is zero). Now, Which of the following statement is true?**
((OPTION_A)) THIS IS MANDATORY OPTION	You will always have test error zero
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	. You can not have test error zero
((OPTION_C)) This is optional	None of the above
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which one of the statement is true regarding residuals in regression analysis?**
((OPTION_A)) THIS IS MANDATORY OPTION	A. Mean of residuals is always zero
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Mean of residuals is always less than zero
((OPTION_C)) This is optional	Mean of residuals is always greater than zero
((OPTION_D)) This is optional	There is no such rule for residuals.
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which of the one is true about Heteroskedasticity?**
((OPTION_A)) THIS IS MANDATORY OPTION	Linear Regression with varying error terms
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Linear Regression with constant error terms
((OPTION_C)) This is optional	Linear Regression with zero error terms
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

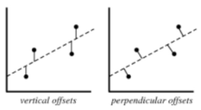
((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which of the following indicates a fairly strong relationship between X and Y?**
((OPTION_A)) THIS IS MANDATORY OPTION	A. Correlation coefficient = 0.9
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	. The p-value for the null hypothesis Beta coefficient =0 is 0.0001
((OPTION_C)) This is optional	The t-statistic for the null hypothesis Beta coefficient=0 is 30
((OPTION_D)) This is optional	None of these
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which of the following assumptions do we make while deriving linear regression param** 1. **The true relationship between dependent y and predictor x is linear** 2. **The model errors are statistically independent** 3. **The errors are normally distributed with a 0 mean and constant standard deviation.**
((OPTION_A)) THIS IS MANDATORY OPTION	1,2&3
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	1&3
((OPTION_C)) This is optional	**All of above**
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**To test linear relationship of y(dependent) and x(independent) continuous variables, which of the following plot best suited?**
((OPTION_A)) THIS IS MANDATORY OPTION	Scatter plot
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Barchart
((OPTION_C)) This is optional	Histograms
((OPTION_D)) This is optional	None of these
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Generally, which of the following method(s) is used for predicting continuous dependent variable?** 1. **Linear Regression** 2. **Logistic Regression**
((OPTION_A)) THIS IS MANDATORY OPTION	1&2
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Only 1
((OPTION_C)) This is optional	Only 2
((OPTION_D)) This is optional	None f the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**. A correlation between age and health of a person found to be -1.09. On the basis of this you would tell the doctors that:**
((OPTION_A)) THIS IS MANDATORY OPTION	. The age is good predictor of health
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	. The age is poor predictor of health
((OPTION_C)) This is optional	None of these
((OPTION_D)) This is optional	All of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which of the following offsets, do we use in case of least square line fit? Suppose horizontal axis is independent variable and vertical axis is dependent variable** 
((OPTION_A)) THIS IS MANDATORY OPTION	Vertical offset
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Perpendicular offset
((OPTION_C)) This is optional	Both but depend on situation
((OPTION_D)) This is optional	Both a&b
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Suppose we have generated the data with help of polynomial regression of degree 3 (degree 3 will perfectly fit this data). Now consider below points and choose the option based on these points.** 1. **Simple Linear regression will have high bias and low variance** 2. **Simple Linear regression will have low bias and high variance** 3. **polynomial of degree 3 will have low bias and high variance** **Polynomial of degree 3 will have low bias and Low variance**
((OPTION_A)) THIS IS MANDATORY OPTION	. Only 1
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	1&3
((OPTION_C)) This is optional	1&4
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**. Suppose you are training a linear regression model. Now consider these points.** 1. **Overfitting is more likely if we have less data** 2. **Overfitting is more likely when the hypothesis space is small** Which of the above statement(s) are correct?
((OPTION_A)) THIS IS MANDATORY OPTION	Both are False
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	1 is False and 2 is True
((OPTION_C)) This is optional	1 is True and 2 is False
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	c
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Suppose we fit "Lasso Regression" to a data set, which has 100 features (X1,X2…X100). Now, we rescale one of these feature by multiplying with 10 (say that feature is X1), and then refit Lasso regression with the same regularization parameter.** **Now, which of the following option will be correct?**
((OPTION_A)) THIS IS MANDATORY OPTION	It is more likely for X1 to be excluded from the model
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	It is more likely for X1 to be included in the model
((OPTION_C)) This is optional	. Can't say
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which of the following is true about "Ridge" or "Lasso" regression methods in case of feature selection?**
((OPTION_A)) THIS IS MANDATORY OPTION	Ridge regression uses subset selection of features
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	. Lasso regression uses subset selection of features
((OPTION_C)) This is optional	Both use subset selection of features
((OPTION_D)) This is optional	All of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

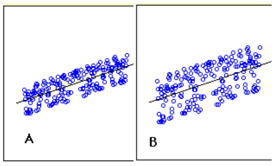
((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	. **Which of the following statement(s) can be true post adding a variable in a linear regression model?** 1. **R-Squared and Adjusted R-squared both increase** 2. **R-Squared increases and Adjusted R-squared decreases** 3. **R-Squared decreases and Adjusted R-squared decreases** 4. **R-Squared decreases and Adjusted R-squared increases**
((OPTION_A)) THIS IS MANDATORY OPTION	. 1 and 2
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	1 and 3
((OPTION_C)) This is optional	2 and 4
((OPTION_D)) This is optional	none of these
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**. Which of the following metrics can be used for evaluating regression models?** 1. **R Squared** 2. **Adjusted R Squared** 3. **F Statistics** 1. **RMSE / MSE / MAE**
((OPTION_A)) THIS IS MANDATORY OPTION	2 and 4
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	1 and 2.
((OPTION_C)) This is optional	. 2, 3 and 4.
((OPTION_D)) This is optional	All of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	D
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**We can also compute the coefficient of linear regression with the help of an analytical method called "Normal Equation". Which of the following is/are true about "Normal Equation"?** 1. **We don't have to choose the learning rate** 2. **It becomes slow when number of features is very large** 3. **No need to iterate**
((OPTION_A)) THIS IS MANDATORY OPTION	1 and 2
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	1&3
((OPTION_C)) This is optional	2&3
((OPTION_D)) This is optional	1,2&3
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	D
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**. The expected value of Y is a linear function of the X(X1,X2….Xn) variables and regression line is defined as:** **$Y = \beta 0 + \beta 1\ X1 + \beta 2\ X2 \ldots\ldots + \beta n\ Xn$** **Which of the following statement(s) are true?** 1. **If Xi changes by an amount ∆Xi, holding other variables constant, then the expected value of Y changes by a proportional amount βi ∆Xi, for some constant βi (which in general could be a positive or negative number).** 2. **The value of βi is always the same, regardless of values of the other X's.** 3. **The total effect of the X's on the expected value of Y is the sum of their separate effects.**
((OPTION_A)) THIS IS MANDATORY OPTION	. 1 and 2
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	1 and 3
((OPTION_C)) This is optional	2 and 3
((OPTION_D)) This is optional	1,2 and 3
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	D
((EXPLANATION)) This is also optional	

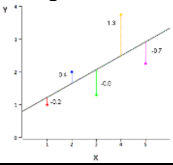
((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**. How many coefficients do you need to estimate in a simple linear regression model (One independent variable)**
((OPTION_A)) THIS IS MANDATORY OPTION	1
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	2
((OPTION_C)) This is optional	CAN'T SAY
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	2
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**. Below graphs show two fitted regression lines (A & B) on randomly generated data. Now, I want to find the sum of residuals in both cases A and B.**  **Which of the following statement is true about sum of residuals of A and B**
((OPTION_A)) THIS IS MANDATORY OPTION	A has higher than B
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	A has lower than B
((OPTION_C)) This is optional	Both have same
((OPTION_D)) This is optional	None of these
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**If two variables are correlated, is it necessary that they have a linear relationsh**
((OPTION_A)) THIS IS MANDATORY OPTION	YES
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	NO
((OPTION_C)) This is optional	Both a&b
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Correlated variables can have zero correlation coeffficient. True or False?**
((OPTION_A)) THIS IS MANDATORY OPTION	TRUE
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	FALSE
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Suppose I applied a logistic regression model on data and got training accuracy X and testing accuracy Y. Now I want to add few new features in data. Select option(s) which are correct in such case.** **Note: Consider remaining parameters are same.** 1. **Training accuracy always decreases.** 2. **Training accuracy always increases or remain same.** 3. **Testing accuracy always decreases** **Testing accuracy always increases or remain same**
((OPTION_A)) THIS IS MANDATORY OPTION	Only 2
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Only 1
((OPTION_C)) This is optional	Only3
((OPTION_D)) This is optional	All of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**The graph below represents a regression line predicting Y from X. The values on the graph shows the residuals for each predictions value. Use this information to compute the SSE.** 
((OPTION_A)) THIS IS MANDATORY OPTION	3.02
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	0.75
((OPTION_C)) This is optional	1.01
((OPTION_D)) This is optional	None of these
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Suppose the distribution of salaries in a company X has median \$35,000, and 25th and 75th percentiles are \$21,000 and \$53,000 respectively.** **Would a person with Salary \$1 be considered an Outlier?**
((OPTION_A)) THIS IS MANDATORY OPTION	YES
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	NO
((OPTION_C)) This is optional	. More information is required
((OPTION_D)) This is optional	None of these
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which of the following option is true regarding "Regression" and "Correlation" ?** **Note: y is dependent variable and x is independent variable.**
((OPTION_A)) THIS IS MANDATORY OPTION	The relationship is symmetric between x and y in both.
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	The relationship is not symmetric between x and y in both.
((OPTION_C)) This is optional	The relationship is not symmetric between x and y in case of correlation but in case of regression it is symmetric.
((OPTION_D)) This is optional	The relationship is symmetric between x and y in case of correlation but in case of regression it is not symmetric.
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**True-False: Is Logistic regression a supervised machine learning algorithm?**
((OPTION_A)) THIS IS MANDATORY OPTION	TRUE
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	FALSE
((OPTION_C)) This is optional	_
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**True-False: Is Logistic regression mainly used for Regression?**
((OPTION_A)) THIS IS MANDATORY OPTION	TRUE
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	FALSE
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**True-False: Is it possible to design a logistic regression algorithm using a Neural Network Algorithm?**
((OPTION_A)) THIS IS MANDATORY OPTION	TRUE
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	FALSE
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**True-False: Is it possible to apply a logistic regression algorithm on a 3-class Classification problem?**
((OPTION_A)) THIS IS MANDATORY OPTION	TRUE
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	FALSE
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which of the following methods do we use to best fit the data in Logistic Regression?**
((OPTION_A)) THIS IS MANDATORY OPTION	Least Square Error
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Maximum Likelihood
((OPTION_C)) This is optional	Jaccard distance
((OPTION_D)) This is optional	Both a&B
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**One of the very good methods to analyze the performance of Logistic Regression is AIC, which is similar to R-Squared in Linear Regression. Which of the following is true about AIC**
((OPTION_A)) THIS IS MANDATORY OPTION	We prefer a model with minimum AIC value
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	We prefer a model with maximum AIC value
((OPTION_C)) This is optional	Both but depend on the situation
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**True-False] Standardisation of features is required before training a Logistic Regression**
((OPTION_A)) THIS IS MANDATORY OPTION	TRUE
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	FALSE
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which of the following algorithms do we use for Variable Selection?**
((OPTION_A)) THIS IS MANDATORY OPTION	) LASSO
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Ridge
((OPTION_C)) This is optional	Both
((OPTION_D)) This is optional	All of these
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Suppose you have been given a fair coin and you want to find out the odds of getting heads. Which of the following option is true for such a case?**
((OPTION_A)) THIS IS MANDATORY OPTION	odds will be 0
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	odds will be 0.5
((OPTION_C)) This is optional	odds will be 1
((OPTION_D)) This is optional	None of the above
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**) The logit function(given as l(x)) is the log of odds function. What could be the range of logit function in the domain x=[0,1]?**
((OPTION_A)) THIS IS MANDATORY OPTION	$(-\infty, \infty)$
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	(0,1)
((OPTION_C)) This is optional	$(0, \infty)$
((OPTION_D)) This is optional	$(-\infty, 0)$
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

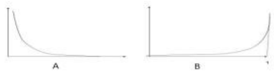
((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which of the following option is true?**
((OPTION_A)) THIS IS MANDATORY OPTION	Linear Regression errors values has to be normally distributed but in case of Logistic Regression it is not the case
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Linear Regression errors values has to be normally distributed but in case of Logistic Regression it is not the case
((OPTION_C)) This is optional	Both Linear Regression and Logistic Regression error values have to be normally distributed
((OPTION_D)) This is optional	Both Linear Regression and Logistic Regression error values have not to be normally distributed
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**17) Which of the following is true regarding the logistic function for any value "x Note:** Logistic(x): is a logistic function of any number "x" Logit(x): is a logit function of any number "x" Logit_inv(x): is a inverse logit function of any number "x"**"?**
((OPTION_A)) THIS IS MANDATORY OPTION	C) A) Logistic(x) = Logit(x)
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Logistic(x) = Logit_inv(x)
((OPTION_C)) This is optional	A) Logistic(x) = Logit(x)
((OPTION_D)) This is optional	None of these
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	2
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Suppose, You applied a Logistic Regression model on a given data and got a training accuracy X and testing accuracy Y. Now, you want to add a few new features in the same data. Select the option(s) which is/are correct in such a case.** **Note: Consider remaining parameters are same.**
((OPTION_A)) THIS IS MANDATORY OPTION	Training accuracy increases
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Training accuracy increases or remains the same
((OPTION_C)) This is optional	Testing accuracy decreases
((OPTION_D)) This is optional	Testing accuracy increases or remains the same
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A&D
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Choose which of the following options is true regarding One-Vs-All method in Logistic Regression.**
((OPTION_A)) THIS IS MANDATORY OPTION	We need to fit n models in n-class classification problem
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	We need to fit n-1 models to classify into n classes
((OPTION_C)) This is optional	We need to fit only 1 model to classify into n classes
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**What would do if you want to train logistic regression on same data that will take less time as well as give the comparatively similar accuracy(may not be same)?** Suppose you are using a Logistic Regression model on a huge dataset. One of the problem you may face on such huge data is that Logistic regression will take very long time to train
((OPTION_A)) THIS IS MANDATORY OPTION	Decrease the learning rate and decrease the number of iteration
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Decrease the learning rate and increase the number of iteration
((OPTION_C)) This is optional	Increase the learning rate and increase the number of iteration
((OPTION_D)) This is optional	Increase the learning rate and decrease the number of iteration
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	D
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	2
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Which of the following image is showing the cost function for y =1.** **Following is the loss function in logistic regression(Y-axis loss function and x axis log probability) for two class classification problem.** **Note: Y is the target class** 
((OPTION_A)) THIS IS MANDATORY OPTION	A
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	B
((OPTION_C)) This is optional	BOTH
((OPTION_D)) This is optional	NON OF THESE
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Logistic regression is used when you want to:**
((OPTION_A)) THIS IS MANDATORY OPTION	Predict a dichotomous variable from continuous or dichotomous variables.
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Predict a continuous variable from dichotomous variables.
((OPTION_C)) This is optional	Predict any categorical variable from several other categorical variables.
((OPTION_D)) This is optional	Predict a continuous variable from dichotomous or continuous variables
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	A
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**The odds ratio is**
((OPTION_A)) THIS IS MANDATORY OPTION	The ratio of the probability of an event not happening to the probability of the event happening.
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	The probability of an event occurring.
((OPTION_C)) This is optional	The ratio of the odds after a unit change in the predictor to the original odds.
((OPTION_D)) This is optional	The ratio of the probability of an event happening to the probability of the event not happening.
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Large values of the log-likelihood statistic indicate:**
((OPTION_A)) THIS IS MANDATORY OPTION	That there are a greater number of explained vs. unexplained observations.
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	That the statistical model fits the data well.
((OPTION_C)) This is optional	That as the predictor variable increases, the likelihood of the outcome occurring decreases.
((OPTION_D)) This is optional	That the statistical model is a poor fit of the data.
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**Logistic regression assumes a:**
((OPTION_A)) THIS IS MANDATORY OPTION	Linear relationship between continuous predictor variables and the outcome variable.
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	Linear relationship between continuous predictor variables and the logit of the outcome variable.
((OPTION_C)) This is optional	Linear relationship between continuous predictor variables.
((OPTION_D)) This is optional	Linear relationship between observations.
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	**In binary logistic regression:**
((OPTION_A)) THIS IS MANDATORY OPTION	The dependent variable is continuous.
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	The dependent variable is divided into two equal subcategories.
((OPTION_C)) This is optional	The dependent variable consists of two categories.
((OPTION_D)) This is optional	There is no dependent variable.
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	1
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	The correlation coefficient is used to determine
((OPTION_A)) THIS IS MANDATORY OPTION	A specific value of the y-variable given a specific value of the x-variable
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	A specific value of the x-variable given a specific value of the y-variable
((OPTION_C)) This is optional	The strength of the relationship between the x and y variables
((OPTION_D)) This is optional	none
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	C
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	B
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	2
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

((MARKS)) QUESTION IS OF HOW MANY MARKS? (1 OR 2 OR 3 UPTO 10)	
((QUESTION)) ENTER CONTENT. QTN CAN HAVE IMAGES ALSO	
((OPTION_A)) THIS IS MANDATORY OPTION	
((OPTION_B)) THIS IS ALSO MANDATORY OPTION	
((OPTION_C)) This is optional	
((OPTION_D)) This is optional	
((OPTION_E)) This is optional. If optional keep empty so that system will skip this option	
((CORRECT_CHOICE)) Either A or B or C or D or E	
((EXPLANATION)) This is also optional	

UNIT I

1. What is classification?
a) when the output variable is a **category,** such as “red” or “blue” or “disease” and “no disease”.
b) when the output variable is a **real value**, such as “dollars” or “weight”.

Ans: Solution A

2. What is regression?
a) When the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.
b) When the output variable is a real value, such as “dollars” or “weight”.

Ans: Solution B

3. What is supervised learning?
a) All data is unlabelled and the algorithms learn to inherent structure from the input data
b) All data is labelled and the algorithms learn to predict the output from the input data
c) It is a framework for learning where an agent interacts with an environment and receives a reward for each interaction
d) Some data is labelled but most of it is unlabelled and a mixture of supervised and unsupervised techniques can be used.

Ans: Solution B

4. What is Unsupervised learning?
a) All data is unlabelled and the algorithms learn to inherent structure from the input data
b) All data is labelled and the algorithms learn to predict the output from the input data
c) It is a framework for learning where an agent interacts with an environment and receives a reward for each interaction
d) Some data is labelled but most of it is unlabelled and a mixture of supervised and unsupervised techniques can be used.

Ans: Solution A

5. What is Semi-Supervised learning?
a) All data is unlabelled and the algorithms learn to inherent structure from the input data
b) All data is labelled and the algorithms learn to predict the output from the input data
c) It is a framework for learning where an agent interacts with an environment and receives a reward for each interaction
d) Some data is labelled but most of it is unlabelled and a mixture of supervised and unsupervised techniques can be used.

Ans: Solution D

6. What is Reinforcement learning?

a) All data is unlabelled and the algorithms learn to inherent structure from the input data
b) All data is labelled and the algorithms learn to predict the output from the input data
c) It is a framework for learning where an agent interacts with an environment and receives a reward for each interaction
d) Some data is labelled but most of it is unlabelled and a mixture of supervised and unsupervised techniques can be used.

Ans: Solution C

7. Sentiment Analysis is an example of:

Regression,

Classification

Clustering

Reinforcement Learning

Options:

A. 1 Only

B. 1 and 2

C. 1 and 3

D. 1, 2 and 4

Ans : Solution D

8. The process of forming general concept definitions from examples of concepts to be learned.

a) Deduction
b) abduction
c) induction
d) conjunction

Ans : Solution C

9. Computers are best at learning
    a) facts.
    b) concepts.
    c) procedures.
    d) principles.

Ans : Solution A

10. Data used to build a data mining model.
    a) validation data
    b) training data
    c) test data
    d) hidden data

Ans : Solution B

11. Supervised learning and unsupervised clustering both require at least one
    a) hidden attribute.
    b) output attribute.
    c) input attribute.
    d) categorical attribute.

Ans : Solution A

12. Supervised learning differs from unsupervised clustering in that supervised learning requires
    a) at least one input attribute.
    b) input attributes to be categorical.
    c) at least one output attribute.
    d) output attributes to be categorical.

Ans : Solution B

13. A regression model in which more than one independent variable is used to predict the dependent variable is called
    a) a simple linear regression model
    b) a multiple regression models
    c) an independent model
    d) none of the above

Ans : Solution C

14. A term used to describe the case when the independent variables in a multiple regression model are correlated is
    a) Regression
    b) correlation
    c) multicollinearity
    d) none of the above

Ans : Solution C

15. A multiple regression model has the form: y = 2 + 3x1 + 4x2. As x1 increases by 1 unit (holding x2 constant), y will
    a) increase by 3 units
    b) decrease by 3 units
    c) increase by 4 units
    d) decrease by 4 units

Ans : Solution C

16. A multiple regression model has
    a) only one independent variable
    b) more than one dependent variable
    c) more than one independent variable
    d) none of the above

Ans : Solution B

17. A measure of goodness of fit for the estimated regression equation is the
    a) multiple coefficient of determination
    b) mean square due to error
    c) mean square due to regression
    d) none of the above

Ans : Solution C

18. The adjusted multiple coefficient of determination accounts for
    a) the number of dependent variables in the model
    b) the number of independent variables in the model
    c) unusually large predictors
    d) none of the above

Ans : Solution D

19. The multiple coefficient of determination is computed by
    a) dividing SSR by SST
    b) dividing SST by SSR
    c) dividing SST by SSE
    d) none of the above

Ans : Solution C

20. For a multiple regression model, SST = 200 and SSE = 50. The multiple coefficient of determination is
    a) 0.25

b) 4.00
c) 0.75
d) none of the above

Ans : Solution B

21. A nearest neighbor approach is best used
    a) with large-sized datasets.
    b) when irrelevant attributes have been removed from the data.
    c) when a generalized model of the data is desirable.
    d) when an explanation of what has been found is of primary importance.

Ans : Solution B

22. Another name for an output attribute.
    a) predictive variable
    b) independent variable
    c) estimated variable
    d) dependent variable

Ans : Solution B

23. Classification problems are distinguished from estimation problems in that
    a) classification problems require the output attribute to be numeric.
    b) classification problems require the output attribute to be categorical.
    c) classification problems do not allow an output attribute.
    d) classification problems are designed to predict future outcome.

Ans : Solution C

24. Which statement is true about prediction problems?
    a) The output attribute must be categorical.
    b) The output attribute must be numeric.
    c) The resultant model is designed to determine future outcomes.
    d) The resultant model is designed to classify current behavior.

Ans : Solution D

25. Which statement about outliers is true?
    a) Outliers should be identified and removed from a dataset.
    b) Outliers should be part of the training dataset but should not be present in the test data.
    c) Outliers should be part of the test dataset but should not be present in the training data.
    d) The nature of the problem determines how outliers are used.

Ans : Solution D

26. Which statement is true about neural network and linear regression models?
    a) Both models require input attributes to be numeric.
    b) Both models require numeric attributes to range between 0 and 1.
    c) The output of both models is a categorical attribute value.
    d) Both techniques build models whose output is determined by a linear sum of weighted input attribute values.

Ans : Solution A

27. Which of the following is a common use of unsupervised clustering?
    a) detect outliers
    b) determine a best set of input attributes for supervised learning
    c) evaluate the likely performance of a supervised learner model
    d) determine if meaningful relationships can be found in a dataset

Ans : Solution A

28. The average positive difference between computed and desired outcome values.
    a) root mean squared error
    b) mean squared error
    c) mean absolute error
    d) mean positive error

Ans : Solution D

29. Selecting data so as to assure that each class is properly represented in both the training and test set.
    a) cross validation
    b) stratification
    c) verification
    d) bootstrapping

Ans : Solution B

30. The standard error is defined as the square root of this computation.
    a) The sample variance divided by the total number of sample instances.
    b) The population variance divided by the total number of sample instances.
    c) The sample variance divided by the sample mean.
    d) The population variance divided by the sample mean.

Ans : Solution A

31. Data used to optimize the parameter settings of a supervised learner model.

   a) Training
   b) Test
   c) Verification
   d) Validation

Ans : Solution D

32. Bootstrapping allows us to

   a) choose the same training instance several times.
   b) choose the same test set instance several times.
   c) build models with alternative subsets of the training data several times.
   d) test a model with alternative subsets of the test data several times.

Ans : Solution A

33. The correlation between the number of years an employee has worked for a company and the salary of the employee is 0.75. What can be said about employee salary and years worked?

   a) There is no relationship between salary and years worked.
   b) Individuals that have worked for the company the longest have higher salaries.
   c) Individuals that have worked for the company the longest have lower salaries.
   d) The majority of employees have been with the company a long time.
   e) The majority of employees have been with the company a short period of time.

Ans : Solution B

34. The correlation coefficient for two real-valued attributes is –0.85. What does this value tell you?

a) The attributes are not linearly related.
b) As the value of one attribute increases the value of the second attribute also increases.
c) As the value of one attribute decreases the value of the second attribute increases.
d) The attributes show a curvilinear relationship.

Ans : Solution C

35. The average squared difference between classifier predicted output and actual output.

   a) mean squared error
   b) root mean squared error
   c) mean absolute error
   d) mean relative error

Ans : Solution A

36. Simple regression assumes a __________ relationship between the input attribute and output attribute.

   a) Linear

b) Quadratic
c) reciprocal
d) inverse

Ans : Solution A

37. Regression trees are often used to model _______ data.
    a) Linear
    b) Nonlinear
    c) Categorical
    d) Symmetrical

Ans : Solution B

38. The leaf nodes of a model tree are
    a) averages of numeric output attribute values.
    b) nonlinear regression equations.
    c) linear regression equations.
    d) sums of numeric output attribute values.

Ans : Solution C

39. Logistic regression is a ________ regression technique that is used to model data having a _____outcome.
    a) linear, numeric
    b) linear, binary
    c) nonlinear, numeric
    d) nonlinear, binary

Ans : Solution D

40. This technique associates a conditional probability value with each data instance.
    a) linear regression
    b) logistic regression
    c) simple regression
    d) multiple linear regression

Ans : Solution B

41. This supervised learning technique can process both numeric and categorical input attributes.
    a) linear regression
    b) Bayes classifier
    c) logistic regression
    d) backpropagation learning

Ans : Solution A

42. With Bayes classifier, missing data items are
    a) treated as equal compares.
    b) treated as unequal compares.
    c) replaced with a default value.
    d) ignored.

Ans : Solution B

43. This clustering algorithm merges and splits nodes to help modify nonoptimal partitions.
    a) agglomerative clustering
    b) expectation maximization
    c) conceptual clustering
    d) K-Means clustering

Ans : Solution D

44. This clustering algorithm initially assumes that each data instance represents a single cluster.

a) agglomerative clustering
b) conceptual clustering
c) K-Means clustering
d) expectation maximization

Ans : Solution C

45. This unsupervised clustering algorithm terminates when mean values computed for the current iteration of the algorithm are identical to the computed mean values for the previous iteration.
    a) agglomerative clustering
    b) conceptual clustering
    c) K-Means clustering
    d) expectation maximization

Ans : Solution C

46. Machine learning techniques differ from statistical techniques in that machine learning methods
    a) typically assume an underlying distribution for the data.
    b) are better able to deal with missing and noisy data.
    c) are not able to explain their behavior.
    d) have trouble with large-sized datasets.

Ans : Solution B

UNIT –II

1.True- False: Over fitting is more likely when you have huge amount of data to train?

A) TRUE

B) FALSE

Ans **Solution: (B)**

With a small training dataset, it's easier to find a hypothesis to fit the training data exactly i.e. over fitting.

2.What is pca.components_ in Sklearn?

Set of all eigen vectors for the projection space

Matrix of principal components

Result of the multiplication matrix

None of the above options

Ans A

3.**Which of the following techniques would perform better for reducing dimensions of a data set?**

A. Removing columns which have too many missing values

B. Removing columns which have high variance in data

C. Removing columns with dissimilar data trends

D. None of these

Ans Solution: **(A)**

If a columns have too many missing values, (say 99%) then we can remove such columns.

4.**It is not necessary to have a target variable for applying dimensionality reduction algorithms.**

A. TRUE

B. FALSE

Ans Solution: **(A)**

LDA is an example of supervised dimensionality reduction algorithm.

5. **PCA can be used for projecting and visualizing data in lower dimensions.**

A. TRUE

B. FALSE

Ans Solution: **(A)**

Sometimes it is very useful to plot the data in lower dimensions. We can take the first 2 principal components and then visualize the data using scatter plot.

6. The most popularly used dimensionality reduction algorithm is Principal Component Analysis (PCA). Which of the following is/are true about PCA?

PCA is an unsupervised method

It searches for the directions that data have the largest variance
Maximum number of principal components <= number of features
All principal components are orthogonal to each other
A. 1 and 2
B. 1 and 3
C. 2 and 3
D. All of the above

Ans D

7. PCA works better if there is?
A linear structure in the data
If the data lies on a curved surface and not on a flat surface
If variables are scaled in the same unit
A. 1 and 2
B. 2 and 3
C. 1 and 3
D. 1 ,2 and 3
Ans Solution: **(C)**

8. What happens when you get features in lower dimensions using PCA?
The features will still have interpretability
The features will lose interpretability
The features must carry all information present in data
The features may not carry all information present in data
A. 1 and 3
B. 1 and 4
C. 2 and 3
D. 2 and 4
Ans Solution: **(D)**
When you get the features in lower dimensions then you will lose some information of data most of the times and you won't be able to interpret the lower dimension data.

9. Which of the following option(s) is / are true?
You need to initialize parameters in PCA
You don't need to initialize parameters in PCA
PCA can be trapped into local minima problem
PCA can't be trapped into local minima problem
A. 1 and 3
B. 1 and 4
C. 2 and 3
D. 2 and 4

Ans Solution: **(D)**

PCA is a deterministic algorithm which doesn't have parameters to initialize and it doesn't have local minima problem like most of the machine learning algorithms has.

10. **What is of the following statement is true about t-SNE in comparison to PCA?**

A. When the data is huge (in size), t-SNE may fail to produce better results.

B. T-NSE always produces better result regardless of the size of the data

C. PCA always performs better than t-SNE for smaller size data.

D. None of these

Ans Solution: **(A)**

Option A is correct

11. **[True or False] PCA can be used for projecting and visualizing data in lower dimensions.**

A. TRUE

B. FALSE

Solution: (A)

Sometimes it is very useful to plot the data in lower dimensions. We can take the first 2 principal components and then visualize the data using scatter plot.

12. **A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college.**

1) Which of the following statement is true in following case?

A) Feature F1 is an example of nominal variable.

B) Feature F1 is an example of ordinal variable.

C) It doesn't belong to any of the above category.

D) Both of these

Solution: (B)

Ordinal variables are the variables which has some order in their categories. For example, grade A should be consider as high grade than grade B.

13. **Which of the following is an example of a deterministic algorithm?**

A) PCA

B) K-Means

C) None of the above

Solution: (A)

A deterministic algorithm is that in which output does not change on different runs. PCA would give the same result if we run again, but not k-means.

UNIT –III

1. Which of the following methods do we use to best fit the data in Logistic Regression?

A) Least Square Error

B) Maximum Likelihood

C) Jaccard distance

D) Both A and B

Ans Solution: B

2. Choose which of the following options is true regarding One-Vs-All method in Logistic Regression.

A) We need to fit n models in n-class classification problem

B) We need to fit n-1 models to classify into n classes

C) We need to fit only 1 model to classify into n classes

D) None of these

Ans **Solution: A**

3. Suppose, You applied a Logistic Regression model on a given data and got a training accuracy X and testing accuracy Y. Now, you want to add a few new features in the same data. Select the option(s) which is/are correct in such a case.

Note: Consider remaining parameters are same.

A) Training accuracy increases

B) Training accuracy increases or remains the same

C) Testing accuracy decreases

D) Testing accuracy increases or remains the same

Ans Solution: A and D

Adding more features to model will increase the training accuracy because model has to consider more data to fit the logistic regression. But testing accuracy increases if feature is found to be significant

4. Which of the following algorithms do we use for Variable Selection?

A) LASSO

B) Ridge

C) Both

D) None of these

Ans Solution: A

In case of lasso we apply a absolute penality, after increasing the penality in lasso some of the coefficient of variables may become zero

5. Which of the following statement is true about outliers in Linear regression?

A) Linear regression is sensitive to outliers

B) Linear regression is not sensitive to outliers

C) Can't say

D) None of these

Ans Solution: (A)

The slope of the regression line will change due to outliers in most of the cases. So Linear Regression is sensitive to outliers.

6. Which of the following methods do we use to find the best fit line for data in Linear Regression?

A) Least Square Error

B) Maximum Likelihood

C) Logarithmic Loss

D) Both A and B

Ans Solution: (A)

In linear regression, we try to minimize the least square errors of the model to identify the line of best fit.

7. Which of the following is true about Residuals?

A) Lower is better

B) Higher is better

C) A or B depend on the situation

D) None of these

Ans Solution: (A)

Residuals refer to the error values of the model. Therefore lower residuals are desired.

8. Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and you found that there is a relationship between them. Which of the following conclusion do you make about this situation?

A) Since the there is a relationship means our model is not good

B) Since the there is a relationship means our model is good

C) Can't say

D) None of these

Ans Solution: (A)

There should not be any relationship between predicted values and residuals. If there exists any relationship between them, it means that the model has not perfectly captured the information in the data.

9. Suppose you have fitted a complex regression model on a dataset. Now, you are using Ridge regression with penalty x.

Choose the option which describes bias in best manner.

A) In case of very large x; bias is low

B) In case of very large x; bias is high

C) We can't say about bias

D) None of these

Ans Solution: (B)

If the penalty is very large it means model is less complex, therefore the bias would be high.

10. Which of the following option is true?

A) Linear Regression errors values has to be normally distributed but in case of Logistic Regression it is not the case

B) Logistic Regression errors values has to be normally distributed but in case of Linear Regression it is not the case

C) Both Linear Regression and Logistic Regression error values have to be normally distributed

D) Both Linear Regression and Logistic Regression error values have not to be normally distributed

Ans Solution: A

11. Suppose you have trained a logistic regression classifier and it outputs a new example x with a prediction ho(x) = 0.2. This means

Our estimate for P(y=1 | x)

Our estimate for P(y=0 | x)

Our estimate for P(y=1 | x)

Our estimate for P(y=0 | x)

Ans Solution: B

12. **True-False: Linear Regression is a supervised machine learning algorithm.**

A) TRUE

B) FALSE

Solution: (A)

Yes, Linear regression is a supervised learning algorithm because it uses true labels for training. Supervised learning algorithm should have input variable (x) and an output variable (Y) for each example.

13. **True-False: Linear Regression is mainly used for Regression.**

A) TRUE

B) FALSE

Solution: (A)

Linear Regression has dependent variables that have continuous values.

14. True-False: It is possible to design a Linear regression algorithm using a neural network?

A) TRUE
B) FALSE

Solution: (A)

True. A Neural network can be used as a universal approximator, so it can definitely implement a linear regression algorithm.

15. **Which of the following methods do we use to find the best fit line for data in Linear Regression?**
A) Least Square Error
B) Maximum Likelihood
C) Logarithmic Loss
D) Both A and B
Solution: (A)
In linear regression, we try to minimize the least square errors of the model to identify the line of best fit.

16. **Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?**
A) AUC-ROC
B) Accuracy
C) Logloss
D) Mean-Squared-Error
Solution: (D)
Since linear regression gives output as continuous values, so in such case we use mean squared error metric to evaluate the model performance. Remaining options are use in case of a classification problem.

17. **True-False: Lasso Regularization can be used for variable selection in Linear Regression.**
A) TRUE
B) FALSE
Solution: (A)
True, In case of lasso regression we apply absolute penalty which makes some of the coefficients zero.

18. **Which of the following is true about Residuals ?**
A) Lower is better
B) Higher is better

C) A or B depend on the situation

D) None of these

Solution: (A)

Residuals refer to the error values of the model. Therefore lower residuals are desired.

19. Suppose that we have N independent variables (X1,X2… Xn) and dependent variable is Y. Now Imagine that you are applying linear regression by fitting the best fit line using least square error on this data.

You found that correlation coefficient for one of it's variable(Say X1) with Y is -0.95.

Which of the following is true for X1?

A) Relation between the X1 and Y is weak

B) Relation between the X1 and Y is strong

C) Relation between the X1 and Y is neutral

D) Correlation can't judge the relationship

Solution: (B)

The absolute value of the correlation coefficient denotes the strength of the relationship. Since absolute correlation is very high it means that the relationship is strong between X1 and Y.

20. **Looking at above two characteristics, which of the following option is the correct for Pearson correlation between V1 and V2?**

If you are given the two variables V1 and V2 and they are following below two characteristics.

1. If V1 increases then V2 also increases
2. If V1 decreases then V2 behavior is unknown

A) Pearson correlation will be close to 1

B) Pearson correlation will be close to -1

C) Pearson correlation will be close to 0

D) None of these

Solution: (D)

We cannot comment on the correlation coefficient by using only statement 1. We need to consider the both of these two statements. Consider V1 as x and V2 as |x|. The correlation coefficient would not be close to 1 in such a case.

21. **Suppose Pearson correlation between V1 and V2 is zero. In such case, is it right to conclude that V1 and V2 do not have any relation between them?**

A) TRUE

B) FALSE

Solution: (B)

Pearson correlation coefficient between 2 variables might be zero even when they have a relationship between them. If the correlation coefficient is zero, it just means that that they don't move together. We can take examples like y=|x| or y=x^2.

22. **True- False: Overfitting is more likely when you have huge amount of data to train?**

A) TRUE

B) FALSE

Solution: (B)

With a small training dataset, it's easier to find a hypothesis to fit the training data exactly i.e. overfitting.

23. **We can also compute the coefficient of linear regression with the help of an analytical method called "Normal Equation". Which of the following is/are true about Normal Equation?**

1. We don't have to choose the learning rate
2. It becomes slow when number of features is very large
3. Thers is no need to iterate

A) 1 and 2

B) 1 and 3

C) 2 and 3

D) 1,2 and 3

Solution: (D)

Instead of gradient descent, Normal Equation can also be used to find coefficients.

Question Context 24-26:

Suppose you have fitted a complex regression model on a dataset. Now, you are using Ridge regression with penality x.

24. Choose the option which describes bias in best manner.

A) In case of very large x; bias is low

B) In case of very large x; bias is high

C) We can't say about bias

D) None of these

Solution: (B)

If the penalty is very large it means model is less complex, therefore the bias would be high.

25. **What will happen when you apply very large penalty?**

A) Some of the coefficient will become absolute zero

B) Some of the coefficient will approach zero but not absolute zero

C) Both A and B depending on the situation

D) None of these

Solution: (B)

In lasso some of the coefficient value become zero, but in case of Ridge, the coefficients become close to zero but not zero.

26. **What will happen when you apply very large penalty in case of Lasso?**

A) Some of the coefficient will become zero

B) Some of the coefficient will be approaching to zero but not absolute zero

C) Both A and B depending on the situation

D) None of these

Solution: (A)

As already discussed, lasso applies absolute penalty, so some of the coefficients will become zero.

27. **Which of the following statement is true about outliers in Linear regression?**

A) Linear regression is sensitive to outliers

B) Linear regression is not sensitive to outliers

C) Can't say

D) None of these

Solution: (A)

The slope of the regression line will change due to outliers in most of the cases. So Linear Regression is sensitive to outliers.

28. **Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and you found that there is a relationship between them. Which of the following conclusion do you make about this situation?**

A) Since the there is a relationship means our model is not good

B) Since the there is a relationship means our model is good

C) Can't say

D) None of these

Solution: (A)

There should not be any relationship between predicted values and residuals. If there exists any relationship between them,it means that the model has not perfectly captured the information in the data.

Question Context 29-31:

Suppose that you have a dataset D1 and you design a linear regression model of degree 3 polynomial and you found that the training and testing error is "0" or in another terms it perfectly fits the data.

29. What will happen when you fit degree 4 polynomial in linear regression?

A) There are high chances that degree 4 polynomial will over fit the data

B) There are high chances that degree 4 polynomial will under fit the data

C) Can't say

D) None of these

Solution: (A)

Since is more degree 4 will be more complex(overfit the data) than the degree 3 model so it will again perfectly fit the data. In such case training error will be zero but test error may not be zero.

30. **What will happen when you fit degree 2 polynomial in linear regression?**

A) It is high chances that degree 2 polynomial will over fit the data

B) It is high chances that degree 2 polynomial will under fit the data

C) Can't say

D) None of these

Solution: (B)

If a degree 3 polynomial fits the data perfectly, it's highly likely that a simpler model(degree 2 polynomial) might under fit the data.

31. **In terms of bias and variance. Which of the following is true when you fit degree 2 polynomial?**

A) Bias will be high, variance will be high

B) Bias will be low, variance will be high

C) Bias will be high, variance will be low

D) Bias will be low, variance will be low

Solution: (C)

Since a degree 2 polynomial will be less complex as compared to degree 3, the bias will be high and variance will be low.

Question Context 32-33:

We have been given a dataset with n records in which we have input attribute as x and output attribute as y. Suppose we use a linear regression method to model this data. To test our linear regressor, we split the data in training set and test set randomly.

32. Now we increase the training set size gradually. As the training set size increases, what do you expect will happen with the mean training error?

A) Increase

B) Decrease

C) Remain constant

D) Can't Say

Solution: (D)

Training error may increase or decrease depending on the values that are used to fit the model. If the values used to train contain more outliers gradually, then the error might just increase.

33. What do you expect will happen with bias and variance as you increase the size of training data?

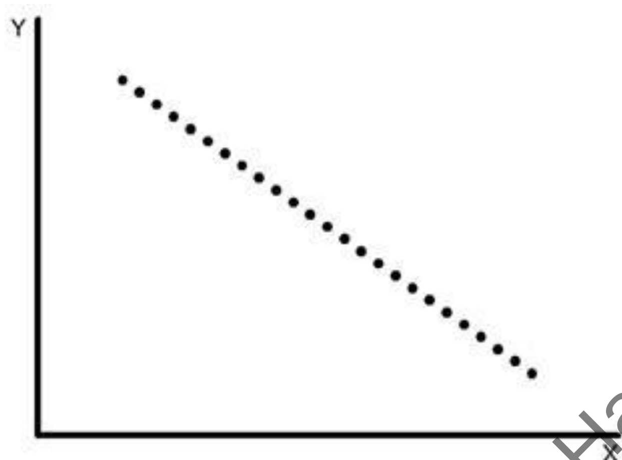
A) Bias increases and Variance increases

B) Bias decreases and Variance increases

C) Bias decreases and Variance decreases

D) Bias increases and Variance decreases
E) Can't Say False
Solution: (D)
As we increase the size of the training data, the bias would increase while the variance would decrease.

Question Context 34:
Consider the following data where one input(X) and one output(Y) is given.



34. What would be the root mean square training error for this data if you run a Linear Regression model of the form (Y = A0+A1X)?

A) Less than 0
B) Greater than zero
C) Equal to 0
D) None of these
Solution: (C)
We can perfectly fit the line on the following data so mean error will be zero.

Question Context 35-36:
Suppose you have been given the following scenario for training and validation error for Linear Regression.

Scenario	Learning Rate	Number of iterations	Training Error	Validation Error
1	0.1	1000	100	110
2	0.2	600	90	105

3	0.3	400	110	110
4	0.4	300	120	130
5	0.4	250	130	150

35. Which of the following scenario would give you the right hyper parameter?

A) 1

B) 2

C) 3

D) 4

Solution: (B)

Option B would be the better option because it leads to less training as well as validation error.

36. Suppose you got the tuned hyper parameters from the previous question. Now, Imagine you want to add a variable in variable space such that this added feature is important. Which of the following thing would you observe in such case?

A) Training Error will decrease and Validation error will increase

B) Training Error will increase and Validation error will increase

C) Training Error will increase and Validation error will decrease

D) Training Error will decrease and Validation error will decrease

E) None of the above

Solution: (D)

If the added feature is important, the training and validation error would decrease.

Question Context 37-38:

Suppose, you got a situation where you find that your linear regression model is under fitting the data.

37. In such situation which of the following options would you consider?

1. I will add more variables
2. I will start introducing polynomial degree variables
3. I will remove some variables

A) 1 and 2

B) 2 and 3

C) 1 and 3

D) 1, 2 and 3

Solution: (A)

In case of under fitting, you need to induce more variables in variable space or you can add some polynomial degree variables to make the model more complex to be able to fir the data better.

38. Now situation is same as written in previous question(under fitting).Which of following regularization algorithm would you prefer?

A) L1
B) L2
C) Any
D) None of these

Solution: (D)

I won't use any regularization methods because regularization is used in case of overfitting.

39. **True-False: Is Logistic regression a supervised machine learning algorithm?**

A) TRUE
B) FALSE

Solution: A

True, Logistic regression is a supervised learning algorithm because it uses true labels for training. Supervised learning algorithm should have input variables (x) and an target variable (Y) when you train the model .

40. **True-False: Is Logistic regression mainly used for Regression?**

A) TRUE
B) FALSE

Solution: B

Logistic regression is a classification algorithm, don't confuse with the name regression.

41. **True-False: Is it possible to design a logistic regression algorithm using a Neural Network Algorithm?**

A) TRUE
B) FALSE

Solution: A

True, Neural network is a is a *universal* approximator so it can implement linear regression algorithm.

42. **True-False: Is it possible to apply a logistic regression algorithm on a 3-class Classification problem?**

A) TRUE
B) FALSE

Solution: A

Yes, we can apply logistic regression on 3 classification problem, We can use One Vs all method for 3 class classification in logistic regression.

43. **Which of the following methods do we use to best fit the data in Logistic Regression?**

A) Least Square Error

B) Maximum Likelihood

C) Jaccard distance

D) Both A and B

Solution: B

Logistic regression uses maximum likely hood estimate for training a logistic regression.

44. **Which of the following evaluation metrics can not be applied in case of logistic regression output to compare with target?**

A) AUC-ROC

B) Accuracy

C) Logloss

D) Mean-Squared-Error

Solution: D

Since, Logistic Regression is a classification algorithm so it's output can not be real time value so mean squared error can not use for evaluating it

45. **One of the very good methods to analyze the performance of Logistic Regression is AIC, which is similar to R-Squared in Linear Regression. Which of the following is true about AIC?**

A) We prefer a model with minimum AIC value

B) We prefer a model with maximum AIC value

C) Both but depend on the situation

D) None of these

Solution: A

We select the best model in logistic regression which can least AIC.

46. **[True-False] Standardisation of features is required before training a Logistic Regression.**

A) TRUE

B) FALSE

Solution: B

Standardization isn't required for logistic regression. The main goal of standardizing features is to help convergence of the technique used for optimization.

47. Which of the following algorithms do we use for Variable Selection?

A) LASSO

B) Ridge

C) Both

D) None of these

Solution: A

In case of lasso we apply a absolute penality, after increasing the penality in lasso some of the coefficient of variables may become zero.

Context: 48-49

Consider a following model for logistic regression: P (y =1|x, w)= g(w0 + w1x)
where g(z) is the logistic function.

In the above equation the P (y =1|x; w) , viewed as a function of x, that we can get by changing the parameters w.

48 What would be the range of p in such case?

A) (0, inf)
B) (-inf, 0)
C) (0, 1)
D) (-inf, inf)

Solution: C

For values of *x* in the range of real number from $-\infty$ to $+\infty$ Logistic function will give the output between (0,1)

49 In above question what do you think which function would make p between (0,1)?

A) logistic function
B) Log likelihood function
C) Mixture of both
D) None of them

Solution: A

Explanation is same as question number 10

50. **Suppose you have been given a fair coin and you want to find out the odds of getting heads. Which of the following option is true for such a case?**

A) odds will be 0
B) odds will be 0.5
C) odds will be 1
D) None of these

Solution: C

Odds are defined as the ratio of the probability of success and the probability of failure. So in case of fair coin probability of success is 1/2 and the probability of failure is 1/2 so odd would be 1

51. **The logit function(given as l(x)) is the log of odds function. What could be the range of logit function in the domain x=[0,1]?**

A) (– ∞ , ∞)
B) (0,1)
C) (0, ∞)
D) (- ∞, 0)

Solution: A

For our purposes, the odds function has the advantage of transforming the probability function, which has values from 0 to 1, into an equivalent function with values between 0 and ∞. When we take the natural log of the odds function, we get a range of values from -∞ to ∞.

52. **Which of the following option is true?**

A) Linear Regression errors values has to be normally distributed but in case of Logistic Regression it is not the case
B) Logistic Regression errors values has to be normally distributed but in case of Linear Regression it is not the case
C) Both Linear Regression and Logistic Regression error values have to be normally distributed
D) Both Linear Regression and Logistic Regression error values have not to be normally distributed

Solution:A

53. **Which of the following is true regarding the logistic function for any value "x"?**

Note:
Logistic(x): is a logistic function of any number "x"

Logit(x): is a logit function of any number "x"

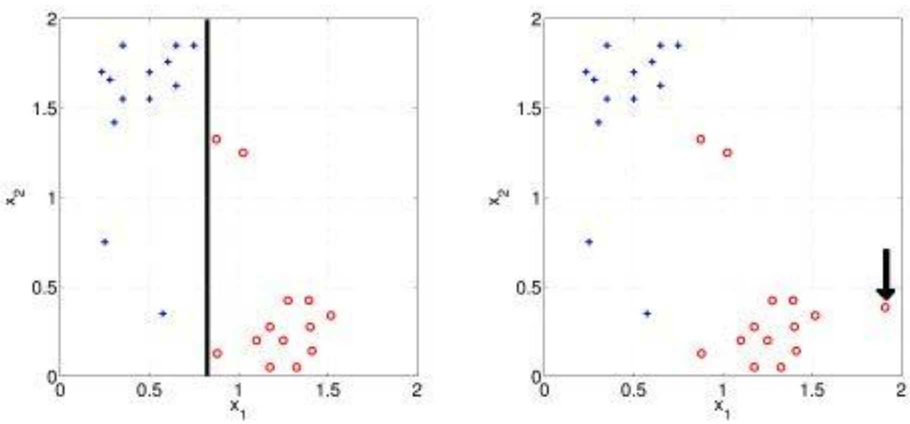
Logit_inv(x): is a inverse logit function of any number "x"

A) Logistic(x) = Logit(x)
B) Logistic(x) = Logit_inv(x)
C) Logit_inv(x) = Logit(x)
D) None of these

Solution: B

54. **How will the bias change on using high(infinite) regularisation?**

Suppose you have given the two scatter plot "a" and "b" for two classes(blue for positive and red for negative class). In scatter plot "a", you correctly classified all data points using logistic regression (black line is a decision boundary).



(a) (b)

A) Bias will be high
B) Bias will be low
C) Can't say
D) None of these

Solution: A

Model will become very simple so bias will be very high.

55. **Suppose, You applied a Logistic Regression model on a given data and got a training accuracy X and testing accuracy Y. Now, you want to add a few new features in the same data. Select the option(s) which is/are correct in such a case.**

Note: Consider remaining parameters are same.

A) Training accuracy increases
B) Training accuracy increases or remains the same
C) Testing accuracy decreases
D) Testing accuracy increases or remains the same

Solution: A and D

Adding more features to model will increase the training accuracy because model has to consider more data to fit the logistic regression. But testing accuracy increases if feature is found to be significant

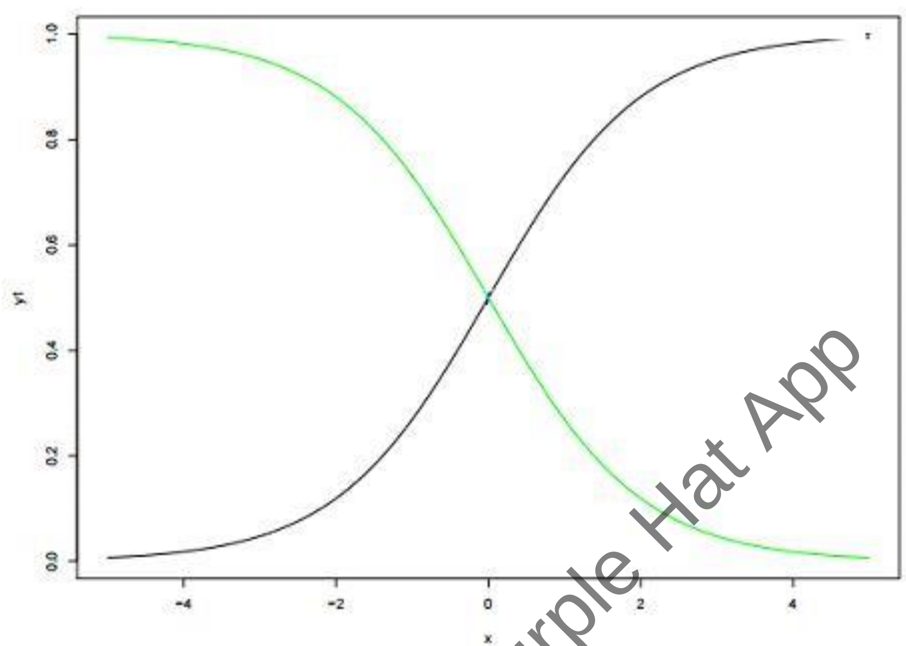
56. **Choose which of the following options is true regarding One-Vs-All method in Logistic Regression.**

A) We need to fit n models in n-class classification problem
B) We need to fit n-1 models to classify into n classes
C) We need to fit only 1 model to classify into n classes
D) None of these

Solution: A

If there are n classes, then n separate logistic regression has to fit, where the probability of each category is predicted over the rest of the categories combined.

57. **Below are two different logistic models with different values for β0 and β1.**



Which of the following statement(s) is true about β0 and β1 values of two logistics models (Green, Black)?

Note: consider Y = β0 + β1*X. Here, β0 is intercept and β1 is coefficient.

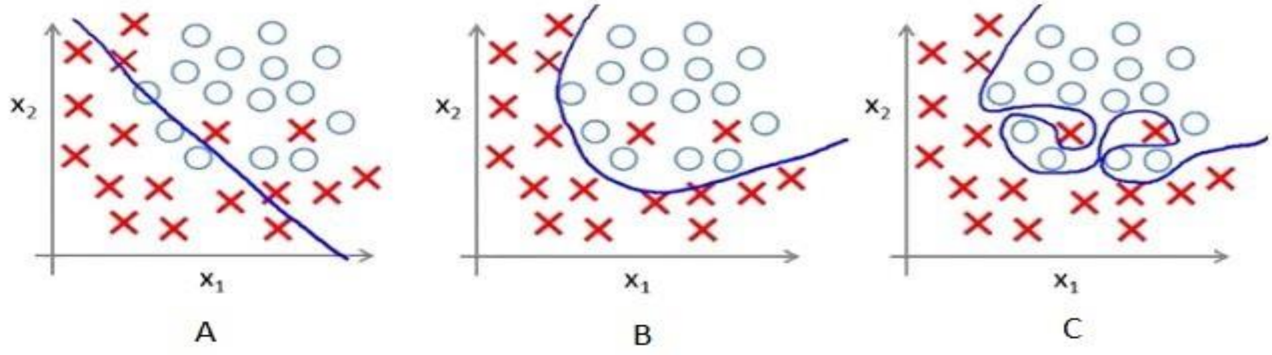
A) β1 for Green is greater than Black
B) β1 for Green is lower than Black
C) β1 for both models is same
D) Can’t Say

Solution: B

β0 and β1: β0 = 0, β1 = 1 is in X1 color(black) and β0 = 0, β1 = −1 is in X4 color (green)

Context 58-60

Below are the three scatter plot(A,B,C left to right) and hand drawn decision boundaries for logistic regression.



58. Which of the following above figure shows that the decision boundary is overfitting the training data?

A) A
B) B
C) C
D)None of these

Solution: C

Since in figure 3, Decision boundary is not smooth that means it will over-fitting the data.

59. What do you conclude after seeing this visualization?

1. The training error in first plot is maximum as compare to second and third plot.
2. The best model for this regression problem is the last (third) plot because it has minimum training error (zero).
3. The second model is more robust than first and third because it will perform best on unseen data.
4. The third model is overfitting more as compare to first and second.
5. All will perform same because we have not seen the testing data.

A) 1 and 3
B) 1 and 3
C) 1, 3 and 4
D) 5

Solution: C

The trend in the graphs looks like a quadratic trend over independent variable X. A higher degree(Right graph) polynomial might have a very high accuracy on the train population but is expected to fail badly on test dataset. But if you see in left graph we will have training error maximum because it underfits the training data

60. Suppose, above decision boundaries were generated for the different value of regularization. Which of the above decision boundary shows the maximum regularization?

A) A
B) B
C) C
D) All have equal regularization

Solution: A

Since, more regularization means more penality means less complex decision boundry that shows in first figure A.

61. **What would do if you want to train logistic regression on same data that will take less time as well as give the comparatively similar accuracy(may not be same)?**

Suppose you are using a Logistic Regression model on a huge dataset. One of the problem you may face on such huge data is that Logistic regression will take very long time to train.

A) Decrease the learning rate and decrease the number of iteration
B) Decrease the learning rate and increase the number of iteration
C) Increase the learning rate and increase the number of iteration
D) Increase the learning rate and decrease the number of iteration

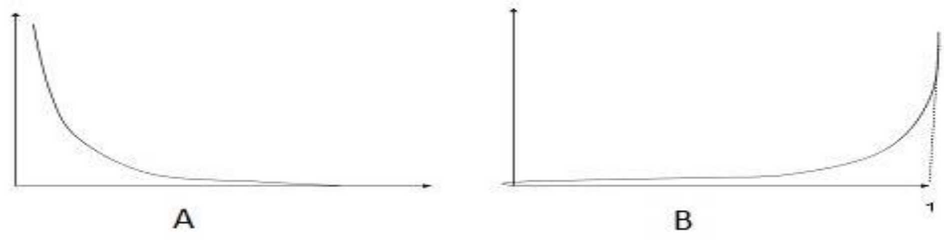
Solution: D

If you decrease the number of iteration while training it will take less time for surly but will not give the same accuracy for getting the similar accuracy but not exact you need to increase the learning rate.

62. **Which of the following image is showing the cost function for y =1.**

Following is the loss function in logistic regression(Y-axis loss function and x axis log probability) for two class classification problem.

Note: Y is the target class

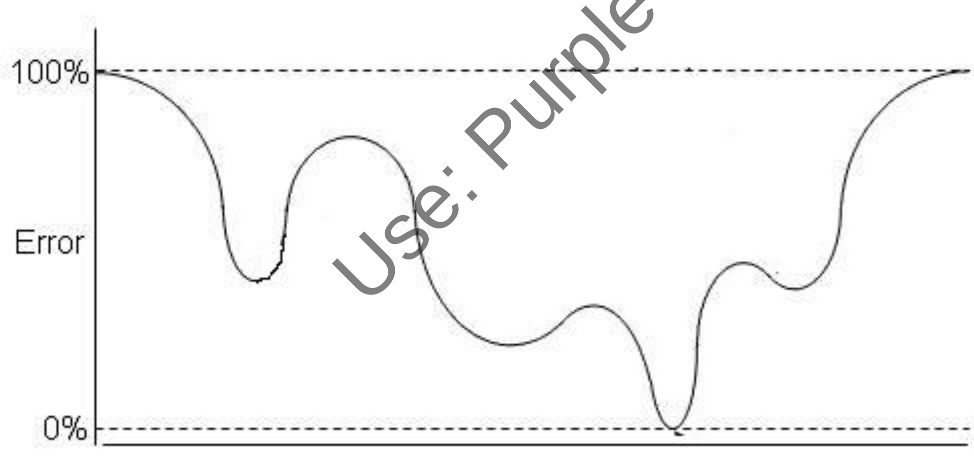


A) A
B) B
C) Both
D) None of these

Solution: A

A is the true answer as loss function decreases as the log probability increases

63. **Suppose, Following graph is a cost function for logistic regression.**



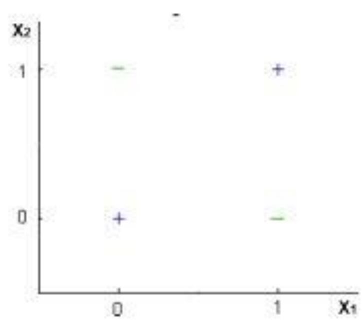
Now, How many local minimas are present in the graph?

A) 1
B) 2
C) 3
D) 4

Solution: C

There are three local minima present in the graph

64. **Can a Logistic Regression classifier do a perfect classification on the below data?**



Note: You can use only X1 and X2 variables where X1 and X2 can take only two binary values(0,1).

A) TRUE
B) FALSE
C) Can’t say
D) None of these

Solution: B

No, logistic regression only forms linear decision surface, but the examples in the figure are not linearly separable.

UNIT IV

1. The SVM's are less effective when:

A) The data is linearly separable
B) The data is clean and ready to use
C) The data is noisy and contains overlapping points

Ans Solution: C

When the data has noise and overlapping points, there is a problem in drawing a clear hyperplane without misclassifying.

2. The cost parameter in the SVM means:

A) The number of cross-validations to be made
B) The kernel to be used
C) The tradeoff between misclassification and simplicity of the model
D) None of the above

Ans Solution: C

The cost parameter decides how much an SVM should be allowed to "bend" with the data. For a low cost, you aim for a smooth decision surface and for a higher cost, you aim to classify more points correctly. It is also simply referred to as the cost of misclassification.

3. Which of the following are real world applications of the SVM?

A) Text and Hypertext Categorization
B) Image Classification
C) Clustering of News Articles
D) All of the above

Ans Solution: D

SVM's are highly versatile models that can be used for practically all real world problems ranging from regression to clustering and handwriting recognitions.

4. Which of the following is true about Naive Bayes ?

Assumes that all the features in a dataset are equally important

Assumes that all the features in a dataset are independent

Both A and B - answer

None of the above options

Ans Solution: C

5 What do you mean by generalization error in terms of the SVM?

A) How far the hyperplane is from the support vectors
B) How accurately the SVM can predict outcomes for unseen data
C) The threshold amount of error in an SVM

Ans Solution: B

Generalisation error in statistics is generally the out-of-sample error which is the measure of how accurately a model can predict values for previously unseen data.

6 The SVM’s are less effective when:

A) The data is linearly separable
B) The data is clean and ready to use
C) The data is noisy and contains overlapping points

Ans Solution: C

When the data has noise and overlapping points, there is a problem in drawing a clear hyperplane without misclassifying.

7 What is/are true about kernel in SVM?

1. Kernel function map low dimensional data to high dimensional space
2. It’s a similarity function

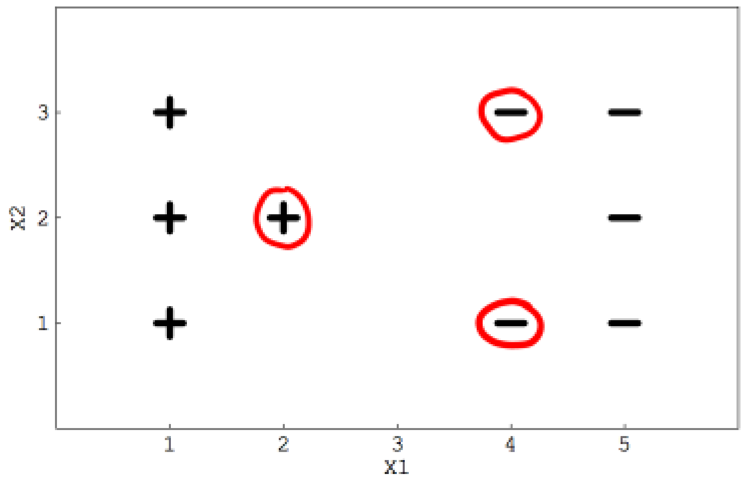
A) 1
B) 2
C) 1 and 2
D) None of these

Ans Solution: C

Both the given statements are correct.

Question Context:8– 9

Suppose you are using a Linear SVM classifier with 2 class classification problem. Now you have been given the following data in which some points are circled red that are representing support vectors.



8. If you remove the following any one red points from the data. Does the decision boundary will change?

A) Yes
B) No

Solution: A

These three examples are positioned such that removing any one of them introduces slack in the constraints. So the decision boundary would completely change.

9. [True or False] If you remove the non-red circled points from the data, the decision boundary will change?

A) True
B) False

Solution: B

On the other hand, rest of the points in the data won't affect the decision boundary much.

10. What do you mean by generalization error in terms of the SVM?

A) How far the hyperplane is from the support vectors
B) How accurately the SVM can predict outcomes for unseen data
C) The threshold amount of error in an SVM

Solution: B

Generalization error in statistics is generally the out-of-sample error which is the measure of how accurately a model can predict values for previously unseen data.

11. When the C parameter is set to infinite, which of the following holds true?

A) The optimal hyperplane if exists, will be the one that completely separates the data
B) The soft-margin classifier will separate the data
C) None of the above

Solution: A

At such a high level of misclassification penalty, soft margin will not hold existence as there will be no room for error.

12. What do you mean by a hard margin?

A) The SVM allows very low error in classification
B) The SVM allows high amount of error in classification
C) None of the above

Solution: A

A hard margin means that an SVM is very rigid in classification and tries to work extremely well in the training set, causing overfitting.

13. The minimum time complexity for training an SVM is O(n2). According to this fact, what sizes of datasets are not best suited for SVM's?

A) Large datasets
B) Small datasets
C) Medium sized datasets
D) Size does not matter

Solution: A

Datasets which have a clear classification boundary will function best with SVM's.

14. The effectiveness of an SVM depends upon:

A) Selection of Kernel
B) Kernel Parameters
C) Soft Margin Parameter C
D) All of the above

Solution: D

The SVM effectiveness depends upon how you choose the basic 3 requirements mentioned above in such a way that it maximises your efficiency, reduces error and overfitting.

15. upport vectors are the data points that lie closest to the decision surface.

A) TRUE
B) FALSE

Solution: A

They are the points closest to the hyperplane and the hardest ones to classify. They also have a direct bearing on the location of the decision surface.

16. The SVM's are less effective when:

A) The data is linearly separable
B) The data is clean and ready to use
C) The data is noisy and contains overlapping points

Solution: C

When the data has noise and overlapping points, there is a problem in drawing a clear hyperplane without misclassifying.

17. Suppose you are using RBF kernel in SVM with high Gamma value. What does this signify?

A) The model would consider even far away points from hyperplane for modeling
B) The model would consider only the points close to the hyperplane for modeling
C) The model would not be affected by distance of points from hyperplane for modeling
D) None of the above

Solution: B

The gamma parameter in SVM tuning signifies the influence of points either near or far away from the hyperplane.

For a low gamma, the model will be too constrained and include all points of the training dataset, without really capturing the shape.

For a higher gamma, the model will capture the shape of the dataset well.

18. The cost parameter in the SVM means:

A) The number of cross-validations to be made
B) The kernel to be used
C) The tradeoff between misclassification and simplicity of the model
D) None of the above

Solution: C

The cost parameter decides how much an SVM should be allowed to “bend” with the data. For a low cost, you aim for a smooth decision surface and for a higher cost, you aim to classify more points correctly. It is also simply referred to as the cost of misclassification.

19. Suppose you are building a SVM model on data X. The data X can be error prone which means that you should not trust any specific data point too much. Now think that you want to build a SVM model which has quadratic kernel function of polynomial degree 2 that uses Slack variable C as one of it’s hyper parameter. Based upon that give the answer for following question.

What would happen when you use very large value of C(C->infinity)?

Note: For small C was also classifying all data points correctly

A) We can still classify data correctly for given setting of hyper parameter C
B) We can not classify data correctly for given setting of hyper parameter C
C) Can’t Say
D) None of these

Solution: A

For large values of C, the penalty for misclassifying points is very high, so the decision boundary will perfectly separate the data if possible.

20. What would happen when you use very small C (C~0)?

A) Misclassification would happen
B) Data will be correctly classified
C) Can’t say
D) None of these

Solution: A

The classifier can maximize the margin between most of the points, while misclassifying a few points, because the penalty is so low.

21. If I am using all features of my dataset and I achieve 100% accuracy on my training set, but ~70% on validation set, what should I look out for?

A) Underfitting
B) Nothing, the model is perfect
C) Overfitting

Solution: C

If we’re achieving 100% training accuracy very easily, we need to check to verify if we’re overfitting our data.

22. **Which of the following are real world applications of the SVM?**

A) Text and Hypertext Categorization
B) Image Classification
C) Clustering of News Articles
D) All of the above

Solution: D

SVM's are highly versatile models that can be used for practically all real world problems ranging from regression to clustering and handwriting recognitions.

Question Context: 23 – 25

Suppose you have trained an SVM with linear decision boundary after training SVM, you correctly infer that your SVM model is under fitting.

23. Which of the following option would you more likely to consider iterating SVM next time?

A) You want to increase your data points
B) You want to decrease your data points
C) You will try to calculate more variables
D) You will try to reduce the features

Solution: C

The best option here would be to create more features for the model.

24. Suppose you gave the correct answer in previous question. What do you think that is actually happening?

1. We are lowering the bias
2. We are lowering the variance
3. We are increasing the bias
4. We are increasing the variance

A) 1 and 2
B) 2 and 3
C) 1 and 4
D) 2 and 4

Solution: C

Better model will lower the bias and increase the variance

25. In above question suppose you want to change one of it's(SVM) hyperparameter so that effect would be same as previous questions i.e model will not under fit?

A) We will increase the parameter C
B) We will decrease the parameter C
C) Changing in C don't effect
D) None of these

Solution: A

Increasing C parameter would be the right thing to do here, as it will ensure regularized model

26. **We usually use feature normalization before using the Gaussian kernel in SVM. What is true about feature normalization?**

1. We do feature normalization so that new feature will dominate other
2. Some times, feature normalization is not feasible in case of categorical variables
3. Feature normalization always helps when we use Gaussian kernel in SVM

A) 1
B) 1 and 2
C) 1 and 3
D) 2 and 3

Solution: B

Statements one and two are correct.

Question Context: 27-29

Suppose you are dealing with 4 class classification problem and you want to train a SVM model on the data for that you are using One-vs-all method. Now answer the below questions?

27. How many times we need to train our SVM model in such case?

A) 1
B) 2
C) 3
D) 4

Solution: D

For a 4 class problem, you would have to train the SVM at least 4 times if you are using a one-vs-all method.

28. Suppose you have same distribution of classes in the data. Now, say for training 1 time in one vs all setting the SVM is taking 10 second. How many seconds would it require to train one-vs-all method end to end?

A) 20
B) 40
C) 60
D) 80

Solution: B

It would take 10×4 = 40 seconds

29 Suppose your problem has changed now. Now, data has only 2 classes. What would you think how many times we need to train SVM in such case?

A) 1
B) 2
C) 3
D) 4

Solution: A

Training the SVM only one time would give you appropriate results

Question context: 30 –31

Suppose you are using SVM with linear kernel of polynomial degree 2, Now think that you have applied this on data and found that it perfectly fit the data that means, Training and testing accuracy is 100%.

30. Now, think that you increase the complexity (or degree of polynomial of this kernel). What would you think will happen?

A) Increasing the complexity will over fit the data
B) Increasing the complexity will under fit the data
C) Nothing will happen since your model was already 100% accurate
D) None of these

Solution: A

Increasing the complexity of the data would make the algorithm overfit the data.

31. In the previous question after increasing the complexity you found that training accuracy was still 100%. According to you what is the reason behind that?

1. Since data is fixed and we are fitting more polynomial term or parameters so the algorithm starts memorizing everything in the data
2. Since data is fixed and SVM doesn't need to search in big hypothesis space

A) 1
B) 2
C) 1 and 2
D) None of these

Solution: C

Both the given statements are correct.

32. **What is/are true about kernel in SVM?**

1. Kernel function map low dimensional data to high dimensional space
2. It's a similarity function

A) 1
B) 2
C) 1 and 2
D) None of these

Solution: C

Both the given statements are correct.

UNIT V

1. **Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?**

a) Decision Tree
b) Regression
c) Classification
d) Random Forest

Ans D

2. Which of the following is a disadvantage of decision trees?

a) Factor analysis
b) Decision trees are robust to outliers
c) Decision trees are prone to be overfit

d) None of the above

Ans C

3. Can decision trees be used for performing clustering?

a. True
b. False

Ans Solution: (A)

Decision trees can also be used to for clusters in the data but clustering often generates natural clusters and is not dependent on any objective function.

4. Which of the following algorithm is most sensitive to outliers?

a. K-means clustering algorithm
b. K-medians clustering algorithm
c. K-modes clustering algorithm
d. K-medoids clustering algorithm

Ans Solution: (A)

5 **Sentiment Analysis is an example of:**

Regression

Classification

Clustering

Reinforcement Learning

Options:

a. 1 Only
b. 1 and 2
c. 1 and 3
d. 1, 2 and 4

Ans D

6 Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:

Capping and flouring of variables

Removal of outliers

Options:

a. 1 only
b. 2 only
c. 1 and 2
d. None of the above

Ans A

7 **Which of the following is/are true about bagging trees?**

1. In bagging trees, individual trees are independent of each other
2. Bagging is the method for improving the performance by aggregating the results of weak learners

A) 1
B) 2
C) 1 and 2
D) None of these

Ans **Solution: C**

Both options are true. In Bagging, each individual trees are independent of each other because they consider different subset of features and samples.

8. **Which of the following is/are true about boosting trees?**

1. In boosting trees, individual weak learners are independent of each other
2. It is the method for improving the performance by aggregating the results of weak learners

A) 1
B) 2
C) 1 and 2
D) None of **these**

Ans **Solution: B**

In boosting tree individual weak learners are not independent of each other because each tree correct the results of previous tree. Bagging and boosting both can be consider as improving the base learners results.

9. In Random forest you can generate hundreds of trees (say T1, T2 …..Tn) and then aggregate the results of these tree. Which of the following is true about individual (Tk) tree in Random Forest?

1. Individual tree is built on a subset of the features
2. Individual tree is built on all the features
3. Individual tree is built on a subset of observations
4. Individual tree is built on full set of observations

A) 1 and 3
B) 1 and 4
C) 2 and 3
D) 2 and 4

Ans Solution: A

Random forest is based on bagging concept, that consider faction of sample and faction of feature for building the individual trees.

10. Suppose you are using a bagging based algorithm say a RandomForest in model building. Which of the following can be true?

1. Number of tree should be as large as possible
2. You will have interpretability after using Random Forest

A) 1
B) 2
C) 1 and 2
D) None of these

Ans Solution: A

Since Random Forest aggregate the result of different weak learners, If It is possible we would want more number of trees in model building. Random Forest is a black box model you will lose interpretability after using it.

11. Which of the following is/are true about Random Forest and Gradient Boosting ensemble methods?

1. Both methods can be used for classification task
2. Random Forest is use for classification whereas Gradient Boosting is use for regression task
3. Random Forest is use for regression whereas Gradient Boosting is use for Classification task
4. Both methods can be used for regression task

A) 1
B) 2
C) 3
D) 4
E) 1 and 4

Solution: E

Both algorithms are design for classification as well as regression task.

12. In Random forest you can generate hundreds of trees (say T1, T2 …..Tn) and then aggregate the results of these tree. Which of the following is true about individual(Tk) tree in Random Forest?

1. Individual tree is built on a subset of the features
2. Individual tree is built on all the features
3. Individual tree is built on a subset of observations
4. Individual tree is built on full set of observations

A) 1 and 3
B) 1 and 4
C) 2 and 3
D) 2 and 4

Solution: A

Random forest is based on bagging concept, that consider faction of sample and faction of feature for building the individual trees.

13. **Which of the following algorithm doesn't uses learning Rate as of one of its hyperparameter?**

1. Gradient Boosting
2. Extra Trees
3. AdaBoost
4. Random Forest

A) 1 and 3
B) 1 and 4
C) 2 and 3
D) 2 and 4

Solution: D

Random Forest and Extra Trees don’t have learning rate as a hyperparameter.

14. **Which of the following algorithm are not an example of ensemble learning algorithm?**

A) Random Forest
B) Adaboost
C) Extra Trees
D) Gradient Boosting
E) Decision Trees

Solution: E

Decision trees doesn’t aggregate the results of multiple trees so it is not an ensemble algorithm.

15. **Suppose you are using a bagging based algorithm say a RandomForest in model building. Which of the following can be true?**

1. Number of tree should be as large as possible
2. You will have interpretability after using RandomForest

A) 1
B) 2
C) 1 and 2
D) None of these

Solution: A

Since Random Forest aggregate the result of different weak learners, If It is possible we would want more number of trees in model building. Random Forest is a black box model you will lose interpretability after using it.

16. **True-False: The bagging is suitable for high variance low bias models?**

A) TRUE
B) FALSE

Solution: A

The bagging is suitable for high variance low bias models or you can say for complex models.

17. **To apply bagging to regression trees which of the following is/are true in such case?**

1. We build the N regression with N bootstrap sample
2. We take the average the of N regression tree
3. Each tree has a high variance with low bias

A) 1 and 2
B) 2 and 3
C) 1 and 3
D) 1,2 and 3

Solution: D

All of the options are correct and self-explanatory

18. **How to select best hyper parameters in tree based models?**

A) Measure performance over training data
B) Measure performance over validation data
C) Both of these
D) None of these

Solution: B

We always consider the validation results to compare with the test result.

19. **In which of the following scenario a gain ratio is preferred over Information Gain?**

A) When a categorical variable has very large number of category
B) When a categorical variable has very small number of category
C) Number of categories is the not the reason
D) None of these

Solution: A

When high cardinality problems, gain ratio is preferred over Information Gain technique.

20. **Suppose you have given the following scenario for training and validation error for Gradient Boosting. Which of the following hyper parameter would you choose in such case?**

Scenario	**Depth**	**Training Error**	**Validation Error**
1	2	100	110
2	4	90	105
3	6	50	100
4	8	45	105

5	10	30	150

A) 1
B) 2
C) 3
D) 4

Solution: B

Scenario 2 and 4 has same validation accuracies but we would select 2 because depth is lower is better hyper parameter.

21. **Which of the following is/are not true about DBSCAN clustering algorithm:**

    1. **For data points to be in a cluster, they must be in a distance threshold to a core point**
    2. **It has strong assumptions for the distribution of data points in** dataspace
    3. **It has substantially high time complexity of order $O(n^3)$**
    4. **It does not require prior knowledge of the no. of desired clusters**
    5. **It is robust to outliers**

Options:

A. 1 only

B. 2 only

C. 4 only

D. 2 and 3

Solution: D

- DBSCAN can form a cluster of any arbitrary shape and does not have strong assumptions for the distribution of data points in the data space.
- DBSCAN has a low time complexity of order O (n log n) only.

22. Point out the correct statement.
a) The choice of an appropriate metric will influence the shape of the clusters
b) Hierarchical clustering is also called HCA
c) In general, the merges and splits are determined in a greedy manner
d) All of the mentioned

Answer: d

Explanation: Some elements may be close to one another according to one distance and farther away according to another.

23. Which of the following is required by K-means clustering?

a) defined distance metric

b) number of clusters

c) initial guess as to cluster centroids

d) all of the mentioned

Answer: d

Explanation: K-means clustering follows partitioning approach.

24. Point out the wrong statement.

a) k-means clustering is a method of vector quantization

b) k-means clustering aims to partition n observations into k clusters

c) k-nearest neighbor is same as k-means

d) none of the mentioned

Answer: c

Explanation: k-nearest neighbour has nothing to do with k-means.

25. Which of the following function is used for k-means clustering?

a) k-means

b) k-mean

c) heat map

d) none of the mentioned

Answer: a

Explanation: K-means requires a number of clusters.

26. K-means is not deterministic and it also consists of number of iterations.

a) True

b) False

Answer: a

Explanation: K-means clustering produces the final estimate of cluster centroids.

27.

Unit IV Naïve Bayes and Support Vector Machine

1. How many terms are required for building a bayes model?

a) 1

b) 2

c) 3

d) 4

Answer: c

2. What is needed to make probabilistic systems feasible in the world?

a) Reliability

b) Crucial robustness

c) Feasibility

d) None of the mentioned

Answer: b

3. Where does the bayes rule can be used?

a) Solving queries

b) Increasing complexity

c) Decreasing complexity

d) Answering probabilistic query

Answer: d

4. What does the bayesian network provides?

a) Complete description of the domain

b) Partial description of the domain

c) Complete description of the problem

d) None of the mentioned

Answer: a

5. How the entries in the full joint probability distribution can be calculated?

a) Using variables

b) Using information

c) Both Using variables & information

d) None of the mentioned

Answer: b

6. How the bayesian network can be used to answer any query?

a) Full distribution

b) Joint distribution

c) Partial distribution

d) All of the mentioned

Answer: b

7. How the compactness of the bayesian network can be described?

a) Locally structured

b) Fully structured

c) Partial structure

d) All of the mentioned

Answer: a

8. To which does the local structure is associated?

a) Hybrid

b) Dependant

c) Linear

d) None of the mentioned

Answer: c

9. Which condition is used to influence a variable directly by all the others?

a) Partially connected

b) Fully connected

c) Local connected

d) None of the mentioned

Answer: b

10. What is the consequence between a node and its predecessors while creating bayesian network?

a) Functionally dependent

b) Dependant

c) Conditionally independent

d) Both Conditionally dependant & Dependant

Answer: c

11.What do you mean by generalization error in terms of the SVM?

A) How far the hyperplane is from the support vectors

B) How accurately the SVM can predict outcomes for unseen data

C) The threshold amount of error in an SVM

Solution: B

12. When the C parameter is set to infinite, which of the following holds true?

A) The optimal hyperplane if exists, will be the one that completely separates the data

B) The soft-margin classifier will separate the data

C) None of the above

Solution: A

13.What do you mean by a hard margin?

A) The SVM allows very low error in classification

B) The SVM allows high amount of error in classification

C) None of the above

Solution: A

14.The minimum time complexity for training an SVM is O(n2). According to this fact, what sizes of datasets are not best suited for SVM's?

A) Large datasets

B) Small datasets

C) Medium sized datasets

D) Size does not matter

Solution: A

15. The effectiveness of an SVM depends upon:

A) Selection of Kernel

B) Kernel Parameters

C) Soft Margin Parameter C

D) All of the above

Solution: D

16. Support vectors are the data points that lie closest to the decision surface.

A) TRUE

B) FALSE

Solution: A

17. The SVM's are less effective when:

A) The data is linearly separable

B) The data is clean and ready to use

C) The data is noisy and contains overlapping points

Solution: C

18. Suppose you are using RBF kernel in SVM with high Gamma value. What does this signify?

A) The model would consider even far away points from hyperplane for modeling

B) The model would consider only the points close to the hyperplane for modeling

C) The model would not be affected by distance of points from hyperplane for modeling

D) None of the above

Solution: B

19. The cost parameter in the SVM means:

A) The number of cross-validations to be made

B) The kernel to be used

C) The tradeoff between misclassification and simplicity of the model

D) None of the above

Solution: C

20.Suppose you are building a SVM model on data X. The data X can be error prone which means that you should not trust any specific data point too much. Now think that you want to build a SVM model which has quadratic kernel function of polynomial degree 2 that uses Slack variable C as one of it's hyper parameter. Based upon that give the answer for following question.

What would happen when you use very large value of C(C->infinity)?

A) We can still classify data correctly for given setting of hyper parameter C

B) We can not classify data correctly for given setting of hyper parameter C

C) Can't Say

D) None of these

Solution: A

21. What would happen when you use very small C (C~0)?

A) Misclassification would happen

B) Data will be correctly classified

C) Can't say

D) None of these

Solution: A

22. If I am using all features of my dataset and I achieve 100% accuracy on my training set, but ~70% on validation set, what should I look out for?

A) Underfitting

B) Nothing, the model is perfect

C) Overfitting

Solution: C

23. Which of the following are real world applications of the SVM?

A) Text and Hypertext Categorization

B) Image Classification

C) Clustering of News Articles

D) All of the above

Solution: D

24.Which of the following option would you more likely to consider iterating SVM next time?

A) You want to increase your data points

B) You want to decrease your data points

C) You will try to calculate more variables

D) You will try to reduce the features

Solution: C

26. We usually use feature normalization before using the Gaussian kernel in SVM. What is true about feature normalization?

1. We do feature normalization so that new feature will dominate other

2. Some times, feature normalization is not feasible in case of categorical variables

3. Feature normalization always helps when we use Gaussian kernel in SVM

A) 1

B) 1 and 2

C) 1 and 3

D) 2 and 3

Solution: B

Question context: 27 – 28

Suppose you are using SVM with linear kernel of polynomial degree 2, Now think that you have applied this on data and found that it perfectly fit the data that means, Training and testing accuracy is 100%.

27) Now, think that you increase the complexity(or degree of polynomial of this kernel). What would you think will happen?

A) Increasing the complexity will overfit the data

B) Increasing the complexity will underfit the data

C) Nothing will happen since your model was already 100% accurate

D) None of these

Solution: A

28) In the previous question after increasing the complexity you found that training accuracy was still 100%. According to you what is the reason behind that?

1. Since data is fixed and we are fitting more polynomial term or parameters so the algorithm starts memorizing everything in the data

2. Since data is fixed and SVM doesn't need to search in big hypothesis space

A) 1

B) 2

C) 1 and 2

D) None of these

Solution: C

29. What is/are true about kernel in SVM?

1. Kernel function map low dimensional data to high dimensional space

2. It's a similarity function

A) 1

B) 2

C) 1 and 2

D) None of these

Solution: C

UNIT V

Decision Trees and Ensemble Learning

1. Predicting with trees evaluate _____________ within each group of data.

a) equality

b) homogeneity

c) heterogeneity

d) all of the mentioned

Answer: b

2. Point out the wrong statement.

a) Training and testing data must be processed in different way

b) Test transformation would mostly be imperfect

c) The first goal is statistical and second is data compression in PCA

d) All of the mentioned

Answer: a

3. Which of the following method options is provided by train function for bagging?

a) bagEarth

b) treebag

c) bagFDA

d) all of the mentioned

Answer: d

4. Which of the following is correct with respect to random forest?

a) Random forest are difficult to interpret but often very accurate

b) Random forest are easy to interpret but often very accurate

c) Random forest are difficult to interpret but very less accurate

d) None of the mentioned

Answer: a

5. Point out the correct statement.

a) Prediction with regression is easy to implement

b) Prediction with regression is easy to interpret

c) Prediction with regression performs well when linear model is correct

Answer: d

6. Which of the following library is used for boosting generalized additive models?

a) gamBoost

b) gbm

c) ada

d) all of the mentioned

Answer: a

7. The principal components are equal to left singular values if you first scale the variables.

a) True

b) False

Answer: b

8. Which of the following is statistical boosting based on additive logistic regression?

a) gamBoost

b) gbm

c) ada

d) mboost

Answer: a

9. Which of the following is one of the largest boost subclass in boosting?

a) variance boosting

b) gradient boosting

c) mean boosting

d) all of the mentioned

Answer: b

10. PCA is most useful for non linear type models.

a) True

b) False

Answer: b

11.varImp is a wrapper around the evimp function in the _______ package.

a) numpy

b) earth

c) plot

d) none of the mentioned

Answer: b

12. Point out the wrong statement.

a) The trapezoidal rule is used to compute the area under the ROC curve

b) For regression, the relationship between each predictor and the outcome is evaluated

c) An argument, para, is used to pick the model fitting technique

d) All of the mentioned

Answer: c

13. Which of the following curve analysis is conducted on each predictor for classification?

a) NOC

b) ROC

c) COC

d) All of the mentioned

Answer: b

14. Which of the following function tracks the changes in model statistics?

a) varImp

b) varImpTrack

c) findTrack

d) none of the mentioned

Answer: a

15. Point out the correct statement.

a) The difference between the class centroids and the overall centroid is used to measure the variable influence

b) The Bagged Trees output contains variable usage statistics

c) Boosted Trees uses different approach as a single tree

d) None of the mentioned

Answer: a

16. The advantage of using a model-based approach is that is more closely tied to the model performance.

a) True

b) False

Answer: a

17. Which of the following model sums the importance over each boosting iteration?

a) Boosted trees

b) Bagged trees

c) Partial least squares

d) None of the mentioned

Answer: a

18. Which of the following argument is used to set importance values?

a) scale

b) set

c) value

d) all of the mentioned

Answer: a

19. For most classification models, each predictor will have a separate variable importance for each class.

a) True

b) False

Answer: a

20. A _________ is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility.

a) Decision tree

b) Graphs

c) Trees

d) Neural Networks

Answer: a

21. Decision Tree is a display of an algorithm.

a) True

b) False

Answer: a

22. What is Decision Tree?

a) Flow-Chart

b) Structure in which internal node represents test on an attribute, each branch represents outcome of test and each leaf node represents class label

c) Flow-Chart & Structure in which internal node represents test on an attribute, each branch represents outcome of test and each leaf node represents class label

d) None of the mentioned

Answer: c

23. Decision Trees can be used for Classification Tasks.

a) True

b) False

View Answer

Answer: a

24. Choose from the following that are Decision Tree nodes?

a) Decision Nodes

b) End Nodes

c) Chance Nodes

d) All of the mentioned

Answer: d

25. Decision Nodes are represented by ____________

a) Disks

b) Squares

c) Circles

d) Triangles

Answer: b

26. Chance Nodes are represented by __________

a) Disks

b) Squares

c) Circles

d) Triangles

Answer: c

27. End Nodes are represented by __________

a) Disks

b) Squares

c) Circles

d) Triangles

Answer: d

28. Which of the following are the advantage/s of Decision Trees?

a) Possible Scenarios can be added

b) Use a white box model, If given result is provided by a model

c) Worst, best and expected values can be determined for different scenarios

d) All of the mentioned

Answer: d

29.Which of the following algorithm is not an example of an ensemble method?

A. Extra Tree Regressor

B. Random Forest

C. Gradient Boosting

D. Decision Tree

Solution: (D)

30. What is true about an ensembled classifier?

1. Classifiers that are more “sure” can vote with more conviction

2. Classifiers can be more “sure” about a particular part of the space

3. Most of the times, it performs better than a single classifier

A. 1 and 2

B. 1 and 3

C. 2 and 3

D. All of the above

Solution: (D)

31. Which of the following option is / are correct regarding benefits of ensemble model?

1. Better performance

2. Generalized models

3. Better interpretability

A. 1 and 3

B. 2 and 3

C. 1 and 2

D. 1, 2 and 3

Solution: (C)

32.Which of the following can be true for selecting base learners for an ensemble?

1. Different learners can come from same algorithm with different hyper parameters

2. Different learners can come from different algorithms

3. Different learners can come from different training spaces

A. 1

B. 2

C. 1 and 3

D. 1, 2 and 3

Solution: (D)

33. True or False: Ensemble learning can only be applied to supervised learning methods.

A. True

B. False

Solution: (B)

34. True or False: Ensembles will yield bad results when there is significant diversity among the models.

A. True

B. False

Solution: (B)

35.Which of the following is / are true about weak learners used in ensemble model?

1. They have low variance and they don’t usually overfit

2. They have high bias, so they can not solve hard learning problems

3. They have high variance and they don’t usually overfit

A. 1 and 2

B. 1 and 3

C. 2 and 3

D. None of these

Solution: (A)

36.True or False: Ensemble of classifiers may or may not be more accurate than any of its individual model.

A. True

B. False

Solution: (A)

37.parameters of all base models to improve the ensemble performance?

A. Yes

B. No

C. can’t say

Solution: (B)

38. Generally, an ensemble method works better, if the individual base models have ____________?

A. Less correlation among predictions

B. High correlation among predictions

C. Correlation does not have any impact on ensemble output

D. None of the above

Solution: (A)

39. Which of the following ensemble method works similar to above-discussed election procedure?

A. Bagging

B. Boosting

C. A Or B

D. None of these

Solution: (A)

40. Suppose you are given ‘n’ predictions on test data by ‘n’ different models (M1, M2, …. Mn) respectively. Which of the following method(s) can be used to combine the predictions of these models?

1. Median

2. Product

3. Average

4. Weighted sum

5. Minimum and Maximum

6. Generalized mean rule

A. 1, 3 and 4

B. 1,3 and 6

C. 1,3, 4 and 6

D. All of above

Solution: (D)

41. If you want to ensemble these models using majority voting method. What will be the maximum accuracy you can get?

A. 100%

B. 78.38 %

C. 44%

D. 70

Solution: (A)

42. If you want to ensemble these models using majority voting. What will be the minimum accuracy you can get?

A. Always greater than 70%

B. Always greater than and equal to 70%

C. It can be less than 70%

D. None of these

Solution: (C)

43. How can we assign the weights to output of different models in an ensemble?

1. Use an algorithm to return the optimal weights

2. Choose the weights using cross validation

3. Give high weights to more accurate models

A. 1 and 2

B. 1 and 3

C. 2 and 3

D. All of above

Solution: (D)

44. Which of the following is true about averaging ensemble?

A. It can only be used in classification problem

B. It can only be used in regression problem

C. It can be used in both classification as well as regression

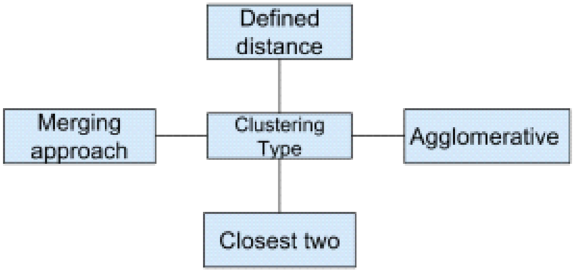
D. None of these

Solution: (C)

UNIT VI

Clustering Techniques

1. Which of the following clustering type has characteristic shown in the below figure?



a) Partitional
b) Hierarchical
c) Naive bayes
d) None of the mentioned
Answer: b

2. Point out the correct statement.
a) The choice of an appropriate metric will influence the shape of the clusters
b) Hierarchical clustering is also called HCA
c) In general, the merges and splits are determined in a greedy manner
d) All of the mentioned
Answer: d

3. Which of the following is finally produced by Hierarchical Clustering?
a) final estimate of cluster centroids
b) tree showing how close things are to each other
c) assignment of each point to clusters
d) all of the mentioned
Answer: b

4. Which of the following is required by K-means clustering?
a) defined distance metric
b) number of clusters
c) initial guess as to cluster centroids
d) all of the mentioned
Answer: d

5. Point out the wrong statement.
a) k-means clustering is a method of vector quantization
b) k-means clustering aims to partition n observations into k clusters
c) k-nearest neighbor is same as k-means
d) none of the mentioned
Answer: c

6. Hierarchical clustering should be primarily used for exploration.
a) True
b) False
Answer: a

8. Which of the following function is used for k-means clustering?
a) k-means
b) k-mean
c) heatmap
d) none of the mentioned
Answer: a

9. Which of the following clustering requires merging approach?
a) Partitional
b) Hierarchical
c) Naive Bayes
d) None of the mentioned
Answer: b

10. K-means is not deterministic and it also consists of number of iterations.
a) True
b) False
Answer: a

This sheet is for 1 Mark questions							
S.r No	**Question**	**Image**	**a**	**b**	**c**	**d**	**Correct Answer**
	Write down question	img.jpg	Option a	Option b	Option c	Option d	a/b/c/d
1	In reinforcement learning if feedback is negative one it is defined as____.		Penalty	Overlearning	Reward	None of above	A
2	According to____ , it's a key success factor for the survival and evolution of all species.		Claude Shannon's theory	Gini Index	Darwin's theory	None of above	C
3	How can you avoid overfitting ?		By using a lot of data	By using inductive machine learning	By using validation only	None of above	A
4	What are the popular algorithms of Machine Learning?		Decision Trees and Neural Networks (back propagation)	Probabilistic networks and Nearest Neighbor	Support vector machines	All	D
5	What is 'Training set'?		Training set is used to test the accuracy of the hypotheses generated by the learner.	A set of data is used to discover the potentially predictive	Both A & B	None of above	B

				relationship.			
6	Common deep learning applications include____		Image classification, Real-time visual tracking	Autonomous car driving, Logistic optimization	Bioinformatics, Speech recognition	All above	D
7	what is the function of ‘Supervised Learning’?		Classifications, Predict time series, Annotate strings	Speech recognition, Regression	Both A & B	None of above	C
8	Commons unsupervised applications include		Object segmentation	Similarity detection	Automatic labeling	All above	D
9	Reinforcement learning is particularly efficient when______________.		the environment is not completely deterministic	it's often very dynamic	it's impossible to have a precise error measure	All above	D
10	if there is only a discrete number of possible outcomes (called categories), the process becomes a______.		Regression	Classification.	Modelfree	Categories	B
11	Which of the following are supervised learning applications		Spam detection, Pattern detection, Natural Language Processing	Image classification, Real-time visual tracking	Autonomous car driving, Logistic optimization	Bioinformatics, Speech recognition	A
12	During the last few years, many ______ algorithms have been applied to deep neural networks to learn the best policy		Logical	Classical	Classification	None of above	D

	for playing Atari video games and to teach an agent how to associate the right action with an input representing the state.						
13	Which of the following sentence is correct?		Machine learning relates with the study, design and development of the algorithms that give computers the capability to learn without being explicitly programmed.	Data mining can be defined as the process in which the unstructured data tries to extract knowledge or unknown interesting patterns.	Both A & B	None of the above	C
14	What is ‘Overfitting’ in Machine learning?		when a statistical model describes random error or noise instead of underlying relationship ‘overfitting’ occurs.	Robots are programed so that they can perform the task based on data they gather	While involving the process of learning ‘overfitting’ occurs.	a set of data is used to discover the potentially predictive relationship	A

				from sensors.			
15	What is 'Test set'?		Test set is used to test the accuracy of the hypotheses generated by the learner.	It is a set of data is used to discover the potentially predictive relationship.	Both A & B	None of above	A
16	________is much more difficult because it's necessary to determine a supervised strategy to train a model for each feature and, finally, to predict their value		Removing the whole line	Creating sub-model to predict those features	Using an automatic strategy to input them according to the other known values	All above	B
17	How it's possible to use a different placeholder through the parameter_______.		regression	classification	random_state	missing_values	D
18	If you need a more powerful scaling feature, with a superior control on outliers and the possibility to select a quantile range, there's also the class________.		RobustScaler	DictVectorizer	LabelBinarizer	FeatureHasher	A
19	scikit-learn also provides a class for per-sample normalization, Normalizer. It can apply________to each element of a dataset		max, l0 and l1 norms	max, l1 and l2 norms	max, l2 and l3 norms	max, l3 and l4 norms	B
20	There are also many univariate methods that can be used in order to select the best features according to specific criteria based on________.		F-tests and p-values	chi-square	ANOVA	All above	A
21	Which of the following selects only a subset of features belonging to a certain percentile		SelectPercentile	FeatureHasher	SelectKBest	All above	A

22	________performs a PCA with non-linearly separable data sets.		SparsePCA	KernelPCA	SVD	None of the Mentioned	B
23	A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college. Which of the following statement is true in following case?		Feature F1 is an example of nominal variable.	Feature F1 is an example of ordinal variable.	It doesn't belong to any of the above category.	Both of these	B
24	What would you do in PCA to get the same projection as SVD?		Transform data to zero mean	Transform data to zero median	Not possible	None of these	A
25	What is PCA, KPCA and ICA used for?		Principal Components Analysis	Kernel based Principal Component Analysis	Independent Component Analysis	All above	D
26	Can a model trained for item based similarity also choose from a given set of items?		YES	NO			A
27	What are common feature selection methods in regression task?		correlation coefficient	Greedy algorithms	All above	None of these	C
28	The parameter______ allows specifying the percentage of elements to put into the test/training set		test_size	training_size	All above	None of these	C
29	In many classification problems, the target ______ is made up of categorical labels which cannot immediately be processed by any algorithm.		random_state	dataset	test_size	All above	B
30	_______adopts a dictionary-oriented approach, associating to each category label a progressive integer number.		LabelEncoder class	LabelBinarizer class	DictVectorizer	Feature Hasher	A
31	If Linear regression model perfectly first i.e., train error is zero, then ______________________		a) Test error is also	b) Test error	c) Couldn't commen	d) Test error is equal to	c

			always zero	is non zero	t on Test error	Train error	
32	Which of the following metrics can be used for evaluating regression models?i) R Squaredii) Adjusted R Squarediii) F Statisticsiv) RMSE / MSE / MAE		a) ii and iv	b) i and ii	c) ii, iii and iv	d) i, ii, iii and iv	d
33	How many coefficients do you need to estimate in a simple linear regression model (One independent variable)?		a) 1	b) 2	c) 3	d) 4	b
34	In a simple linear regression model (One independent variable), If we change the input variable by 1 unit. How much output variable will change?		a) by 1	b) no change	c) by intercept	d) by its slope	d
35	Function used for linear regression in R is ___________		a) lm(formula, data)	b) lr(formula, data)	c) lrm(formula, data)	d) regression.linear (formula, data)	a
36	In syntax of linear model lm(formula,data,..), data refers to ______		a) Matrix	b) Vector	c) Array	d) List	b
37	In the mathematical Equation of Linear Regression Y = β1 + β2X + ϵ, (β1, β2) refers to ___________		a) (X-intercept, Slope)	b) (Slope, X-Intercept)	c) (Y-Intercept, Slope)	d) (slope, Y-Intercept)	c
38	Linear Regression is a supervised machine learning algorithm.		A) TRUE	B) FALSE			a
39	It is possible to design a Linear regression algorithm using a neural network?		A) TRUE	B) FALSE			a
40	Which of the following methods do we use to find the best fit line for data in Linear Regression?		A) Least Square Error	B) Maximum Likelihood	C) Logarithmic Loss	D) Both A and B	a
41	Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?		A) AUC-ROC	B) Accuracy	C) Logloss	D) Mean-Squared-Error	d
42	Which of the following is true about Residuals ?		A) Lower is better	B) Higher is better	C) A or B depend on the situation	D) None of these	a
43	Overfitting is more likely when you have huge amount of data to train?		A) TRUE	B) FALSE			b
44	Which of the following statement is true about outliers in Linear regression?		A) Linear regression is	B) Linear regression	C) Can't say	D) None of these	a

			sensitive to outliers	is not sensitive to outliers			
45	Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and you found that there is a relationship between them. Which of the following conclusion do you make about this situation?		A) Since the there is a relationship means our model is not good	B) Since the there is a relationship means our model is good	C) Can't say	D) None of these	a
46	Naive Bayes classifiers are a collection ------------------of algorithms		Classification	Clustering	Regression	All	a
47	Naive Bayes classifiers is ________________ Learning		Supervised	Unsupervised	Both	None	a
48	Features being classified is independent of each other in Naïve Bayes Classifier		False	TRUE			b
49	Features being classified is ___________ of each other in Naïve Bayes Classifier		Independent	Dependent	Partial Dependent	None	a
50	Bayes Theorem is given by where 1. P(H) is the probability of hypothesis H being true. 2. P(E) is the probability of the evidence(regardless of the hypothesis). 3. P(E|H) is the probability of the evidence given that hypothesis is true. 4. P(H|E) is the probability of the hypothesis given that the evidence is there.	bayes.jpg	True	FALSE			a
51	In given image, P(H|E) is___________probability.	bayes.jpg	Posterior	Prior			a
52	In given image, P(H) is___________probability.	bayes.jpg	Posterior	Prior			b
53	Conditional probability is a measure of the probability of an event given that another event has already occurred.		True	FALSE			a
54	Bayes' theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event.		True	FALSE			a

55	Bernoulli Naïve Bayes Classifier is ___________distribution		Continuous	Discrete	Binary		c
56	Multinomial Naïve Bayes Classifier is ___________distribution		Continuous	Discrete	Binary		b
57	Gaussian Naïve Bayes Classifier is ___________distribution		Continuous	Discrete	Binary		a
58	Binarize parameter in BernoulliNB scikit sets threshold for binarizing of sample features.		True	FALSE			a
59	Gaussian distribution when plotted, gives a bell shaped curve which is symmetric about the _______ of the feature values.		Mean	Variance	Discrete	Random	a
60	SVMs directly give us the posterior probabilities P(y = 1jx) and P(y = □1jx)		True	FALSE			b
61	Any linear combination of the components of a multivariate Gaussian is a univariate Gaussian.		True	FALSE			a
62	Solving a non linear separation problem with a hard margin Kernelized SVM (Gaussian RBF Kernel) might lead to overfitting		True	FALSE			a
63	SVM is a ----------------- algorithm		Classification	Clustering	Regression	All	a
64	SVM is a ----------------- learning		Supervised	Unsupervised	Both	None	a
65	The linear **SVM** classifier works by drawing a straight line between two classes		True	FALSE			a
66	Which of the following function provides unsupervised prediction ?	--	cl_forecastB	cl_nowcastC	cl_precastD	None of the Mentioned	D
67	Which of the following is characteristic of best machine learning method ?	--	fast	accuracy	scalable	All above	D
68	What are the different Algorithm techniques in Machine Learning?	--	Supervised Learning and Semi-supervised Learning	Unsupervised Learning and Transduction	Both A & B	None of the Mentioned	C

69	What is the standard approach to supervised learning?	--	split the set of example into the training set and the test	group the set of example into the training set and the test	a set of observed instances tries to induce a general rule	learns programs from data	A
70	Which of the following is not Machine Learning?	--	Artificial Intelligence	Rule based inference	Both A & B	None of the Mentioned	B
71	What is Model Selection in Machine Learning?	--	The process of selecting models among different mathematical models, which are used to describe the same data set	when a statistical model describes random error or noise instead of underlying relationship	Find interesting directions in data and find novel observations/ database cleaning	All above	A
72	Which are two techniques of Machine Learning ?	--	Genetic Programming and Inductive Learning	Speech recognition and Regression	Both A & B	None of the Mentioned	A
73	Even if there are no actual supervisors _________ learning is also based on feedback provided by the environment	--	Supervised	Reinforcement	Unsupervised	None of the above	B

74	What does learning exactly mean?	--	Robots are programed so that they can perform the task based on data they gather from sensors.	A set of data is used to discover the potentially predictive relationship.	Learning is the ability to change according to external stimuli and remembering most of all previous experiences.	It is a set of data is used to discover the potentially predictive relationship.	C
75	When it is necessary to allow the model to develop a generalization ability and avoid a common problem called______.	--	Overfitting	Overlearning	Classification	Regression	A
76	Techniques involve the usage of both labeled and unlabeled data is called___.	--	Supervised	Semi-supervised	Unsupervised	None of the above	B
77	In reinforcement learning if feedback is negative one it is defined as____.	--	Penalty	Overlearning	Reward	None of above	A
78	According to____ , it's a key success factor for the survival and evolution of all species.	--	Claude Shannon's theory	Gini Index	Darwin's theory	None of above	C
79	A supervised scenario is characterized by the concept of a _____.	--	Programmer	Teacher	Author	Farmer	B
80	overlearning causes due to an excessive ______.	--	Capacity	Regression	Reinforcement	Accuracy	A
81	Which of the following is an example of a deterministic algorithm?	--	PCA	K-Means	None of the above		A
82	Which of the following model model include a backwards elimination feature selection routine?	--	MCV	MARS	MCRS	All above	B
83	Can we extract knowledge without apply feature selection	--	YES	NO			A
84	While using feature selection on the data, is the number of features decreases.	--	NO	YES			B
85	Which of the following are several models for feature extraction	--	regression	classification	None of the above		C

86	_____ provides some built-in datasets that can be used for testing purposes.	--	scikit-learn	classification	regression	None of the above	A
87	While using _____ all labels are turned into sequential numbers.	--	LabelEncoder class	LabelBinarizer class	DictVectorizer	FeatureHasher	A
88	_______produce sparse matrices of real numbers that can be fed into any machine learning model.	--	DictVectorizer	FeatureHasher	Both A & B	None of the Mentioned	C
89	scikit-learn offers the class______, which is responsible for filling the holes using a strategy based on the mean, median, or frequency	--	LabelEncoder	LabelBinarizer	DictVectorizer	Imputer	D
90	Which of the following scale data by removing elements that don't belong to a given range or by considering a maximum absolute value.	--	MinMaxScaler	MaxAbsScaler	Both A & B	None of the Mentioned	C
91	scikit-learn also provides a class for per-sample normalization,_____	--	Normalizer	Imputer	Classifier	All above	A
92	______dataset with many features contains information proportional to the independence of all features and their variance.	--	normalized	unnormalized	Both A & B	None of the Mentioned	B
93	In order to assess how much information is brought by each component, and the correlation among them, a useful tool is the_____.	--	Concuttent matrix	Convergance matrix	Supportive matrix	Covariance matrix	D
94	The_____ parameter can assume different values which determine how the data matrix is initially processed.	--	run	start	init	stop	C
95	______allows exploiting the natural sparsity of data while extracting principal components.	--	SparsePCA	KernelPCA	SVD	init parameter	A
96	Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?	--	AUC-ROC	Accuracy	Logloss	Mean-Squared-Error	D
97	Which of the following is true about Residuals ?	--	Lower is better	Higher is better	A or B depend on the situation	None of these	A
98	Overfitting is more likely when you have huge amount of data to train?	--	TRUE	FALSE			B

99	Which of the following statement is true about outliers in Linear regression?	--	Linear regression is sensitive to outliers	Linear regression is not sensitive to outliers	Can’t say	None of these	A
100	Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and you found that there is a relationship between them. Which of the following conclusion do you make about this situation?	--	Since the there is a relationship means our model is not good	Since the there is a relationship means our model is good	Can’t say	None of these	A
101	Let’s say, a “Linear regression” model perfectly fits the training data (train error is zero). Now, Which of the following statement is true?	--	You will always have test error zero	You can not have test error zero	None of the above		C
102	In a linear regression problem, we are using “R-squared” to measure goodness-of-fit. We add a feature in linear regression model and retrain the same model.Which of the following option is true?	--	If R Squared increases, this variable is significant.	If R Squared decreases, this variable is not significant.	Individually R squared cannot tell about variable importance. We can’t say anything about it right now.	None of these.	C
103	Which of the one is true about Heteroskedasticity?	--	Linear Regression with varying error terms	Linear Regression with constant error terms	Linear Regression with zero error terms	None of these	A

104	Which of the following assumptions do we make while deriving linear regression parameters?1. The true relationship between dependent y and predictor x is linear2. The model errors are statistically independent3. The errors are normally distributed with a 0 mean and constant standard deviation4. The predictor x is non-stochastic and is measured error-free	--	1,2 and 3.	1,3 and 4.	1 and 3.	All of above.	D
105	To test linear relationship of y(dependent) and x(independent) continuous variables, which of the following plot best suited?	--	Scatter plot	Barchart	Histograms	None of these	A
106	which of the following step / assumption in regression modeling impacts the trade-off between under-fitting and over-fitting the most.	--	The polynomial degree	Whether we learn the weights by matrix inversion or gradient descent	The use of a constant-term		A
107	Can we calculate the skewness of variables based on mean and median?	--	TRUE	FALSE			B
108	Which of the following is true about "Ridge" or "Lasso" regression methods in case of feature selection?	--	Ridge regression uses subset selection of features	Lasso regression uses subset selection of features	Both use subset selection of features	None of above	B
109	Which of the following statement(s) can be true post adding a variable in a linear regression model?1. R-Squared and Adjusted R-squared both increase2. R-Squared increases and Adjusted R-squared decreases3. R-	--	1 and 2	1 and 3	2 and 4	None of the above	A

	Squared decreases and Adjusted R-squared decreases4. R-Squared decreases and Adjusted R-squared increases						
110	How many coefficients do you need to estimate in a simple linear regression model (One independent variable)?	--	1	2	Can't Say		B
111	In given image, P(H) is___________probability.	bayes.jpg	Posterior	Prior			B
112	Conditional probability is a measure of the probability of an event given that another event has already occurred.	--	True	FALSE			A
113	Gaussian distribution when plotted, gives a bell shaped curve which is symmetric about the _________ of the feature values.	--	Mean	Variance	Discrete	Random	A
114	SVMs directly give us the posterior probabilities P(y = 1jx) and P(y = 1jx)	--	True	FALSE			B
115	SVM is a ----------------- algorithm	--	Classification	Clustering	Regression	All	A
116	What is/are true about kernel in SVM?1. Kernel function map low dimensional data to high dimensional space2. It's a similarity function	--	1	2	1 and 2	None of these	C
117	Suppose you are building a SVM model on data X. The data X can be error prone which means that you should not trust any specific data point too much. Now think that you want to build a SVM model which has quadratic kernel function of polynomial degree 2 that uses Slack variable C as one of it's hyper parameter.What would happen when you use very small C (C~0)?	--	Misclassification would happen	Data will be correctly classified	Can't say	None of these	A
118	The cost parameter in the SVM means:	--	The number of cross-validations to be made	The kernel to be used	The tradeoff between misclassification and simplicity of the model	None of the above	C
119	Bayes' theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event.	--	True	FALSE			A

120	Bernoulli Naïve Bayes Classifier is ____________distribution	--	Continuous	Discrete	Binary		C
121	If you remove the non-red circled points from the data, the decision boundary will change?	svm.jpg	TRUE	FALSE			B
122	How do you handle missing or corrupted data in a dataset?	--	a. Drop missing rows or columns	b. Replace missing values with mean/median/mode	c. Assign a unique category to missing values	d. All of the above	D
123	Binarize parameter in BernoulliNB scikit sets threshold for binarizing of sample features.	--	True	FALSE			A
124	Which of the following statements about Naive Bayes is incorrect?	--	A. Attributes are equally important.	B. Attributes are statistically dependent of one another given the class value.	C. Attributes are statistically independent of one another given the class value.	D. Attributes can be nominal or numeric	B
125	The SVM's are less effective when:	--	The data is linearly separable	The data is clean and ready to use	The data is noisy and contains overlapping points		C
126	Naive Bayes classifiers is _________________ Learning	--	Supervised	Unsupervised	Both	None	A
127	Features being classified is independent of each other in Naïve Bayes Classifier	--	False	TRUE			B

128	Features being classified is __________ of each other in Naïve Bayes Classifier	--	Independent	Dependent	Partial Dependent	None	A
129	Bayes Theorem is given by where 1. P(H) is the probability of hypothesis H being true. 2. P(E) is the probability of the evidence(regardless of the hypothesis). 3. P(E|H) is the probability of the evidence given that hypothesis is true. 4. P(H|E) is the probability of the hypothesis given that the evidence is there.	bayes.jpg	True	FALSE			A
130	Any linear combination of the components of a multivariate Gaussian is a univariate Gaussian.	--	True	FALSE			A

This sheet is for 2 Mark questions							
S.r No	**Question**	**Image**	**a**	**b**	**c**	**d**	**Correct Answer**
e.g 1	Write down question	img.jpg	Option a	Option b	Option c	Option d	a/b/c/d
1	A supervised scenario is characterized by the concept of a _____.		Programmer	Teacher	Author	Farmer	B
2	overlearning causes due to an excessive ______.		Capacity	Regression	Reinforcement	Accuracy	A
3	If there is only a discrete number of		Modelfree	Categories	Prediction	None of above	B

	possible outcomes called ______.						
4	What is the standard approach to supervised learning?		split the set of example into the training set and the test	group the set of example into the training set and the test	a set of observed instances tries to induce a general rule	learns programs from data	A
5	Some people are using the term ___ instead of prediction only to avoid the weird idea that machine learning is a sort of modern magic.		Inferenc e	Interfere nce	Accuracy	None of above	A
6	The term ______ can be freely used, but with the same meaning adopted in physics or system theory.		Accuracy	Cluster	Regression	Prediction	D
7	Which are two techniques of Machine Learning ?		Genetic Program ming and Inductive Learning	Speech recogniti on and Regressio n	Both A & B	None of the Mentioned	A
8	Even if there are no actual supervisors _________ learning is also based on feedback provided by		Supervis ed	Reinforce ment	Unsupervis ed	None of the above	B

	the environment						
9	Common deep learning applications / problems can also be solved using_____		Real-time visual object identification	Classic approaches	Automatic labeling	Bio-inspired adaptive systems	B
10	Identify the various approaches for machine learning.		Concept Vs Classification Learning	Symbolic Vs Statistical Learning	Inductive Vs Analytical Learning	All above	D
11	what is the function of 'Unsupervised Learning'?		Find clusters of the data and find low-dimensional representations of the data	Find interesting directions in data and find novel observations/ database cleaning	Interesting coordinates and correlations	All	D
12	What are the two methods used for the calibration in Supervised Learning?		Platt Calibration and Isotonic Regression	Statistics and Informal Retrieval			A
13	What is the standard approach to supervised learning?		split the set of example into the training set and the test	group the set of example into the training set and the test	a set of observed instances tries to induce a general rule	learns programs from data	A
14	Which of the following is not Machine Learning?		Artificial Intelligence	Rule based inference	Both A & B	None of the Mentioned	B
15	What is Model Selection in		The process of	when a statistical model	Find interesting directions	All above	A

	Machine Learning?		selecting models among different mathematical models, which are used to describe the same data set	describes random error or noise instead of underlying relationship	in data and find novel observations/ database cleaning		
16	_____ provides some built-in datasets that can be used for testing purposes.		scikit-learn	classification	regression	None of the above	A
17	While using _____ all labels are turned into sequential numbers.		LabelEncoder class	LabelBinarizer class	DictVectorizer	FeatureHasher	A
18	_______produce sparse matrices of real numbers that can be fed into any machine learning model.		DictVectorizer	FeatureHasher	Both A & B	None of the Mentioned	C
19	scikit-learn offers the class______, which is responsible for filling the holes using a strategy based on the mean, median, or frequency		LabelEncoder	LabelBinarizer	DictVectorizer	Imputer	D

20	Which of the following scale data by removing elements that don't belong to a given range or by considering a maximum absolute value.		MinMax Scaler	MaxAbsS caler	Both A & B	None of the Mentioned	C
21	Which of the following model model include a backwards elimination feature selection routine?		MCV	MARS	MCRS	All above	B
22	Can we extract knowledge without apply feature selection		YES	NO			A
23	While using feature selection on the data, is the number of features decreases.		NO	YES			B
24	Which of the following are several models for feature extraction		regressio n	classificat ion	None of the above		C
25	scikit-learn also provides a class for per-sample		Normaliz er	Imputer	Classifier	All above	A

	normalization,______						
26	_______data set with many features contains information proportional to the independence of all features and their variance.		normalized	unnormalized	Both A & B	None of the Mentioned	B
27	In order to assess how much information is brought by each component, and the correlation among them, a useful tool is the______.		Concuttent matrix	Convergance matrix	Supportive matrix	Covariance matrix	D
28	The______ parameter can assume different values which determine how the data matrix is initially processed.		run	start	init	stop	C
29	_______allows exploiting the natural sparsity of data while extracting principal components.		SparsePCA	KernelPCA	SVD	init parameter	A
30	Which of the		PCA	K-Means	None of the above		A

	following is an example of a deterministic algorithm?						
31	Let’s say, a “Linear regression” model perfectly fits the training data (train error is zero). Now, Which of the following statement is true?		A. You will always have test error zero	B. You can not have test error zero	C. None of the above		c
32	In a linear regression problem, we are using “R-squared” to measure goodness-of-fit. We add a feature in linear regression model and retrain the same model.Which of the following option is true?		A. If R Squared increases, this variable is significant.	B. If R Squared decreases, this variable is not significant.	C. Individually R squared cannot tell about variable importance. We can’t say anything about it right now.	D. None of these.	c
33	Which of the one is true about Heteroskedasticity?		A. Linear Regression with varying error terms	B. Linear Regression with constant error terms	C. Linear Regression with zero error terms	D. None of these	a
34	Which of the following assumptions do we make		A. 1,2 and 3.	B. 1,3 and 4.	C. 1 and 3.	D. All of above.	d

	while deriving linear regression parameters ?1. The true relationship between dependent y and predictor x is linear2. The model errors are statistically independent3. The errors are normally distributed with a 0 mean and constant standard deviation4. The predictor x is non-stochastic and is measured error-free						
35	To test linear relationship of y(dependent) and x(independent) continuous variables, which of the following plot best suited?		A. Scatter plot	B. Barchart	C. Histograms	D. None of these	a
36	Generally, which of the following method(s) is used for		A. 1 and 2	B. only 1	C. only 2	D. None of these.	b

	predicting continuous dependent variable?1. Linear Regression2. Logistic Regression						
37	Suppose you are training a linear regression model. Now consider these points.1. Overfitting is more likely if we have less data2. Overfitting is more likely when the hypothesis space is small.Which of the above statement(s) are correct?		A. Both are False	B. 1 is False and 2 is True	C. 1 is True and 2 is False	D. Both are True	c
38	Suppose we fit "Lasso Regression" to a data set, which has 100 features (X1,X2…X100). Now, we rescale one of these feature by multiplying with 10 (say that feature is X1), and then refit Lasso		A. It is more likely for X1 to be excluded from the model	B. It is more likely for X1 to be included in the model	C. Can't say	D. None of these	b

	regression with the same regularization parameter. Now, which of the following option will be correct?						
39	Which of the following is true about “Ridge” or “Lasso” regression methods in case of feature selection?		A. Ridge regression uses subset selection of features	B. Lasso regression uses subset selection of features	C. Both use subset selection of features	D. None of above	b
40	Which of the following statement(s) can be true post adding a variable in a linear regression model?1. R-Squared and Adjusted R-squared both increase2. R-Squared increases and Adjusted R-squared decreases3. R-Squared decreases and Adjusted R-squared decreases4. R-Squared		A. 1 and 2	B. 1 and 3	C. 2 and 4	D. None of the above	a

	decreases and Adjusted R-squared increases						
41	We can also compute the coefficient of linear regression with the help of an analytical method called "Normal Equation". Which of the following is/are true about "Normal Equation"?1. We don't have to choose the learning rate2. It becomes slow when number of features is very large3. No need to iterate		A. 1 and 2	B. 1 and 3.	C. 2 and 3.	D. 1,2 and 3.	d
42	How many coefficients do you need to estimate in a simple linear regression model (One independent variable)?		A. 1	B. 2	C. Can't Say		b
43	If two variables are correlated,		A. Yes	B. No			b

	is it necessary that they have a linear relationship ?						
44	Correlated variables can have zero correlation coeffficient. True or False?		A. True	B. False			a
45	Which of the following option is true regarding “Regression ” and “Correlation ” ?Note: y is dependent variable and x is independen t variable.		A. The relations hip is symmetr ic between x and y in both.	B. The relations hip is not symmetri c between x and y in both.	C. The relationshi p is not symmetric between x and y in case of correlation but in case of regression it is symmetric.	D. The relationshi p is symmetric between x and y in case of correlation but in case of regression it is not symmetric.	d
46	What is/are true about kernel in SVM?1. Kernel function map low dimensional data to high dimensional space2. It’s a similarity function		1	2	1 and 2	None of these	c
47	Suppose you are building a SVM model on data X. The data X can be error prone which		Misclassi fication would happen	Data will be correctly classified	Can’t say	None of these	a

	means that you should not trust any specific data point too much. Now think that you want to build a SVM model which has quadratic kernel function of polynomial degree 2 that uses Slack variable C as one of it's hyper parameter. What would happen when you use very small C (C~0)?						
48	Suppose you are using a Linear SVM classifier with 2 class classification problem. Now you have been given the following data in which some points are circled red that are representing support vectors.If you remove the	svm.jpg	yes	no			a

	following any one red points from the data. Does the decision boundary will change?						
49	If you remove the non-red circled points from the data, the decision boundary will change?	svm.jpg	TRUE	FALSE			b
50	When the C parameter is set to infinite, which of the following holds true?		The optimal hyperplane if exists, will be the one that completely separates the data	The soft-margin classifier will separate the data	None of the above		a
51	Suppose you are building a SVM model on data X. The data X can be error prone which means that you should not trust any specific data point too much. Now think that you want to build a SVM model which has quadratic		We can still classify data correctly for given setting of hyper parameter C	We can not classify data correctly for given setting of hyper parameter C	Can't Say	None of these	a

	kernel function of polynomial degree 2 that uses Slack variable C as one of it’s hyper parameter. What would happen when you use very large value of C(C->infinity)?						
52	SVM can solve linear and non-linear probl ems		TRUE	FALSE			a
53	The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.		TRUE	FALSE			a
54	Hyperplanes are _______________boundar ies that help classify the data points.		usual	decision	parallel		b
55	The ______of the hyperplane depends		dimensio n	classificat ion	reduction		a

	upon the number of features.						
56	Hyperplanes are decision boundaries that help classify the data points.		TRUE	FALSE			a
57	SVM algorit hms use a set of mathematic al functions that are defined as the kernel.		TRUE	FALSE			a
58	In SVM, Kernel function is used to map a lower dimensional data into a higher dimensional data.		TRUE	FALSE			a
59	*In SVR we try to fit the error within a certain threshold.*		TRUE	FALSE			a
60	When the C parameter is set to infinite, which of the following holds true?		The optimal hyperpla ne if exists, will be the one that complet ely separate s the data	The soft-margin classifier will separate the data	None of the above		a
61	How do you handle missing or corrupted data in a dataset?		a. Drop missing rows or columns	b. Replace missing values with mean/m	c. Assign a unique category to missing values	d. All of the above	d

				edian/mode			
62	What is the purpose of performing cross-validation?		a. To assess the predictive performance of the models	b. To judge how the trained model performs outside the sample on test data	c. Both A and B		c
63	Which of the following is true about Naive Bayes ?		a. Assumes that all the features in a dataset are equally important	b. Assumes that all the features in a dataset are independent	c. Both A and B	d. None of the above option	c
64	Which of the following statements about Naive Bayes is incorrect?		A. Attributes are equally important.	B. Attributes are statistically dependent of one another given the class value.	C. Attributes are statistically independent of one another given the class value.	D. Attributes can be nominal or numeric	b
65	Which of the following is **not** supervised learning?		PCA	Decision Tree	Naive Bayesian	Linerar regression	a
66	How can you avoid overfitting ?	--	By using a lot of data	By using inductive machine learning	By using validation only	None of above	A
67	What are the popular algorithms of Machine Learning?	--	Decision Trees and Neural Networks (back	Probabilistic networks and Nearest Neighbor	Support vector machines	All	D

			propagation)				
68	What is 'Training set'?	--	Training set is used to test the accuracy of the hypotheses generated by the learner.	A set of data is used to discover the potentially predictive relationship.	Both A & B	None of above	B
69	Identify the various approaches for machine learning.	--	Concept Vs Classification Learning	Symbolic Vs Statistical Learning	Inductive Vs Analytical Learning	All above	D
70	what is the function of 'Unsupervised Learning'?	--	Find clusters of the data and find low-dimensional representations of the data	Find interesting directions in data and find novel observations/ database cleaning	Interesting coordinates and correlations	All	D
71	What are the two methods used for the calibration in Supervised Learning?	--	Platt Calibration and Isotonic Regression	Statistics and Informal Retrieval			A
72	_______can be adopted when it's necessary to categorize a large amount of data with a few complete examples or when there's the	--	Supervised	Semi-supervised	Reinforcement	Clusters	B

	need to impose some constraints to a clustering algorithm.						
73	In reinforcement learning, this feedback is usually called as___.	--	Overfitting	Overlearning	Reward	None of above	C
74	In the last decade, many researchers started training bigger and bigger models, built with several different layers that's why this approach is called_____.	--	Deep learning	Machine learning	Reinforcement learning	Unsupervised learning	A
75	there's a growing interest in pattern recognition and associative memories whose structure and functioning are similar to what happens in the neocortex. Such an approach also allows	--	Regression	Accuracy	Modelfree	Scalable	C

	simpler algorithms called ______						
76	_______ showed better performance than other approaches, even without a context-based model	--	Machine learning	Deep learning	Reinforcement learning	Supervised learning	B
77	Common deep learning applications / problems can also be solved using_____	--	Real-time visual object identification	Classic approaches	Automatic labeling	Bio-inspired adaptive systems	B
78	Some people are using the term ____ instead of prediction only to avoid the weird idea that machine learning is a sort of modern magic.	--	Inference	Interference	Accuracy	None of above	A
79	The term ______ can be freely used, but with the same meaning adopted in physics or system theory.	--	Accuracy	Cluster	Regression	Prediction	D
80	If there is only a discrete	--	Modelfree	Categories	Prediction	None of above	B

	number of possible outcomes called ______.						
81	A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college. Which of the following statement is true in following case?	--	Feature F1 is an example of nominal variable.	Feature F1 is an example of ordinal variable.	It doesn't belong to any of the above category.	Both of these	B
82	What would you do in PCA to get the same projection as SVD?	--	Transform data to zero mean	Transform data to zero median	Not possible	None of these	A
83	What is PCA, KPCA and ICA used for?	--	Principal Components Analysis	Kernel based Principal Component Analysis	Independent Component Analysis	All above	D
84	Can a model trained for item based similarity also choose from a given set of items?	--	YES	NO			A
85	What are common feature selection methods in regression task?	--	correlation coefficient	Greedy algorithms	All above	None of these	C

86	The parameter_ ______ allows specifying the percentage of elements to put into the test/training set	--	test_size	training_ size	All above	None of these	C
87	In many classification problems, the target _______ is made up of categorical labels which cannot immediately be processed by any algorithm.	--	random_ state	dataset	test_size	All above	B
88	________adopts a dictionary-oriented approach, associating to each category label a progressive integer number.	--	LabelEncoder class	LabelBinarizer class	DictVectorizer	FeatureHasher	A
89	_________is much more difficult because it's necessary to determine a supervised strategy to train a model for each feature and, finally, to	--	Removing the whole line	Creating sub-model to predict those features	Using an automatic strategy to input them according to the other known values	All above	B

	predict their value						
90	How it's possible to use a different placeholder through the parameter_ _______.	--	regression	classification	random_state	missing_values	D
91	If you need a more powerful scaling feature, with a superior control on outliers and the possibility to select a quantile range, there's also the class_________.	--	RobustScaler	DictVectorizer	LabelBinarizer	FeatureHasher	A
92	scikit-learn also provides a class for per-sample normalization, Normalizer. It can apply_________to each element of a dataset	--	max, l0 and l1 norms	max, l1 and l2 norms	max, l2 and l3 norms	max, l3 and l4 norms	B
93	There are also many univariate methods that can be used in	--	F-tests and p-values	chi-square	ANOVA	All above	A

	order to select the best features according to specific criteria based on_________.						
94	Which of the following selects only a subset of features belonging to a certain percentile	--	SelectPercentile	FeatureHasher	SelectKBest	All above	A
95	_________performs a PCA with non-linearly separable data sets.	--	SparsePCA	KernelPCA	SVD	None of the Mentioned	B
96	If two variables are correlated, is it necessary that they have a linear relationship?	--	Yes	No			B
97	Correlated variables can have zero correlation coeffficient. True or False?	--	TRUE	FALSE			A
98	Suppose we fit "Lasso Regression" to a data set, which has 100 features	--	It is more likely for X1 to be excluded from the model	It is more likely for X1 to be included in the model	Can't say	None of these	B

	(X1,X2…X100). Now, we rescale one of these feature by multiplying with 10 (say that feature is X1), and then refit Lasso regression with the same regularization parameter. Now, which of the following option will be correct?						
99	If Linear regression model perfectly first i.e., train error is zero, then _______________________	--	Test error is also always zero	Test error is non zero	Couldn't comment on Test error	Test error is equal to Train error	C
100	Which of the following metrics can be used for evaluating regression models?i) R Squaredii) Adjusted R Squarediii) F Statisticsiv) RMSE / MSE / MAE	--	ii and iv	i and ii	ii, iii and iv	i, ii, iii and iv	D
101	In syntax of linear model lm(formula, data,..),	--	Matrix	Vector	Array	List	B

	data refers to _______						
102	Linear Regression is a supervised machine learning algorithm.	--	TRUE	FALSE			A
103	It is possible to design a Linear regression algorithm using a neural network?	--	TRUE	FALSE			A
104	Which of the following methods do we use to find the best fit line for data in Linear Regression?	--	Least Square Error	Maximum Likelihood	Logarithmic Loss	Both A and B	A
105	Suppose you are training a linear regression model. Now consider these points.1. Overfitting is more likely if we have less data2. Overfitting is more likely when the hypothesis space is small.Which of the above statement(s	--	Both are False	1 is False and 2 is True	1 is True and 2 is False	Both are True	C

	) are correct?						
106	We can also compute the coefficient of linear regression with the help of an analytical method called “Normal Equation”. Which of the following is/are true about “Normal Equation”?1. We don’t have to choose the learning rate2. It becomes slow when number of features is very large3. No need to iterate	--	1 and 2	1 and 3.	2 and 3.	1,2 and 3.	D
107	Which of the following option is true regarding “Regression” and “Correlation” ?Note: y is dependent variable and x is independent variable.	--	The relationship is symmetric between x and y in both.	The relationship is not symmetric between x and y in both.	The relationship is not symmetric between x and y in case of correlation but in case of regression it is symmetric.	The relationship is symmetric between x and y in case of correlation but in case of regression it is not symmetric.	D

108	In a simple linear regression model (One independent variable), If we change the input variable by 1 unit. How much output variable will change?	--	by 1	no change	by intercept	by its slope	D
109	Generally, which of the following method(s) is used for predicting continuous dependent variable?1. Linear Regression2. Logistic Regression	--	1 and 2	only 1	only 2	None of these.	B
110	How many coefficients do you need to estimate in a simple linear regression model (One independent variable)?	--	1	2	3	4	B
111	Suppose you are building a SVM model on data X. The data X can be error prone which means that you should not trust any specific data point	--	We can still classify data correctly for given setting of hyper parameter C	We can not classify data correctly for given setting of hyper parameter C	Can't Say	None of these	A

	too much. Now think that you want to build a SVM model which has quadratic kernel function of polynomial degree 2 that uses Slack variable C as one of it's hyper parameter. What would happen when you use very large value of C(C->infinity)?						
112	SVM can solve linear and non-linear probl ems	--	TRUE	FALSE			A
113	The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.	--	TRUE	FALSE			A
114	Hyperplanes are ___________	--	usual	decision	parallel		B

	___boundaries that help classify the data points.						
115	When the C parameter is set to infinite, which of the following holds true?	--	The optimal hyperplane if exists, will be the one that completely separates the data	The soft-margin classifier will separate the data	None of the above		A
116	SVM is a ---------------- learning	--	Supervised	Unsupervised	Both	None	A
117	The linear **SVM** classifier works by drawing a straight line between two classes	--	True	FALSE			A
118	In a real problem, you should check to see if the SVM is separable and then include slack variables if it is not separable.	--	TRUE	FALSE			B
119	Which of the following are real world applications of the SVM?	--	Text and Hypertext Categorization	Image Classification	Clustering of News Articles	All of the above	D
120	The _____of the hyperplane depends	--	dimension	classification	reduction		A

	upon the number of features.						
121	Hyperplanes are decision boundaries that help classify the data points.	--	TRUE	FALSE			A
122	SVM algorithms use a set of mathematical functions that are defined as the kernel.	--	TRUE	FALSE			A
123	Naive Bayes classifiers are a collection ----------------- of algorithms	--	Classification	Clustering	Regression	All	A
124	In given image, P(H|E) is___________probability.	bayes.jpg	Posterior	Prior			A
125	Solving a non linear separation problem with a hard margin Kernelized SVM (Gaussian RBF Kernel) might lead to overfitting		True	FALSE			A
126	100 people are at party. Given data gives information about how many wear pink or not,	man.jpg	TRUE	FALSE			A

	and if a man or not. Imagine a pink wearing guest leaves, was it a man?						
127	For the given weather data, Calculate probability of playing	weather data.jpg	0.4	0.64	0.29	0.75	B
128	In SVM, Kernel function is used to map a lower dimensional data into a higher dimensional data.	--	TRUE	FALSE			A
129	In SVR we try to fit the error within a certain threshold.	--	TRUE	FALSE			A
130	When the C parameter is set to infinite, which of the following holds true?	--	The optimal hyperplane if exists, will be the one that completely separates the data	The soft-margin classifier will separate the data	None of the above		A

Question	**Image**	**a**	**b**	**c**	**d**	**Correct Answer**
Write down question	img.jpg	Option a	Option b	Option c	Option d	a/b/c/d
Which of the following is		fast	accuracy	scalable	All above	D

characteristic of best machine learning method ?						
What are the different Algorithm techniques in Machine Learning?		Supervised Learning and Semi-supervised Learning	Unsupervised Learning and Transduction	Both A & B	None of the Mentioned	C
_______can be adopted when it's necessary to categorize a large amount of data with a few complete examples or when there's the need to impose some constraints to a clustering algorithm.		Supervised	Semi-supervised	Reinforcement	Clusters	B
In reinforcement learning, this feedback is usually called as____.		Overfitting	Overlearning	Reward	None of above	C
In the last decade, many researchers started training bigger and bigger models, built with several different layers that's why this approach is		Deep learning	Machine learning	Reinforcement learning	Unsupervised learning	A

called______.						
What does learning exactly mean?		Robots are programed so that they can perform the task based on data they gather from sensors.	A set of data is used to discover the potentially predictive relationship.	Learning is the ability to change according to external stimuli and remembering most of all previous experiences.	It is a set of data is used to discover the potentially predictive relationship.	C
When it is necessary to allow the model to develop a generalization ability and avoid a common problem called______.		Overfitting	Overlearning	Classification	Regression	A
Techniques involve the usage of both labeled and unlabeled data is called___.		Supervised	Semi-supervised	Unsupervised	None of the above	B
there's a growing interest in pattern recognition and associative memories whose structure and functioning are similar to what happens in the		Regression	Accuracy	Modelfree	Scalable	C

neocortex. Such an approach also allows simpler algorithms called _____						
______ showed better performance than other approaches, even without a context-based model		Machine learning	Deep learning	Reinforcement learning	Supervised learning	B
Which of the following sentence is correct?	--	Machine learning relates with the study, design and development of the algorithms that give computers the capability to learn without being explicitly programmed.	Data mining can be defined as the process in which the unstructured data tries to extract knowledge or unknown interesting patterns.	Both A & B	None of the above	C
What is ‘Overfitting’ in Machine learning?	--	when a statistical model describes random error or noise instead of underlying relationship ‘overfitting’ occurs.	Robots are programed so that they can perform the task based on data they gather from sensors.	While involving the process of learning ‘overfitting’ occurs.	a set of data is used to discover the potentially predictive relationship	A

What is 'Test set'?	--	Test set is used to test the accuracy of the hypotheses generated by the learner.	It is a set of data is used to discover the potentially predictive relationship.	Both A & B	None of above	A
what is the function of 'Supervised Learning'?	--	Classifications, Predict time series, Annotate strings	Speech recognition, Regression	Both A & B	None of above	C
Commons unsupervised applications include	--	Object segmentation	Similarity detection	Automatic labeling	All above	D
Reinforcement learning is particularly efficient when_____ __________.	--	the environment is not completely deterministic	it's often very dynamic	it's impossible to have a precise error measure	All above	D
During the last few years, many _______ algorithms have been applied to deep neural networks to learn the best policy for playing Atari video games and to teach an agent how to associate the right action with an input representing the state.	--	Logical	Classical	Classification	None of above	D

Common deep learning applications include____	--	Image classification, Real-time visual tracking	Autonomous car driving, Logistic optimization	Bioinformatics, Speech recognition	All above	D
if there is only a discrete number of possible outcomes (called categories), the process becomes a______.	--	Regression	Classification.	Modelfree	Categories	B
Which of the following are supervised learning applications	--	Spam detection, Pattern detection, Natural Language Processing	Image classification, Real-time visual tracking	Autonomous car driving, Logistic optimization	Bioinformatics, Speech recognition	A
Let's say, you are working with categorical feature(s) and you have not looked at the distribution of the categorical variable in the test data. You want to apply one hot encoding (OHE) on the categorical feature(s). What challenges you may face if you have	--	All categories of categorical variable are not present in the test dataset.	Frequency distribution of categories is different in train as compared to the test dataset.	Train and Test always have same distribution.	Both A and B	D

applied OHE on a categorical variable of train dataset?						
Which of the following sentence is FALSE regarding regression?	--	It relates inputs to outputs.	It is used for prediction.	It may be used for interpretation.	It discovers causal relationships.	D
Which of the following method is used to find the optimal features for cluster analysis	--	k-Means	Density-Based Spatial Clustering	Spectral Clustering Find clusters	All above	D
scikit-learn also provides functions for creating dummy datasets from scratch:	--	make_classification()	make_regression()	make_blobs()	All above	D
______which can accept a NumPy RandomState generator or an integer seed.	--	make_blobs	random_state	test_size	training_size	B
In many classification problems, the target dataset is made up of categorical labels which cannot immediately be processed	--	1	2	3	4	B

by any algorithm. An encoding is needed and scikit-learn offers at least_____valid options						
In which of the following each categorical label is first turned into a positive integer and then transformed into a vector where only one feature is 1 while all the others are 0.	--	LabelEncoder class	DictVectorizer	LabelBinarizer class	FeatureHasher	C
_______is the most drastic one and should be considered only when the dataset is quite large, the number of missing features is high, and any prediction could be risky.	--	Removing the whole line	Creating sub-model to predict those features	Using an automatic strategy to input them according to the other known values	All above	A
It's possible to specify if the scaling process must include both mean and standard deviation using the	--	with_mean=True/False	with_std=True/False	Both A & B	None of the Mentioned	C

parameters_ ________.						
Which of the following selects the best K high-score features.	--	SelectPercentile	FeatureHasher	SelectKBest	All above	C
How does number of observations influence overfitting? Choose the correct answer(s). Note: Rest all parameters are same1. In case of fewer observations, it is easy to overfit the data.2. In case of fewer observations, it is hard to overfit the data.3. In case of more observations, it is easy to overfit the data.4. In case of more observations, it is hard to overfit the data.	--	1 and 4	2 and 3	1 and 3	None of theses	A

Suppose you have fitted a complex regression model on a dataset. Now, you are using Ridge regression with tuning parameter lambda to reduce its complexity. Choose the option(s) below which describes relationship of bias and variance with lambda.	--	In case of very large lambda; bias is low, variance is low	In case of very large lambda; bias is low, variance is high	In case of very large lambda; bias is high, variance is low	In case of very large lambda; bias is high, variance is high	C
What is/are true about ridge regression? 1. When lambda is 0, model works like linear regression model2. When lambda is 0, model doesn't work like linear regression model3. When lambda goes to	--	1 and 3	1 and 4	2 and 3	2 and 4	A

infinity, we get very, very small coefficients approaching 04. When lambda goes to infinity, we get very, very large coefficients approaching infinity						
Which of the following method(s) does not have closed form solution for its coefficients?	--	Ridge regression	Lasso	Both Ridge and Lasso	None of both	B
Function used for linear regression in R is ____________	--	lm(formula, data)	lr(formula, data)	lrm(formula, data)	regression.linear(formula, data)	A
In the mathematical Equation of Linear Regression Y = β1 + β2X + ϵ, (β1, β2) refers to ____________	--	(X-intercept, Slope)	(Slope, X-Intercept)	(Y-Intercept, Slope)	(slope, Y-Intercept)	C

Suppose that we have N independent variables (X1,X2…Xn) and dependent variable is Y. Now Imagine that you are applying linear regression by fitting the best fit line using least square error on this data. You found that correlation coefficient for one of it's variable(Say X1) with Y is -0.95.Which of the following is true for X1?	--	Relation between the X1 and Y is weak	Relation between the X1 and Y is strong	Relation between the X1 and Y is neutral	Correlation can't judge the relationship	B
We have been given a dataset with n records in which we have input attribute as x and output attribute as y. Suppose we use a	--	Increase	Decrease	Remain constant	Can't Say	D

linear regression method to model this data. To test our linear regressor, we split the data in training set and test set randomly. Now we increase the training set size gradually. As the training set size increases, what do you expect will happen with the mean training error?						
We have been given a dataset with n records in which we have input attribute as x and output attribute as y. Suppose we use a linear regression method to model this data. To test our linear	--	Bias increases and Variance increases	Bias decreases and Variance increases	Bias decreases and Variance decreases	Bias increases and Variance decreases	D

regressor, we split the data in training set and test set randomly. What do you expect will happen with bias and variance as you increase the size of training data?						
Suppose, you got a situation where you find that your linear regression model is under fitting the data. In such situation which of the following options would you consider?1. I will add more variables2. I will start introducing polynomial degree variables3. I will remove some variables	--	1 and 2	2 and 3	1 and 3	1, 2 and 3	A

Problem: Players will play if weather is sunny. Is this statement is correct?	weather data.jpg	TRUE	FALSE			A
Multinomial Naïve Bayes Classifier is ____________distribution		Continuous	Discrete	Binary		B
For the given weather data, Calculate probability of not playing	weather data.jpg	0.4	0.64	0.36	0.5	C
Suppose you have trained an SVM with linear decision boundary after training SVM, you correctly infer that your SVM model is under fitting.Which of the following option would you more likely to consider iterating SVM next time?	--	You want to increase your data points	You want to decrease your data points	You will try to calculate more variables	You will try to reduce the features	C

The minimum time complexity for training an SVM is O(n2). According to this fact, what sizes of datasets are not best suited for SVM's?	--	Large datasets	Small datasets	Medium sized datasets	Size does not matter	A
The effectiveness of an SVM depends upon:	--	Selection of Kernel	Kernel Parameters	Soft Margin Parameter C	All of the above	D
What do you mean by generalization error in terms of the SVM?	--	How far the hyperplane is from the support vectors	How accurately the SVM can predict outcomes for unseen data	The threshold amount of error in an SVM		B
What do you mean by a hard margin?	--	The SVM allows very low error in classification	The SVM allows high amount of error in classification	None of the above		A
We usually use feature normalization before using the Gaussian kernel in SVM. What is true about feature	--	1	1 and 2	1 and 3	2 and 3	B

normalization? 1. We do feature normalization so that new feature will dominate other 2. Some times, feature normalization is not feasible in case of categorical variables3. Feature normalization always helps when we use Gaussian kernel in SVM						
Support vectors are the data points that lie closest to the decision surface.	--	TRUE	FALSE			A
Which of the following is **not** supervised learning?	--	PCA	Decision Tree	Naive Bayesian	Linerar regression	A
Suppose you are using RBF kernel in SVM with high Gamma value. What does	--	The model would consider even far away points from hyperpla	The model would consider only the points close to the hyperplane for	The model would not be affected by distance of points from hyperpla	None of the above	B

this signify?		ne for modeling	modelin g	ne for modelin g		
Gaussian Naïve Bayes Classifier is ___________ __distribution	--	Continuous	Discrete	Binary		A
If I am using all features of my dataset and I achieve 100% accuracy on my training set, but ~70% on validation set, what should I look out for?	--	Underfitting	Nothing, the model is perfect	Overfitting		C
What is the purpose of performing cross-validation?	--	a. To assess the predictive performance of the models	b. To judge how the trained model performs outside the sample on test data	c. Both A and B		C

Which of the following is true about Naive Bayes ?	--	a. Assumes that all the features in a dataset are equally important	b. Assumes that all the features in a dataset are independent	c. Both A and B	d. None of the above option	C
Suppose you are using a Linear SVM classifier with 2 class classification problem. Now you have been given the following data in which some points are circled red that are representing support vectors.If you remove the following any one red points from the data. Does the decision boundary will change?	svm.jpg	yes	no			A
Linear SVMs have no	--	TRUE	FALSE			B

hyperparameters that need to be set by cross-validation						
For the given weather data, what is the probability that players will play if weather is sunny	weather data.jpg	0.5	0.26	0.73	0.6	D
100 people are at party. Given data gives information about how many wear pink or not, and if a man or not. Imagine a pink wearing guest leaves, what is the probability of being a man	man.jpg	0.4	0.2	0.6	0.45	B
Problem: Players will play if weather is sunny. Is this statement is correct?	weather data.jpg	TRUE	FALSE			a
For the given weather	weather data.jpg	0.4	0.64	0.29	0.75	b

data, Calculate probability of playing						
For the given weather data, Calculate probability of not playing	weather data.jpg	0.4	0.64	0.36	0.5	c
For the given weather data, what is the probability that players will play if weather is sunny	weather data.jpg	0.5	0.26	0.73	0.6	d
100 people are at party. Given data gives informatio n about how many wear pink or not, and if a man or not. Imagine a pink wearing guest leaves, what is the probability of being a man	man.jpg	0.4	0.2	0.6	0.45	b
100 people are at party.	man.jpg	TRUE	FALSE			a

Given data gives information about how many wear pink or not, and if a man or not. Imagine a pink wearing guest leaves, was it a man?						
What do you mean by generalization error in terms of the SVM?		How far the hyperplane is from the support vectors	How accurately the SVM can predict outcomes for unseen data	The threshold amount of error in an SVM		b
What do you mean by a hard margin?		The SVM allows very low error in classification	The SVM allows high amount of error in classification	None of the above		a
The minimum time complexity for training an SVM is O(n2). According to this fact, what sizes of datasets are not best suited for SVM's?		Large datasets	Small datasets	Medium sized datasets	Size does not matter	a
The effectiveness of an SVM		Selection of Kernel	Kernel Parameters	Soft Margin Parameter C	All of the above	d

depends upon:						
Support vectors are the data points that lie closest to the decision surface.		TRUE	FALSE			a
The SVM's are less effective when:		The data is linearly separable	The data is clean and ready to use	The data is noisy and contains overlapping points		c
Suppose you are using RBF kernel in SVM with high Gamma value. What does this signify?		The model would consider even far away points from hyperplane for modeling	The model would consider only the points close to the hyperplane for modeling	The model would not be affected by distance of points from hyperplane for modeling	None of the above	b
The cost parameter in the SVM means:		The number of cross-validations to be made	The kernel to be used	The tradeoff between misclassification and simplicity of the model	None of the above	c
If I am using all features of my dataset and I achieve 100% accuracy on my training set, but ~70% on validation set, what		Underfitting	Nothing, the model is perfect	Overfitting		c

should I look out for?						
Which of the following are real world applications of the SVM?		Text and Hypertext Categorization	Image Classification	Clustering of News Articles	All of the above	d
Suppose you have trained an SVM with linear decision boundary after training SVM, you correctly infer that your SVM model is under fitting.Which of the following option would you more likely to consider iterating SVM next time?		You want to increase your data points	You want to decrease your data points	You will try to calculate more variables	You will try to reduce the features	c
We usually use feature normalization before using the Gaussian kernel in SVM. What is true about feature normalization? 1. We do feature normalization so that new feature will		1	1 and 2	1 and 3	2 and 3	b

dominate other 2. Some times, feature normalization is not feasible in case of categorical variables3. Feature normalization always helps when we use Gaussian kernel in SVM						
Linear SVMs have no hyperparameters that need to be set by cross-validation		TRUE	FALSE			b
In a real problem, you should check to see if the SVM is separable and then include slack variables if it is not separable.		TRUE	FALSE			b

This sheet is for 1 Mark questions							
S.r No	Question	Image	a	b	c	d	Correct Answer
	Write down question	img.jpg	Option a	Option b	Option c	Option d	a/b/c/d
1	In reinforcement learning if feedback is negative one it is defined as____.		Penalty	Overlearning	Reward	None of above	A
2	According to____ , it's a key success factor for the survival and evolution of all species.		Claude Shannon's theory	Gini Index	Darwin's theory	None of above	C
3	How can you avoid overfitting ?		By using a lot of data	By using inductive machine learning	By using validation only	None of above	A
4	What are the popular algorithms of Machine Learning?		Decision Trees and Neural Networks (back propagation)	Probabilistic networks and Nearest Neighbor	Support vector machines	All	D
5	What is 'Training set'?		Training set is used to test the accuracy of the hypotheses generated by the learner.	A set of data is used to discover the potentially predictive relationship.	Both A & B	None of above	B

6	Common deep learning applications include____		Image classification, Real-time visual tracking	Autonomous car driving, Logistic optimization	Bioinformatics, Speech recognition	All above	D
7	what is the function of ‘Supervised Learning’?		Classifications, Predict time series, Annotate strings	Speech recognition, Regression	Both A & B	None of above	C
8	Commons unsupervised applications include		Object segmentation	Similarity detection	Automatic labeling	All above	D
9	Reinforcement learning is particularly efficient when______________.		the environment is not completely deterministic	it's often very dynamic	it's impossible to have a precise error measure	All above	D
10	if there is only a discrete number of possible outcomes (called categories), the process becomes a______.		Regression	Classification.	Modelfree	Categories	B
11	Which of the following are supervised learning applications		Spam detection, Pattern detection, Natural Language Processing	Image classification, Real-time visual tracking	Autonomous car driving, Logistic optimization	Bioinformatics, Speech recognition	A
12	During the last few years, many ______ algorithms		Logical	Classical	Classification	None of above	D

	have been applied to deep neural networks to learn the best policy for playing Atari video games and to teach an agent how to associate the right action with an input representing the state.						
13	Which of the following sentence is correct?		Machine learning relates with the study, design and development of the algorithms that give computers the capability to learn without being explicitly programmed.	Data mining can be defined as the process in which the unstructured data tries to extract knowledge or unknown interesting patterns.	Both A & B	None of the above	C
14	What is ‘Overfitting’ in Machine learning?		when a statistical model describes random error or noise instead of underlying relationship ‘overfitti	Robots are programed so that they can perform the task based on data they gather from sensors.	While involving the process of learning ‘overfitting’ occurs.	a set of data is used to discover the potentially predictive relationship	A

			ng' occurs.				
15	What is 'Test set'?		Test set is used to test the accuracy of the hypotheses generated by the learner.	It is a set of data is used to discover the potentially predictive relationship.	Both A & B	None of above	A
16	________is much more difficult because it's necessary to determine a supervised strategy to train a model for each feature and, finally, to predict their value		Removing the whole line	Creating sub-model to predict those features	Using an automatic strategy to input them according to the other known values	All above	B
17	How it's possible to use a different placeholder through the parameter________.		regression	classification	random_state	missing_values	D
18	If you need a more powerful scaling feature, with a superior control on outliers and the possibility to select a quantile range, there's also the class_________.		RobustScaler	DictVectorizer	LabelBinarizer	FeatureHasher	A

19	scikit-learn also provides a class for per-sample normalization, Normalizer. It can apply_________to each element of a dataset		max, l0 and l1 norms	max, l1 and l2 norms	max, l2 and l3 norms	max, l3 and l4 norms	B
20	There are also many univariate methods that can be used in order to select the best features according to specific criteria based on_________.		F-tests and p-values	chi-square	ANOVA	All above	A
21	Which of the following selects only a subset of features belonging to a certain percentile		SelectPercentile	FeatureHasher	SelectKBest	All above	A
22	_________performs a PCA with non-linearly separable data sets.		SparsePCA	KernelPCA	SVD	None of the Mentioned	B
23	A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college.		Feature F1 is an example of nominal variable.	Feature F1 is an example of ordinal variable.	It doesn’t belong to any of the above category.	Both of these	B

	Which of the following statement is true in following case?						
24	What would you do in PCA to get the same projection as SVD?		Transform data to zero mean	Transform data to zero median	Not possible	None of these	A
25	What is PCA, KPCA and ICA used for?		Principal Components Analysis	Kernel based Principal Component Analysis	Independent Component Analysis	All above	D
26	Can a model trained for item based similarity also choose from a given set of items?		YES	NO			A
27	What are common feature selection methods in regression task?		correlation coefficient	Greedy algorithms	All above	None of these	C
28	The parameter______ allows specifying the percentage of elements to put into the test/training set		test_size	training_size	All above	None of these	C
29	In many classification problems, the target ______ is made up of categorical labels which cannot immediately		random_state	dataset	test_size	All above	B

	be processed by any algorithm.						
30	________adopts a dictionary-oriented approach, associating to each category label a progressive integer number.		LabelEncoder class	LabelBinarizer class	DictVectorizer	FeatureHasher	A
31	If Linear regression model perfectly first i.e., train error is zero, then ______________________		a) Test error is also always zero	b) Test error is non zero	c) Couldn't comment on Test error	d) Test error is equal to Train error	c
32	Which of the following metrics can be used for evaluating regression models?i) R Squaredii) Adjusted R Squarediii) F Statisticsiv) RMSE / MSE / MAE		a) ii and iv	b) i and ii	c) ii, iii and iv	d) i, ii, iii and iv	d
33	How many coefficients do you need to estimate in a simple linear regression model (One independent variable)?		a) 1	b) 2	c) 3	d) 4	b
34	In a simple linear		a) by 1	b) no change	c) by intercept	d) by its slope	d

	regression model (One independent variable), If we change the input variable by 1 unit. How much output variable will change?							
35	Function used for linear regression in R is ___________		a) lm(formula, data)	b) lr(formula, data)	c) lrm(formula, data)	d) regression.linear(formula, data)	a
36	In syntax of linear model lm(formula, data,..), data refers to _______		a) Matrix	b) Vector	c) Array	d) List	b
37	In the mathematical Equation of Linear Regression $Y = \beta1 + \beta2X + \epsilon$, $(\beta1, \beta2)$ refers to ___________		a) (X-intercept, Slope)	b) (Slope, X-Intercept)	c) (Y-Intercept, Slope)	d) (slope, Y-Intercept)	c
38	Linear Regression is a supervised machine learning algorithm.		A) TRUE	B) FALSE			a
39	It is possible to design a Linear regression algorithm using a neural network?		A) TRUE	B) FALSE			a

40	Which of the following methods do we use to find the best fit line for data in Linear Regression?		A) Least Square Error	B) Maximum Likelihood	C) Logarithmic Loss	D) Both A and B	a
41	Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?		A) AUC-ROC	B) Accuracy	C) Logloss	D) Mean-Squared-Error	d
42	Which of the following is true about Residuals ?		A) Lower is better	B) Higher is better	C) A or B depend on the situation	D) None of these	a
43	Overfitting is more likely when you have huge amount of data to train?		A) TRUE	B) FALSE			b
44	Which of the following statement is true about outliers in Linear regression?		A) Linear regression is sensitive to outliers	B) Linear regression is not sensitive to outliers	C) Can’t say	D) None of these	a
45	Suppose you plotted a scatter plot between the residuals and		A) Since the there is a relationship means our	B) Since the there is a relationship means our	C) Can’t say	D) None of these	a

	predicted values in linear regression and you found that there is a relationship between them. Which of the following conclusion do you make about this situation?		model is not good	model is good			
46	Naive Bayes classifiers are a collection ----------------- of algorithms		Classification	Clustering	Regression	All	a
47	Naive Bayes classifiers is ________________ Learning		Supervised	Unsupervised	Both	None	a
48	Features being classified is independent of each other in Naïve Bayes Classifier		False	TRUE			b
49	Features being classified is ____________ of each other in Naïve Bayes Classifier		Independent	Dependent	Partial Dependent	None	a
50	Bayes Theorem is given by where 1. P(H) is the	bayes.jpg	True	FALSE			a

	probability of hypothesis H being true. 2. P(E) is the probability of the evidence(regardless of the hypothesis). 3. P(E|H) is the probability of the evidence given that hypothesis is true. 4. P(H|E) is the probability of the hypothesis given that the evidence is there.						
51	In given image, P(H|E) is___________probability.	bayes.jpg	Posterior	Prior			a
52	In given image, P(H) is___________probability.	bayes.jpg	Posterior	Prior			b
53	Conditional probability is a measure of the probability of an event given that another event has already occurred.		True	FALSE			a

54	Bayes' theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event.		True	FALSE			a
55	Bernoulli Naïve Bayes Classifier is ___________ _distribution		Continuous	Discrete	Binary		c
56	Multinomial Naïve Bayes Classifier is ___________ _distribution		Continuous	Discrete	Binary		b
57	Gaussian Naïve Bayes Classifier is ___________ _distribution		Continuous	Discrete	Binary		a
58	Binarize parameter in BernoulliNB scikit sets threshold for binarizing of sample features.		True	FALSE			a
59	Gaussian distribution when plotted, gives a bell shaped curve which is symmetric		Mean	Variance	Discrete	Random	a

	about the ________ of the feature values.						
60	SVMs directly give us the posterior probabilities P(y = 1jx) and P(y = 1jx)		True	FALSE			b
61	Any linear combination of the components of a multivariate Gaussian is a univariate Gaussian.		True	FALSE			a
62	Solving a non linear separation problem with a hard margin Kernelized SVM (Gaussian RBF Kernel) might lead to overfitting		True	FALSE			a
63	SVM is a ------------------ algorithm		Classification	Clustering	Regression	All	a
64	SVM is a ------------------ learning		Supervised	Unsupervised	Both	None	a
65	The linear **SVM** classifier works by drawing a straight line between two classes		True	FALSE			a
66	Which of the following	--	cl_forecastB	cl_nowcastC	cl_precast D	None of the Mentioned	D

	function provides unsupervised prediction ?						
67	Which of the following is characteristic of best machine learning method ?	--	fast	accuracy	scalable	All above	D
68	What are the different Algorithm techniques in Machine Learning?	--	Supervised Learning and Semi-supervised Learning	Unsupervised Learning and Transduction	Both A & B	None of the Mentioned	C
69	What is the standard approach to supervised learning?	--	split the set of example into the training set and the test	group the set of example into the training set and the test	a set of observed instances tries to induce a general rule	learns programs from data	A
70	Which of the following is not Machine Learning?	--	Artificial Intelligence	Rule based inference	Both A & B	None of the Mentioned	B
71	What is Model Selection in Machine Learning?	--	The process of selecting models among different mathematical models, which	when a statistical model describes random error or noise instead of underlying	Find interesting directions in data and find novel observations/ database cleaning	All above	A

			are used to describe the same data set	relationship			
72	Which are two techniques of Machine Learning ?	--	Genetic Programming and Inductive Learning	Speech recognition and Regression	Both A & B	None of the Mentioned	A
73	Even if there are no actual supervisors ________ learning is also based on feedback provided by the environment	--	Supervised	Reinforcement	Unsupervised	None of the above	B
74	What does learning exactly mean?	--	Robots are programed so that they can perform the task based on data they gather from sensors.	A set of data is used to discover the potentially predictive relationship.	Learning is the ability to change according to external stimuli and remembering most of all previous experiences.	It is a set of data is used to discover the potentially predictive relationship.	C
75	When it is necessary to allow the model to develop	--	Overfitting	Overlearning	Classification	Regression	A

	a generalization ability and avoid a common problem called_______.						
76	Techniques involve the usage of both labeled and unlabeled data is called___.	--	Supervised	Semi-supervised	Unsupervised	None of the above	B
77	In reinforcement learning if feedback is negative one it is defined as_____.	--	Penalty	Overlearning	Reward	None of above	A
78	According to_____ , it's a key success factor for the survival and evolution of all species.	--	Claude Shannon's theory	Gini Index	Darwin's theory	None of above	C
79	A supervised scenario is characterized by the concept of a ______.	--	Programmer	Teacher	Author	Farmer	B
80	overlearning causes due to an excessive _______.	--	Capacity	Regression	Reinforcement	Accuracy	A
81	Which of the following	--	PCA	K-Means	None of the above		A

	is an example of a deterministic algorithm?						
82	Which of the following model model include a backwards elimination feature selection routine?	--	MCV	MARS	MCRS	All above	B
83	Can we extract knowledge without apply feature selection	--	YES	NO			A
84	While using feature selection on the data, is the number of features decreases.	--	NO	YES			B
85	Which of the following are several models for feature extraction	--	regression	classification	None of the above		C
86	______ provides some built-in datasets that can be used for testing purposes.	--	scikit-learn	classification	regression	None of the above	A

87	While using ______ all labels are turned into sequential numbers.	--	LabelEncoder class	LabelBinarizer class	DictVectorizer	FeatureHasher	A
88	________produce sparse matrices of real numbers that can be fed into any machine learning model.	--	DictVectorizer	FeatureHasher	Both A & B	None of the Mentioned	C
89	scikit-learn offers the class_______, which is responsible for filling the holes using a strategy based on the mean, median, or frequency	--	LabelEncoder	LabelBinarizer	DictVectorizer	Imputer	D
90	Which of the following scale data by removing elements that don't belong to a given range or by considering a maximum absolute value.	--	MinMaxScaler	MaxAbsScaler	Both A & B	None of the Mentioned	C

91	scikit-learn also provides a class for per-sample normalization,______	--	Normalizer	Imputer	Classifier	All above	A
92	_______data set with many features contains information proportional to the independence of all features and their variance.	--	normalized	unnormalized	Both A & B	None of the Mentioned	B
93	In order to assess how much information is brought by each component, and the correlation among them, a useful tool is the______.	--	Concuttent matrix	Convergance matrix	Supportive matrix	Covariance matrix	D
94	The______ parameter can assume different values which determine how the data matrix is initially processed.	--	run	start	init	stop	C
95	_______allows exploiting	--	SparsePCA	KernelPCA	SVD	init parameter	A

	the natural sparsity of data while extracting principal components.						
96	Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?	--	AUC-ROC	Accuracy	Logloss	Mean-Squared-Error	D
97	Which of the following is true about Residuals ?	--	Lower is better	Higher is better	A or B depend on the situation	None of these	A
98	Overfitting is more likely when you have huge amount of data to train?	--	TRUE	FALSE			B
99	Which of the following statement is true about outliers in Linear regression?	--	Linear regression is sensitive to outliers	Linear regression is not sensitive to outliers	Can't say	None of these	A
100	Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and you found that there is a relationship	--	Since the there is a relationship means our model is not good	Since the there is a relationship means our model is good	Can't say	None of these	A

	between them. Which of the following conclusion do you make about this situation?						
101	Let’s say, a “Linear regression” model perfectly fits the training data (train error is zero). Now, Which of the following statement is true?	--	You will always have test error zero	You can not have test error zero	None of the above		C
102	In a linear regression problem, we are using “R-squared” to measure goodness-of-fit. We add a feature in linear regression model and retrain the same model.Which of the following option is true?	--	If R Squared increases, this variable is significant.	If R Squared decreases, this variable is not significant.	Individually R squared cannot tell about variable importance. We can’t say anything about it right now.	None of these.	C
103	Which of the one is true about Heteroskedasticity?	--	Linear Regression with varying error terms	Linear Regression with constant error terms	Linear Regression with zero error terms	None of these	A
104	Which of the following assumptions	--	1,2 and 3.	1,3 and 4.	1 and 3.	All of above.	D

	do we make while deriving linear regression parameters? 1. The true relationship between dependent y and predictor x is linear2. The model errors are statistically independent 3. The errors are normally distributed with a 0 mean and constant standard deviation4. The predictor x is non-stochastic and is measured error-free						
105	To test linear relationship of y(dependent) and x(independent) continuous variables, which of the following plot best suited?	--	Scatter plot	Barchart	Histograms	None of these	A

106	which of the following step / assumption in regression modeling impacts the trade-off between under-fitting and over-fitting the most.	--	The polynomial degree	Whether we learn the weights by matrix inversion or gradient descent	The use of a constant-term		A
107	Can we calculate the skewness of variables based on mean and median?	--	TRUE	FALSE			B
108	Which of the following is true about “Ridge” or “Lasso” regression methods in case of feature selection?	--	Ridge regression uses subset selection of features	Lasso regression uses subset selection of features	Both use subset selection of features	None of above	B
109	Which of the following statement(s) can be true post adding a variable in a linear regression model?1. R-Squared and Adjusted R-squared both increase2. R-Squared increases and	--	1 and 2	1 and 3	2 and 4	None of the above	A

	Adjusted R-squared decreases3. R-Squared decreases and Adjusted R-squared decreases4. R-Squared decreases and Adjusted R-squared increases						
110	How many coefficients do you need to estimate in a simple linear regression model (One independent variable)?	--	1	2	Can’t Say		B
111	In given image, P(H) is___________probability.	bayes.jpg	Posterior	Prior			B
112	Conditional probability is a measure of the probability of an event given that another event has already occurred.	--	True	FALSE			A
113	Gaussian distribution when plotted, gives a bell shaped curve which is	--	Mean	Variance	Discrete	Random	A

	symmetric about the ________ of the feature values.						
114	SVMs directly give us the posterior probabilities P(y = 1jx) and P(y = 1jx)	--	True	FALSE			B
115	SVM is a ----------------- algorithm	--	Classification	Clustering	Regression	All	A
116	What is/are true about kernel in SVM?1. Kernel function map low dimensional data to high dimensional space2. It's a similarity function	--	1	2	1 and 2	None of these	C
117	Suppose you are building a SVM model on data X. The data X can be error prone which means that you should not trust any specific data point too much. Now think	--	Misclassification would happen	Data will be correctly classified	Can't say	None of these	A

	that you want to build a SVM model which has quadratic kernel function of polynomial degree 2 that uses Slack variable C as one of it’s hyper parameter. What would happen when you use very small C (C~0)?						
118	The cost parameter in the SVM means:	--	The number of cross-validations to be made	The kernel to be used	The tradeoff between misclassification and simplicity of the model	None of the above	C
119	Bayes’ theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event.	--	True	FALSE			A
120	Bernoulli Naïve	--	Continuous	Discrete	Binary		C

	Bayes Classifier is ___________ distribution						
121	If you remove the non-red circled points from the data, the decision boundary will change?	svm.jpg	TRUE	FALSE			B
122	How do you handle missing or corrupted data in a dataset?	--	a. Drop missing rows or columns	b. Replace missing values with mean/median/mode	c. Assign a unique category to missing values	d. All of the above	D
123	Binarize parameter in BernoulliNB scikit sets threshold for binarizing of sample features.	--	True	FALSE			A
124	Which of the following statements about Naive Bayes is incorrect?	--	A. Attributes are equally important.	B. Attributes are statistically dependent of one another given the class value.	C. Attributes are statistically independent of one another given the class value.	D. Attributes can be nominal or numeric	B

125	The SVM's are less effective when:	--	The data is linearly separable	The data is clean and ready to use	The data is noisy and contains overlapping points		C
126	Naive Bayes classifiers is ________________ Learning	--	Supervised	Unsupervised	Both	None	A
127	Features being classified is independent of each other in Naïve Bayes Classifier	--	False	TRUE			B
128	Features being classified is ___________ of each other in Naïve Bayes Classifier	--	Independent	Dependent	Partial Dependent	None	A
129	Bayes Theorem is given by where 1. P(H) is the probability of hypothesis H being true. 2. P(E) is the probability of the evidence(regardless of the hypothesis).	bayes.jpg	True	FALSE			A

	3. P(E|H) is the probability of the evidence given that hypothesis is true. 4. P(H|E) is the probability of the hypothesis given that the evidence is there.						
130	Any linear combination of the components of a multivariate Gaussian is a univariate Gaussian.	--	True	FALSE			A

Subject: Machine Learning (BE-A&B)

Short Answer Questions

1) Why overfitting happens?
The possibility of overfitting exists as the criteria used for training the model is not the same as the criteria used to judge the efficacy of a model.

2) What are the five popular algorithms of Machine Learning?

- Decision Trees
- Neural Networks (back propagation)
- Probabilistic networks
- Nearest Neighbor
- Support vector machines

3) What are the three stages to build the hypotheses or model in machine learning?

- Model building
- Model testing
- Applying the model

4) What is the standard approach to supervised learning?
The standard approach to supervised learning is to split the set of example into the training set and the test.

5) What is not Machine Learning?

- Artificial Intelligence
- Rule based inference

6) What is classifier in machine learning?
A classifier in a Machine Learning is a system that inputs a vector of discrete or continuous feature values and outputs a single discrete value, the class.

7) Give a popular application of machine learning that you see on day to day basis?
The recommendation engine implemented by major ecommerce websites uses Machine Learning.

8. What are Recommender Systems?

Ans. It is a subclass of information filtering system that seeks to predict the "rating" or "preference" that a user would give to an item.

9 What is an Eigenvalue and Eigenvector?
Ans. Eigenvectors are used for understanding linear transformations.
Eigenvalue can be referred to as the strength of the transformation in the direction of eigenvector or the factor by which the compression occurs.

10 Which technique is used to predict categorical responses?
Ans. The classification techniques is used to predict categorical responses.

11. What is the difference between data science and big data?
Ans. Data science is a field applicable to any data sizes. Big data refers to the large amount of data which cannot be analysed by traditional methods.

12. Name some of the prominent resampling methods in data science
Ans. The Bootstrap, Permutation Tests, Cross-validation and Jackknife

13. What packages are used for Machine learning in Python and R?
Ans. There are various packages in Python and R:
Python – Orange, Pandas, NLTK, Matplotlib, and Scikit-learn are some of them
R – Arules, tm, Forecast and GGPlot are some of the packages

14. How do you check for data quality?
Ans. Some of the definitions used to check for data quality are:

- Completeness
- Consistency
- Uniqueness
- Integrity
- Conformity
- Accuracy

15. What is the difference between squared error and absolute error?
Ans. Squared error measures the average of the squares of the errors or deviations—that is, the difference between the estimator and what is estimated.
Absolute error is the difference between the measured or inferred value of a quantity and its actual value.

16. Differentiate between wide and long data formats?
Ans. In wide format, categorical data is always grouped.
Long data format is in which there are a number of instances with many variable and subject variable

17. What do you understand by Recall and Precision?
Ans. Precision is the fraction of retrieved instances that are relevant, while Recall is the fraction of relevant instances that are retrieved.

18. Explain the difference between type I and type II error.
Ans. Type I error is the rejection of a true null hypothesis or false positive finding, while the Type II error is the non-rejection of a false null hypothesis or false negative finding.

19 What are Recommender Systems?
Ans. Recommender systems are information filtering systems that predict which products will attract customers, but these systems are not ideal for every business situation. These systems are used in movies, news, research articles, products, etc. These systems are content and collaborative filtering based.

20. What is the main difference between overfitting and underfitting?
Ans. Overfitting – In overfitting, a statistical model describes any random error or noise, and occurs when a model is super complex. An overfit model has a poor predictive performance as it overreacts to minor fluctuations in training data.Underfitting – In underfitting, a statistical model is unable to capture the underlying data trend. This type of model also shows a poor predictive performance.

Multiple Choice Questions

1. Supervised learning and and unsupervised clustering both require t least one

A. **Hideen attribute**
B. Output attribute
C. Input attribute
D. Categorical attribute

2.Supervised learning and and unsupervised clustering in that supervised learning requires

A. At least one input attribute
B. Input attribute to be categorical
C. At least one output attribute
D. output attribute to be categorical

3.Select all multi class classification techniques from given techniques

(A) One-versus-all (OVA)
(B) All-versus-all (AVA)
(C) Error-Correcting Output-Coding (ECOC)
(D) None of these

4.Causes of overfitting

1.Small training dataset
2. Large number of features in a dataset
3. Noise in the dataset
4.None of the Above

5.Which of the following is characteristic of best machine learning method ?

1. Fast
2. Accuracy
3. Scalable
4. All of the Mentioned

6. Different learning methods does not include?
a) Memorization
b) Analogy
c) Deduction
d) Introduction

7. Which of the factors affect the performance of learner system does not include?
a) Representation scheme used
b) Training scenario
c) Type of feedback
d) Good data structures

8.. Which of the following is a categorical outcome?
a) RMSE
b) RSquared

c) Accuracy
d) All of the mentioned

9. Point out the wrong combination.
a) True negative=correctly rejected
b) False negative=correctly rejected
c) False positive=correctly identified
d) All of the mentioned

10. Which of the following is a common error measure?
a) Sensitivity
b) Median absolute deviation
c) Specificity
d) All of the mentioned

11. Predictive analytics is same as forecasting.
a) True
b) False

12 Maximum aposteriori classifier is also known as:
A. Decision tree classifier
B. Bayes classifier
C. Gaussian classifier
D. Maximum margin classifier

13. Which of the following is an example of continuous attribute?
A. Weight of a person
B. Shoe size of a person
C. Gender of a person
D. None of the above

14.Rows of a data matrix storing record data usually represents?
A. Metadata
B. Objects

C. Attributes
D. Aggregates

15. Sales database of items in a supermarket can be considered as an example of:
A. Record data
B. Tree data
C. Graph data
D. None of the above

16.User rating given to a movie in a scale 1-10, can be considered as an attribute of type?
A. Nominal
B. Ordinal
C. Interval

D. Ratio

17.. Name of a movie, can be considered as an attribute of type?
A. Nominal
B. Ordinal
C. Interval
D. Ratio

18 Average squared difference between classifier predicted output and actual output is

A. Mean Squared error
B. Root mean squared error
C. Mean absolute error
D. Mean Relative error

19. Machine learning differs from statistical techniques in that machine learning methods

A. Typically assume an underlying distribution of data
B.are better able to deal with missing and noisy data
C. are not able to explain their behavior
D. None of the above.

20.Suppose a model is overfitting,which is not a valid way to reduce overfitting

A. Increase the amount of training data
B. Decrease the model complexity
C. Reduce the noise of training data
D. Improve the optimization algorithm being used for error minimization

21.WHICH OF THE FOLLOWING STATEMENTS IS/ARE TRUE ABOUT “TYPE-1” AND “TYPE-2” ERRORS?

1 TYPE1 IS KNOWN AS FALSE POSITIVE AND TYPE2 IS KNOWN AS FALSE NEGATIVE.
2 TYPE1 IS KNOWN AS FALSE NEGATIVE AND TYPE2 IS KNOWN AS FALSE POSITIVE.
3 TYPE1 ERROR OCCURS WHEN WE REJECT A NULL HYPOTHESIS WHEN IT IS ACTUALLY

A 1 and 2
B 1 and 3
C 1

22.How many types are available in machine learning
A 2
B 4
C 1
D 3

23.Which of the following is the model used for learning:
A Neural networks
B Prepositional and FOL rules
C All of the above
D Decision tree

24.Automated vehicle is an example of

A Unsupervised learning
B None of above
C Active learning
D Supervised learning

25. Which Statement is true about prediction problems
A. Only one independent variable
B. More than one independent variable
C. . More than one dependent variable
D. None of the above

26. Adaptive system

1. Uses machine-learning techniques in which program can learn from past experience and adapt themselves to new situations.
2. Is a computational procedure that takes some value as input and produces some value as output.
3. Is a science of making machines performs tasks that would require intelligence when performed by humans
4. None of these

ANSWER: 1

27. In the representation of machine learning algorithm with Y = f(X) + e, e represents ;

1. Reducible error specifying, model not having enough attributes to sufficiently characterize the best mapping from X to Y.
2. Irreducible error specifying, model not having enough attributes to sufficiently characterize the best mapping from X to Y.
3. Propagation error specifying, model not having enough attributes to sufficiently characterize the best mapping from X to Y.
4. Transmission error specifying, model not having enough attributes to sufficiently characterize the best mapping from X to Y.

ANSWER: 2

28. Which of the following is a supervised learning problem? (multiple options may be correct)

1. Predicting credit approval based on historical data
2. Grouping people in a social network.
3. Predicting the gender of a person from his/her image. You are given the data of 1 Million

4. Images along the gender.
5. Given the class labels of old news articles, predicting the class of a new news article from its content. Class of a news article can be such as sports, politics, technology, etc.

ANSWER: (1), (3), (4)

29. Which of the following are classification problems? (multiple options may be correct)
    1. Predicting the temperature (in Celsius) of a room from other environmental features (such as atmospheric pressure, humidity etc).
    2. Predicting if a cricket player is a batsman or bowler given his playing records.
    3. Finding the shorter route between two existing routes between two points.
    4. Predicting if a particular route between two points has traffic jam or not based on the
    5. travel time of vehicles.
    6. Filtering of spam messages

ANSWER : (2),(4), (5)

30. Which of the following is an unsupervised task?
    1. Learning to play chess.
    2. Predicting if an edible item is sweet or spicy based on the information of the ingredients and their quantities.
    3. Grouping related documents from an unannotated corpus.
    4. all of the above

ANSWER : (3)

31. Which of the following are true about bias and variance of overfitted and underfitted models? (multiple options may be correct)
    1. Underfitted models have high bias.
    2. Underfitted models have low bias.
    3. Overfitted models have high variance.
    4. Overfitted models have low variance.
    5. none of these

ANSWER : (1), (3)

32. Which of the following is a categorical feature?
    1. Number of legs of an animal
    2. Number of hours you study in a day
    3. Branch of an engineering student
    4. Your weekly expenditure in rupees.
    5. Ethnicity of a person
    6. Height of a person in inches

ANSWER : (3) and (5)

33. Which of the following is a regression task? (multiple options may be correct)
    1. Predicting the monthly sales of a cloth store in rupees.
    2. Predicting if a user would like to listen to a newly released song or not based on historical data.
    3. Predicting the confirmation probability (in fraction) of your train ticket whose current status is waiting list based on historical data.
    4. Predicting if a patient has diabetes or not based on historical medical records.
    5. Predicting the gender of a human

ANSWER : (1) and (3)

34. What happens when your model complexity increases?
    1. Model bias increases
    2. Model bias decreases
    3. Variance of the model increases
    4. Variance of the model decreases

ANSWER : 1

35. Supervised learning problems can be further grouped into
    1. Regression and classification problems
    2. Clustering and association problems
    3. Both of the mentioned
    4. None of the mentioned

ANSWER : 1

36. Unsupervised learning problems can be further grouped into
    1. Regression and classification problems
    2. Clustering and association problems.
    3. Both of the mentioned
    4. None of the mentioned

ANSWER : 2

37. Problems where you have a large amount of input data (X) and only some of the data is labeled (Y) are called
    1. Un supervised learning problems
    2. Supervised learning problems.
    3. Semi supervised learning problems
    4. None of the mentioned

ANSWER : 3

38. Reinforcement learning is based on _______________ provided by the environment. This ___________________is usually called reward .
    1. Positive feedback, Feedback
    2. Feedback, Positive feedback
    3. Feedback, Negative feedback

ANSWER : 2

39. Machine learning concerned with algorithms inspired by the structure and function of the brain called
    1. Neural networks.
    2. Deep neural networks.
    3. Artificial neural networks.
    4. Option 1 & 2

ANSWER : 3

40. __________________ are typically feedforward networks in which data flows from the input layer to the output layer without looping back.

1. Convolutional Neural Networks
2. Deep Neural Network
3. Recurrent Neural Networks
4. Artificial Neural Networks

ANSWER : 2

41. __________________ are the networks in which data can flow in any direction and are used for applications for language purposes.

1. Convolutional Neural Networks
2. Deep Neural Network
3. Recurrent Neural Networks
4. Artificial Neural Networks

ANSWER : 3

42. Common Deep learning applications include:

1. Real-time visual tracking
2. Logistic optimization
3. Bioinformatics, Speech recognition
4. Only 1 & 3
5. All of the mentioned

ANSWER : 5

43. The most widely used metrics and tools to assess a classification model are:

1. Confusion matrix
2. Cost-sensitive accuracy
3. Area under the ROC curve
4. All of the above

ANSWER :4

44. Which of the following is a good test dataset characteristic?

1. Large enough to yield meaningful results
2. Is representative of the dataset as a whole
3. Both A and B
4. None of the above

ANSWER : 3

45. How do you handle missing or corrupted data in a dataset?

1. Drop missing rows or columns
2. Replace missing values with mean/median/mode
3. Assign a unique category to missing values
4. All of the above

ANSWER : 4

46. Choose the options that are correct regarding machine learning (ML) and artificial intelligence (AI),

1. ML is an alternate way of programming intelligent machines.
2. ML and AI have very different goals.
3. ML is a set of techniques that turns a dataset into a software.
4. AI is a software that can emulate the human mind.

Answer: (1), (3), (4)

47. Unsupervised learning is where you only have input data (X) and

1. No corresponding output variables.
2. No corresponding model to define the set of variables
3. No corresponding training dataset.
4. None of the above.

ANSWER: 1

48. Self-organizing maps are the examples of

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning
4. Missing data imputation

ANSWER: 2

49. Task of inferring a model from the labeled training data is called as :

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning
4. Missing data imputation

ANSWER: 1

50. A classification problem is when the output variable

1. Is a real value, such as “dollars” or “weight”.
2. Discovers the inherent groupings in the data.
3. Is a category, such as “red” or “blue” or “disease” and “no disease”
4. Discovers rules that describe large portions of your data.

ANSWER: 3

Subject: Machine Learning (BE-A&B)

UNIT II- SYLLABUS : Feature Selection

Scikit- learn Dataset, Creating training and test sets, managing categorical data, Managing missing features, Data scaling and normalization.

Feature selection and Filtering, Principle Component Analysis(PCA)-non negative matrix factorization, Sparse PCA, Kernel PCA. Atom Extraction and Dictionary Learning.

1. Tick the odd one out
    A. Scikit-learn dataset provide some built-in datasets that can be used for testing purposes.
    B. Built-in datasets are available in the package sklearn.datasets and have a common structure
    C. scikit-learn comes with a few standard datasets, for instance the iris and digits datasets for regression and the boston house prices dataset for classification
    D. Data instance variable contains the whole input set X while target contains the labels for classification or target values for regression.

ANSWER: C

2. Arrange in correct sequence
    A. Loading the dataset.
    B. Installing the Python and SciPy platform.-Download Miniconda
    C. Summarizing the dataset.
    D. Evaluating some algorithms.
    E. Making some predictions
    F. Visualizing the dataset.

ANSWER: B-A-C-F-D-E

3. Requirements for working with data in scikit-learn
    A. Features and response are separate objects
    B. Features and response should be numeric
    C. Features and response should have specific shapes
    D. All of the options mentioned
    E. Only option 1to 3

ANSWER: D

4. Correct project name for project name for scikit learn dataset
    A. scikit-learn
    B. scikit_learn
    C. SciKit learn
    D. sci-kit learn.

ANSWER: A

5. **ndarray is [TICK THE CORRECT OPTION]**
    A. a fast and space-efficient multidimensional array providing vectorized arithmetic operation
    B. a generic multidimensional container for homogeneous data that is all of the elements must be the same type.
    C. Every array has a shape a tuple indicating the size of each dimension, and a dtype, an object describing the *data type* of the array

ANSWER: A,B,C

6. When a dataset is large enough, it's a good practice to split it into training and test sets by rules such as (tick the odd one out)
    A. Both datasets must reflect the original distribution
    B. The original dataset must be randomly shuffled before the split phase in order to avoid a correlation between consequent elements
    C. Percentage of elements to put into the test/training set can be of a ratio as 90 percent for training and 10 percent for the test phase.
    D. All of the mentioned

ANSWER: C

7. To convert categorical features to such integer codes, we can use the
    E. OrdinalEncoder
    F. LabelEncoder
    G. Both A and B can be used alternatively
    H. None of the above

ANSWER: A

8. LabelEncoder can be used to normalize labels
    A. TRUE
    B. FALSE

ANSWER: A

9 Different ways of encoding categorical features are [tick the odd one out]
    A. OneHotEncoder
    B. DictVectorizer
    C. Pandas get_dummies
    D. OrdinalEncoder

ANSWER: D

10. One-Hot Encoding is
    A. Applied for categorical variables where no ordinal relationship exists
    B. Applied where the integer encoding is not enough.
    C. Used to add new binary variable for each unique integer value.
    D. None of the options mentioned
    E. All of the options mentioned

ANSWER: E

11. A dataset can contain missing features, with the options to be considered as:
    A. Removing the whole line- dataset is quite large, the number of missing features is high, and any prediction could be risky.
    B. Creating sub-model to predict those features-more difficult because it's necessary to determine a supervised strategy to train a model for each feature and, finally, to predict their value.
    C. Using an automatic strategy to input them according to the other known values-likely to be the best choice.
    D. Option 1 & 2 only
    E. All of the options mentioned

ANSWER: E

12. The command to create an environment in skilearn with a specific version of Python is
   A. $ conda create -n myenv python=3.4
   B. $ conda create -n myenv scipy
   C. $ conda create -n myenv scipy=0.15.0
   D. None of the mentioned

ANSWER: A

13. To create an environment in skilearn with a specific package:
   A. $ conda create -n myenv python=3.4
   B. $ conda create -n myenv scipy
   C. $ conda create -n myenv scipy=0.15.0
   D. None of the mentioned

ANSWER: B

14. To create an environment with a specific version of a package in skilearn
   A. $ conda create -n myenv python=3.4
   B. $ conda create -n myenv scipy
   C. $ conda create -n myenv scipy=0.15.0
   D. None of the mentioned

ANSWER: C

15. Command to load boston dataset in python is
   A. from sklearn.datasets import load_boston
   B. from sklear_datasets import load_boston
   C. from sklearn.datasets.import _load_boston
   D. All of the above

ANSWER: A

16 .Which of the following are true about forward subset selection?
O(2d) models must be trained during the algorithm, where d is the number of features
It greedily adds the feature that most improves cross-validation accuracy
It finds the subset of features that give the lowest test error
Forward selection is faster than backward selection if few features are relevant to prediction

17. Principal component analysis (PCA) can be used with variables of any mathematical types: quantitative, qualitative, or a mixture of these types. – True, False

18. Principal component analysis (PCA) requires quantitative multivariate data. – True, False.

19. For variables with physical dimensions (e.g. kg), their variances also have physical dimensions. – True, False

20. The variables subjected to PCA must all have the same physical dimensions. – True, False.

21. When the variables have different physical dimensions, they must be made dimensionless by standardization or ranging before PCA. – True, False.

22. Which of the following is the second goal of PCA?
a) data compression
b) statistical analysis
c) data dredging
d) all of the mentioned
Answer: a

23.T or F The goal of PCA is to interpret the underlying structure of the data in terms of the principal components that are best at predicting the output variable.

24 T or F The output of PCA is a new representation of the data that is always of lower dimensionality than the original feature representation.

25. In principal component analysis, a smaller eigenvalue indicates that
A. A given variable in the original data set, say Xj, is more important
B. A given variable in the original data set, say Xj, is less important
C. A given principal component, say Yj, is more important
D. A given principal component, say Yj, is less important

26. Why do we often pick just the first two principal components?
A. Because we can graph them in a scatterplot
B. Because they explain most of the variance
C. Because they are uncorrelated
D. Because of the Kaiser criterion

27. Taking a bootstrap sample of n data points in p dimensions means:

A. Sampling p features with replacement.
B. Sampling √p features without replacement.
C. Sampling n samples with replacement.
D. Sampling k < n samples without replacement

28.PCA properties are

A. Unsupervised dimensionality reduction
B. Linear representation that gives best squared error fit
C. No local minima (exact)
D. Orthogonal vectors

29. NMF properties are

A. Unsupervised dimensionality reduction
B. Non-negative coefficients
C. Iterative (the presented algorithm)
D. “Parts-based”; easier to interpret

30. Kernel PCA applications including

A. denoising,
B. compression
C. structured prediction

31. Which of the following is an example of a deterministic algorithm?
A) PCA
B) K-Means
C) None of the above

32. [True or False] A Pearson correlation between two variables is zero but, still their values can still be related to each other.

A) TRUE

B) FALSE

33. What would you do in PCA to get the same projection as SVD?

A) Transform data to zero mean

B) Transform data to zero median

C) Not possible

D) None of these

34. Which of the folllowing is an example of feature extraction?

A all of the above

B Constructing bag of words vector from an email

C Applying PCA projects to a large high-dimensional data

D Removing stopwords in a sentence

35.When performing regression or classification, which of the following is the correct way to preprocess the data?

A none of the above

B Normalize the data → PCA → normalize PCA output → training

C Normalize the data → PCA → training

D PCA → normalize PCA output → training

36. Which of the following techniques would perform better for reducing dimensions of a data set?

A.none of these

B Removing columns with dissimilar data trends

CRemoving columns which have high variance in data

D Removing columns which have too many missing values

37.Dimensionality reduction algorithms are one of the possible ways to reduce the computation time required to build a model.

TRUE/False

38. Which of the following is/are one of the important step(s) to pre-process the text in NLP based projects?

1 STEMMING 2 STOP WORD REMOVAL 3 OBJECT STANDARDIZATION

A: 1,2

B1,2,3

C 2,3

D 1,3

39. Which of the following is an example of a deterministic algorithm?

A PCA

B none of the above

C K means

40. What is pca.components_ in Sklearn?

A None of the above

B Result of the multiplication matrix

C Set of all eigen vectors for the projection space

D Matrix of principal components

Short Question and Answer

1 Explain Principal Component Analysis (PCA).
PCA is a dimensionality-reduction technique which mathematically transforms a set of correlated variables into a smaller set of uncorrelated variables called principal components.

2.Explain Principal Component Analysis (PCA).
PCA is a dimensionality-reduction technique which mathematically transforms a set of correlated variables into a smaller set of uncorrelated variables called principal components.

3 What is an Incremental Learning algorithm in the ensemble?
The incremental learning method is the ability of an algorithm to learn from new data that may be available after the classifier has already been generated from the already available dataset.

4. What are PCA, KPCA, and ICA used for?
PCA (Principal Components Analysis), KPCA (Kernel-based Principal Component Analysis) and ICA (Independent Component Analysis) are important feature extraction techniques used for dimensionality reduction.

5. What is dimension reduction in Machine Learning?
In Machine Learning and statistics, dimension reduction is the process of reducing the number of random variables under considerations and can be divided into feature selection and feature extraction

6 How to Standardize data?
Standardization is the method that is used for rescaling data attributes. The attributes would likely have a value of mean as **0** and the value of standard deviation as **1**. The main objective of standardization is to prompt the mean and standard deviation for the attributes.

7. **what is data normalization and why do we need it?**
. Data normalization is a very important preprocessing step, used to rescale values to fit in a specific range to assure better convergence during backpropagation. In general, it boils down to subtracting the mean of each data point and dividing by its standard deviation. If we don't do this then some of the features (those with high magnitude) will be weighted more in the cost function (if a higher-magnitude feature changes by 1%, then that change is pretty big, but for smaller features it's quite insignificant). The data normalization makes all features weighted equally.

8. explain dimensionality reduction, where it's used, and its benefits?
Dimensionality reduction is the process of reducing the number of feature variables under consideration by obtaining a set of principal variables which are basically the important features. Importance of a feature depends on how much the feature variable contributes to the information representation of the data and depends on which technique you decide to use. Deciding which technique to use comes down to trial-and-error and preference. It's common to start with a linear technique and move to non-linear techniques when results suggest inadequate fit. Benefits of dimensionality reduction for a data set may be:

9. how do you handle missing or corrupted data in a dataset?
You could find missing/corrupted data in a dataset and either drop those rows or columns, or decide to replace them with another value. In Pandas, there are two very useful methods: isnull() and dropna() that will help you find columns of data with missing or corrupted data and drop those values. If you want to fill the invalid values with a placeholder value (for example, 0), you could use the fillna() method.

10 How are NumPy and SciPy related?

- NumPy is part of SciPy.
- NumPy defines arrays along with some basic numerical functions like indexing, sorting, reshaping, etc.
- SciPy implements computations such as numerical integration, optimization and machine learning using NumPy's functionality.

11. Is rotation necessary in PCA? If yes, Why? What will happen if you don't rotate the components?
Answer: Yes, rotation (orthogonal) is necessary because it maximizes the difference between variance captured by the component. This makes the components easier to interpret. Not to forget, that's the motive of doing PCA where, we aim to select fewer components (than features) which can explain the maximum variance in the data set. By doing rotation, the relative location of the components doesn't change, it only changes the actual coordinates of the points.
If we don't rotate the components, the effect of PCA will diminish and we'll have to select more number of components to explain variance in the data set.

12. **While working on a data set, how do you select important variables? Explain your methods.**
Answer: Following are the methods of variable selection you can use:
1. Remove the correlated variables prior to selecting important variables
2. Use linear regression and select variables based on p values
3. Use Forward Selection, Backward Selection, Stepwise Selection
4. Use Random Forest, Xgboost and plot variable importance chart
5. Use Lasso Regression
6. Measure information gain for the available set of features and select top n features accordingly.

13 **We know that one hot encoding increasing the dimensionality of a data set. But, label encoding doesn't. How ?**
Answer: Using one hot encoding, the dimensionality (a.k.a features) in a data set get increased because it creates a new variable for each level present in categorical variables. For example: let's say we have a variable 'color'. The variable has 3 levels namely Red, Blue and Green. One hot encoding 'color' variable will generate three new variables as Color.Red, Color.Blue and Color.Green containing 0 and 1 value.
In label encoding, the levels of a categorical variables gets encoded as 0 and 1, so no new variable is created. Label encoding is majorly used for binary variables.

14.. What do you understand by Eigenvectors and Eigenvalues?
- **Eigenvectors:** *Eigenvectors are those vectors whose direction remains unchanged even when a linear transformation is performed on them.*
- **Eigenvalues:** *Eigenvalue is the scalar that is used for the transformation of an Eigenvector.*

$$\begin{bmatrix} 3 & 4 & -2 \\ 1 & 4 & -1 \\ 2 & 6 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 4 & -2 \\ 1 & 4 & -1 \\ 2 & 6 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

Eigenvalue & Eigenvectors
In the above example, 3 is an Eigenvalue, with the original vector in the multiplication problem being an eigenvector.

15 Name a few libraries in Python used for Data Analysis and Scientific Computations.
Here is a list of Python libraries mainly used for Data Analysis:
- NumPy
- SciPy
- Pandas
- SciKit
- Matplotlib
- Seaborn
- Bokeh

Subject: Machine Learning (BE-A&B)

UNIT III- SYLLABUS : Regression
Linear regression- Linear models, A bi-dimensional example, Linear Regression and higher dimensionality, Ridge, Lasso and ElasticNet, Robust regression with random sample consensus, Polynomial regression, Isotonic regression.
Logistic regression-Linear classification, Logistic regression, Implementation and Optimizations, Stochastic gradient descendent algorithms, Finding the optimal hyper-parameters through grid search, Classification metric, ROC Curve.
Home

SECTION A: MCQS

1. Which one of the statement is true regarding residuals in regression analysis?
   A Mean of residuals is always zero
   B Mean of residuals is always less than zero
   C Mean of residuals is always greater than zero
   D There is no such rule for residuals.

ANSWER: A

2. The correlation coefficient is used to determine:
   A. A specific value of the y-variable given a specific value of the x-variable
   B. A specific value of the x-variable given a specific value of the y-variable
   C. The strength of the relationship between the x and y variables
   D. All of the above

ANSWER: C

3. To test linear relationship of y(dependent) and x(independent) continuous variables, which of the following plot best suited?
   A Scatter plot
   B Barchart
   C Histograms
   D All of the above

ANSWER: A

4. Which of the following method(s) does not have closed form solution for its coefficients?
   A Ridge regression
   B Lasso
   C Both Ridge and Lasso
   D None of both

ANSWER: B

5. Suppose we fit “Lasso Regression” to a data set, which has 100 features (X1,X2…X100). Now, we rescale one of these feature by multiplying with 10 (say that feature is X1), and then refit Lasso regression with the same regularization parameter. Now, which of the following option will be correct?
   A It is more likely for X1 to be excluded from the model
   B It is more likely for X1 to be included in the model
   C Can’t say
   D All of these

ANSWER: B

6. In a simple linear regression model (One independent variable), If we change the input variable by 1 unit. How much output variable will change?
   A By 1
   B No change
   C By its Slope
   D None of the above

ANSWER: C

7. It is possible to design a Linear regression algorithm using a neural network?

   A. TRUE
   B. FALSE

ANSWER: A

8. What will happen when you apply very large penalty in case of Lasso?
   A. Some of the coefficient will become zero
   B. Some of the coefficient will be approaching to zero but not absolute zero
   C. Both A and B depending on the situation
   D. None of these

ANSWER: A

9. Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and you found that there is a relationship between them. Which of the following conclusion do you make about this situation?
   A. Since the there is a relationship means our model is not good
   B. Since the there is a relationship means our model is good
   C. Can't say
   D. None of these

ANSWER: A

10. If the slope of the regression equation y = bo + b1x is positive, then;
   A. as x increases y decreases
   B. as x increases so does y
   C. Either a or b is correct
   D. as x decreases y increases

ANSWER: B

11. Ridge regression uses __________regularization which adds the penalty term to the OLS equation
   A. L2
   B. L1
   C. Uses L1 and L2 alternatively
   D. None of the mentioned

ANSWER: A

12. TICK THE ODD ONE OUT
   A. Linear regression is a linear model which assumes a linear relationship between the input variables (x) and the single output variable (y).
   B. In linear regression y can be calculated from a linear combination of the input variables (x).

C. The regression analysis helps us to understand how much the dependent variable changes with a change in one or more independent variables.
D. Linear regression is similar to Kernal PCA.

ANSWER: A

13. Match the following

1. Simple linear regression	a. 1 dependent variable (dichotomous), 2+ independent variable(s) (interval or ratio or dichotomous)
2. Multiple linear regression	b. 1 dependent variable (interval or ratio), 1 independent variable (interval or ratio or dichotomous)
3. Logistic regression	c. 1 dependent variable (nominal), 1+ independent variable(s) (interval or ratio or dichotomous)
4. Multinomial regression	d. 1 dependent variable (interval or ratio) , 2+ independent variables (interval or ratio or dichotomous)

A. 1-b, 2-a, 3-d, 4-c
B. 1-d, 2-b, 3-a, 4-c
C. 1-b, 2-d, 3-a, 4-c
D. 1-c, 2-d, 3-a, 4-b

ANSWER: C

14. Let's say, a "Linear regression" model perfectly fits the training data (train error is zero). Which of the following statement is true?

A. You will always have test error zero
B. You cannot have test error zero
C. None of the above

ANSWER: C

15. What would be then consequences for the OLS estimator if heteroscedasticity is present in a regression model but ignored?

A. It will be biased
B. It will be inconsistent
C. It will be inefficient
D. All of the options mentioned

ANSWER: C

16. TICK THE ODD ONE OUT

A. Multiple linear regression requires at least two independent variables, which can be nominal, ordinal, or interval/ratio level variables.
B. Multiple linear regression requires the relationship between the independent and dependent variables to be linear.
C. Multiple linear regression analysis requires that the errors between observed and predicted values should be normally distributed.
D. Multiple linear regression assumes that there is multicollinearity in the data

ANSWER: D

17. Multicollinearity may be checked multiple as

A. Using correlation matrix
B. Using Variance Inflation Factor (VIF)
C. By centering the data.
D. Option 1 and 2 only
E. All of the options mentioned

ANSWER: E

18. To check the accuracy of a regression, scikit-learn provides ______________ function which evaluates the model on test data.
   A. lr.score(X_test, Y_test)
   B. lr.accute(X_test, Y_test)
   C. lr.accuracy(X_test, Y_test)
   D. lr.acc.reg(X_test, Y_test)

ANSWER: A

19. If the points on the scatter diagram indicate that as one variable increases the other variable tends to decrease the value of r will be:
   A. Perfect positive
   B. Perfect negative
   C. Negative
   D. Zero

ANSWER: C

20. General equation of a polynomial regression is
   A. $Y=\theta_0 + \theta_1 X + \theta_2 X^2 + \ldots + \theta_m X^m$ + transmission error
   B. $Y=\theta_0 + \theta_1 X + \theta_2 X^2 + \ldots + \theta_m X^m$ + residual error
   C. $Y=\theta_0 + \theta_1 X + \theta_2 X^2 + \ldots + \theta_m X^m$ + gross error
   D. None of the options mentioned

ANSWER: B

21. Ridge Regression performs L2 regularization which
   A. Adds penalty equivalent to the magnitude of coefficients
   B. Adds penalty equivalent to squareroot of the magnitude of coefficients
   C. Adds penalty equivalent to square of the magnitude of coefficients
   D. Adds penalty equivalent to cube root of the magnitude of coefficients

ANSWER: C

22. **ElasticNet**, which combines both Lasso and Ridge into a single model with two penalty factors, one proportional to $L1$ norm and the other to $L2$ norm.
   A. TRUE
   B. FALSE

ANSWER: A

23. L1 Regularization which is ___________ adds regularization terms in the model which are function of ______________________of the coefficients of parameters.

   A. Lasso Regularization, absolute value
   B. Ridge Regularization , square
   C. Lasso Regularization , square
   D. Ridge Regularization , absolute value

ANSWER: A

24. Steps followed in RANSAC algorithm

a. Select a random subset of the original data. Call this subset the hypothetical inliers.
b. All other data are then tested against the fitted model. Those points that fit the estimated model well, according to some model-specific loss function, are considered as part of the consensus set.
c. A model is fitted to the set of hypothetical inliers.
d. The estimated model is reasonably good if sufficiently many points have been classified as part of the consensus set.
e. Afterwards, the model may be improved by reestimating it using all members of the consensus set.

ANSWER: a-c-b-d-e

25. TICK THE ODD ONE OUT

A. The isotonic regression finds a non-decreasing approximation of a function while minimizing the mean squared error on the training data.
B. The benefit of isotonic regression model is that it does not assume any form for the target function such as linearity.
C. Isotonic regression equation becomes polynomial regression equation if the power of independent variable is more than 1.
D. Isotonic regression produces a piecewise interpolating functions minimizing the functional.

ANSWER: C

1) True-False: Is Logistic regression a supervised machine learning algorithm?
A) TRUE
B) FALSE

2) True-False: Is Logistic regression mainly used for Regression?
A) TRUE
B) FALSE

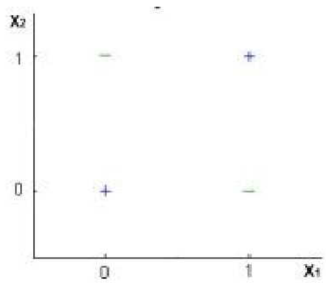
3) True-False: Is it possible to design a logistic regression algorithm using a Neural Network Algorithm?
A) TRUE
B) FALSE

4) True-False: Is it possible to apply a logistic regression algorithm on a 3-class Classification problem?
A) TRUE
B) FALSE

5) [True-False] Standardisation of features is required before training a Logistic Regression.
A) TRUE
B) FALSE
ANSWER: B

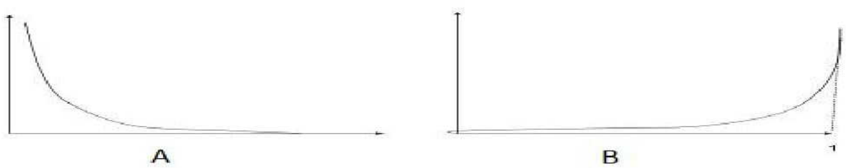
6) Choose which of the following options is true regarding One-Vs-All method in Logistic Regression.
A) We need to fit n models in n-class classification problem
B) We need to fit n-1 models to classify into n classes
C) We need to fit only 1 model to classify into n classes
D) None of these
ANSWER: **A**

7) Can a Logistic Regression classifier do a perfect classification on the below data?



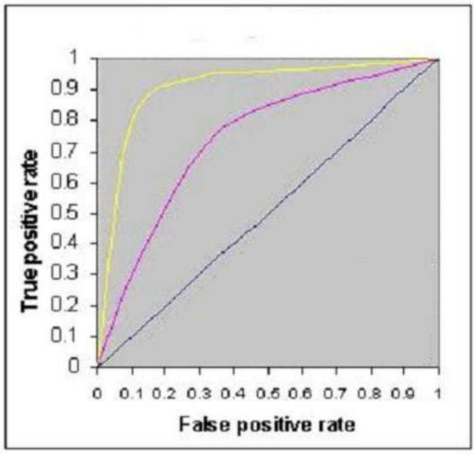
Note: You can use only X1 and X2 variables where X1 and X2 can take only two binary values(0,1).
A) TRUE
B) FALSE
C) Can't say
D) None of these
ANSWER: **B**

8) Which of the following image is showing the cost function for y =1.
Following is the loss function in logistic regression(Y-axis loss function and x axis log probability) for two class classification problem.
Note: Y is the target class



A) A
B) B
C) Both
D) None of these
ANSWER: A

9) The below figure shows AUC-ROC curves for three logistic regression models. Different colors show curves for different hyper parameters values. Which of the following AUC-ROC will give best result?



A) Yellow
B) Pink
C) Black
D) All are same
ANSWER: A
Q10. Generally, which of the following method(s) is used for predicting continuous dependent variable?

1. **Linear Regression**
2. **Logistic Regression**

A. 1 and 2
B. only 1
C. only 2
D. None of these.
ANSWER: B

Q11. Suppose I applied a logistic regression model on data and got training accuracy X and testing accuracy Y. Now I want to add few new features in data. Select option(s) which are correct in such case. Note: Consider remaining parameters are same.

1. **Training accuracy always decreases.**
2. **Training accuracy always increases or remain same.**
3. **Testing accuracy always decreases**
4. **Testing accuracy always increases or remain same**

A. Only 2
B. Only 1
C. Only 3
D. Only 4
ANSWER: A

Q12. Suppose, we are using Logistic regression model for n-class classification problem. In this case, we can use One-vs-rest method. Choose which of the following option is true regarding this?
A. We need to fit n model in n-class classification problem.
B. We need to fit n-1 models to classify into n classes.
C. We need to fit only 1 model to classify into n classes.
D. None of these.
ANSWER: A

13) Which of the following hyper parameter(s), when increased may cause random forest to over fit the data?

1. **Number of Trees**
2. **Depth of Tree**
3. **Learning Rate**

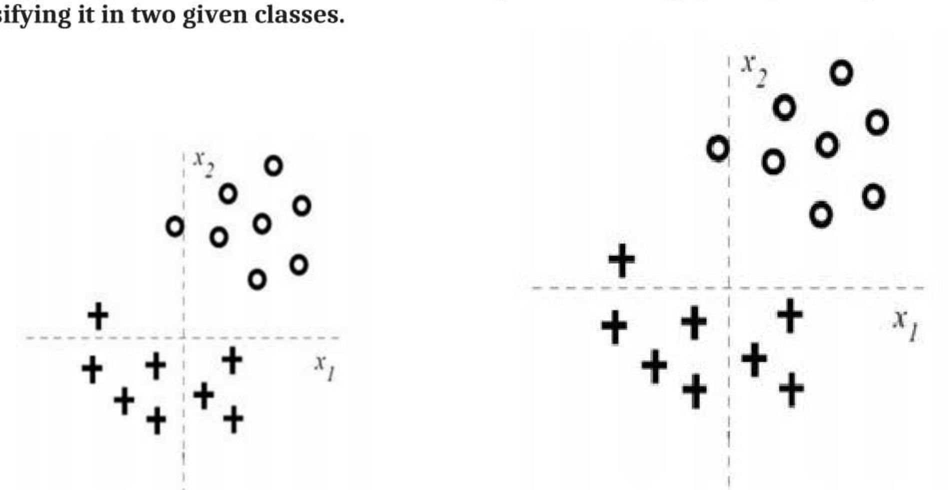
A) Only 1
B) Only 2
C) Only 3
D) 1 and 2
E) 2 and 3
F) 1,2 and 3
ANSWER: (B)
Usually, if we increase the depth of tree it will cause overfitting. Learning rate is not an hyperparameter in random forest. Increase in the number of tree will cause under fitting.

14) Imagine, you are solving a classification problems with highly imbalanced class. The majority class is observed 99% of times in the training data.
Your model has 99% accuracy after taking the predictions on test data. Which of the following is true in such a case?

1. Accuracy metric is not a good idea for imbalanced class problems.
2. Accuracy metric is a good idea for imbalanced class problems.
3. Precision and recall metrics are good for imbalanced class problems.
4. Precision and recall metrics aren't good for imbalanced class problems.

A) 1 and 3
B) 1 and 4
C) 2 and 3
D) 2 and 4
ANSWER: (A)

15.Suppose you are given the below data and you want to apply a logistic regression model for classifying it in two given classes.



You are using logistic regression with L1 regularization.

$$\sum_{i=1}^{n} \log P(y_i|x_i, w_0, w_1, w_2) - C(|w_1| + |w_2|).$$

Where C is the regularization parameter and w1 & w2 are the coefficients of x1 and x2.
Which of the following option is correct when you increase the value of C from zero to a very large value?
A) First w2 becomes zero and then w1 becomes zero

B) First w1 becomes zero and then w2 becomes zero
C) Both becomes zero at the same time
D) Both cannot be zero even after very large value of C
ANSWER: (B)

16.If searching among a large number of hyperparameters, you should try values in a grid rather than random values, so that you can carry out the search more systematically and not rely on chance. True or False?
FALSE

17.Every hyperparameter, if set poorly, can have a huge negative impact on training, and so all hyperparameters are about equally important to tune well. True or False?
FALSE

18.During hyperparameter search, whether you try to babysit one model ("Panda" strategy) or train a lot of models in parallel ("Caviar") is largely determined by:

- **A.** Whether you use batch or mini-batch optimization
- **B.** The presence of local minima (and saddle points) in your neural network
- **C.** The amount of computational power you can access
- **D.** The number of hyperparameters you have to tune

19.Finding good hyperparameter values is very time-consuming. So typically you should do it once at the start of the project, and try to find very good hyperparameters so that you don't ever have to revisit tuning them again. True or false?
FALSE

20.In the normalization formula, why do we use epsilon?
To avoid division by zero. True or false?
TRUE

SECTION B:Short question and answer

1. What is linear regression?

In simple terms, linear regression is a method of finding the best straight line fitting to the given data, i.e. finding the best linear relationship between the independent and dependent variables. In technical terms, linear regression is a machine learning algorithm that finds the best linear-fit relationship on any given data, between independent and dependent variables. It is mostly done by the Sum of Squared Residuals Method.

2. State the assumptions in a linear regression model.

There are three main assumptions in a linear regression model:

1. The assumption about the form of the model:

   It is assumed that there is a linear relationship between the dependent and independent variables. It is known as the 'linearity assumption'.
2. Assumptions about the residuals:

a. Normality assumption: It is assumed that the error terms, $\varepsilon^{(i)}$, are normally distributed.

b. Zero mean assumption: It is assumed that the residuals have a mean value of zero.

c. Constant variance assumption: It is assumed that the residual terms have the same (but unknown) variance, σ^2 This assumption is also known as the assumption of homogeneity or homoscedasticity.

d. Independent error assumption: It is assumed that the residual terms are independent of each other, i.e. their pair-wise covariance is zero.

3. Assumptions about the estimators:

   a. The independent variables are measured without error.

   b. The independent variables are linearly independent of each other, i.e. there is no multicollinearity in the data.

Explanation:

1. This is self-explanatory.
2. If the residuals are not normally distributed, their randomness is lost, which implies that the model is not able to explain the relation in the data.

   Also, the mean of the residuals should be zero.

   $Y^{(i)i}= \beta_0+ \beta_1 x^{(i)} + \varepsilon^{(i)}$

   This is the assumed linear model, where ε is the residual term.

   $E(Y) = E(\beta_0+ \beta_1 x^{(i)} + \varepsilon^{(i)})$

   $\quad = E(\beta_0+ \beta_1 x^{(i)} + \varepsilon^{(i)})$

   If the expectation(mean) of residuals, $E(\varepsilon^{(i)})$, is zero, the expectations of the target variable and the model become the same, which is one of the targets of the model.

   The residuals (also known as error terms) should be independent. This means that there is no correlation between the residuals and the predicted values, or among the residuals themselves. If some correlation is present, it implies that there is some relation that the regression model is not able to identify.
3. If the independent variables are not linearly independent of each other, the uniqueness of the least squares solution (or normal equation solution) is lost.

3. What is feature engineering? How do you apply it in the process of modelling?

Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data.

In layman terms, feature engineering means the development of new features that may help you understand and model the problem in a better way. Feature engineering is of two kinds — business driven and data-driven. Business-driven feature engineering revolves around the inclusion of features from a business point of view. The job here is to transform the business variables into features of the problem. In case of data-driven feature engineering, the features you add do not have any significant physical interpretation, but they help the model in the prediction of the target variable.

To apply feature engineering, one must be fully acquainted with the dataset. This involves knowing what the given data is, what it signifies, what the raw features are, etc. You must also have a crystal clear idea of the problem, such as what factors affect the target variable, what the physical interpretation of the variable is, etc.

5 Breakthrough Applications of Machine Learning

4. What is the use of regularization? Explain L1 and L2 regularizations.

Regularization is a technique that is used to tackle the problem of overfitting of the model. When a very complex model is implemented on the training data, it overfits. At times, the simple model might not be able to generalise the data and the complex model overfits. To address this problem, regularisation is used. Regularisation is nothing but adding the coefficient terms (betas) to the cost function so that the terms are penalised and are small in magnitude. This essentially helps in capturing the trends in the data and at the same time prevents overfitting by not letting the model become too complex.

- L1 or LASSO regularisation: Here, the absolute values of the coefficients are added to the cost function. This can be seen in the following equation; the highlighted part corresponds to the L1 or LASSO regularisation. This regularisation technique gives sparse results, which lead to feature selection as well.

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

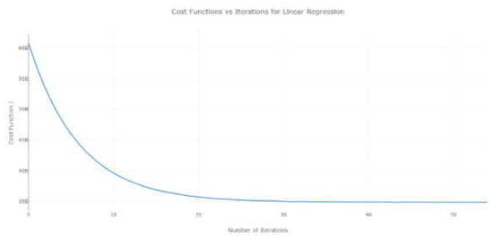
- L2 or Ridge regularisation: Here, the squares of the coefficients are added to the cost function. This can be seen in the following equation, where the highlighted part corresponds to the L2 or Ridge regularisation.

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

5. How to choose the value of the parameter learning rate (α)?

Selecting the value of learning rate is a tricky business. If the value is too small, the gradient descent algorithm takes ages to converge to the optimal solution. On the other hand, if the value of the learning rate is high, the gradient descent will overshoot the optimal solution and most likely never converge to the optimal solution.

To overcome this problem, you can try different values of alpha over a range of values and plot the cost vs the number of iterations. Then, based on the graphs, the value corresponding to the graph showing the rapid decrease can be chosen.



The aforementioned graph is an ideal cost vs the number of iterations curve. Note that the cost initially decreases as the number of iterations increases, but after certain iterations, the gradient descent converges and the cost does not decrease anymore.If you see that the cost is increasing with the number of iterations, your learning rate parameter is high and it needs to be decreased.

6. How to choose the value of the regularisation parameter (λ)?

Selecting the regularisation parameter is a tricky business. If the value of λ is too high, it will lead to extremely small values of the regression coefficient β, which will lead to the model underfitting (high bias – low variance). On the other hand, if the value of λ is 0 (very small), the model will tend to overfit the training data (low bias – high variance).

There is no proper way to select the value of λ. What you can do is have a sub-sample of data and run the algorithm multiple times on different sets. Here, the person has to decide how much variance can be tolerated. Once the user is satisfied with the variance, that value of λ can be chosen for the full dataset.

One thing to be noted is that the value of λ selected here was optimal for that subset, not for the entire training data.

7. Can we use linear regression for time series analysis?

One can use linear regression for time series analysis, but the results are not promising. So, it is generally not advisable to do so. The reasons behind this are —

a. Time series data is mostly used for the prediction of the future, but linear regression seldom gives good results for future prediction as it is not meant for extrapolation.
b. Mostly, time series data have a pattern, such as during peak hours, festive seasons, etc., which would most likely be treated as outliers in the linear regression analysis.

8. What value is the sum of the residuals of a linear regression close to? Justify.

Ans The sum of the residuals of a linear regression is 0. Linear regression works on the assumption that the errors (residuals) are normally distributed with a mean of 0, i.e.

$\mathbf{Y = \beta^T X + \varepsilon}$

Here, Y is the target or dependent variable,

β is the vector of the regression coefficient, X is the feature matrix containing all the features as the columns,

ε is the residual term such that $\varepsilon \sim N(0,\sigma^2)$.

So, the sum of all the residuals is the expected value of the residuals times the total number of data points. Since the expectation of residuals is 0, the sum of all the residual terms is zero.

Note: $N(\mu,\sigma^2)$ is the standard notation for a normal distribution having mean μ and standard deviation σ^2.

9. How does multicollinearity affect the linear regression?

Ans Multicollinearity occurs when some of the independent variables are highly correlated (positively or negatively) with each other. This multicollinearity causes a problem as it is against the basic assumption of linear regression. The presence of multicollinearity does not affect the predictive capability of the model. So, if you just want predictions, the presence of multicollinearity does not affect your output. However, if you want to draw some insights from the model and apply them in, let's say, some business model, it may cause problems.

One of the major problems caused by multicollinearity is that it leads to incorrect interpretations and provides wrong insights. The coefficients of linear regression suggest the mean change in the target value if a feature is changed by one unit. So, if multicollinearity exists, this does not hold true as changing one feature will lead to changes in the correlated variable and consequent changes in the target variable. This leads to wrong insights and can produce hazardous results for a business.

A highly effective way of dealing with multicollinearity is the use of VIF (Variance Inflation Factor). Higher the value of VIF for a feature, more linearly correlated is that feature. Simply remove the feature with very high VIF value and re-train the model on the remaining dataset.

10. What is the normal form (equation) of linear regression? When should it be preferred to the gradient descent method?

The normal equation for linear regression is —

$\beta=(X^TX)^{-1}.X^TY$

Here, $Y=\beta^TX$ is the model for the linear regression,

Y is the target or dependent variable,

β is the vector of the regression coefficient, which is arrived at using the normal equation,

X is the feature matrix containing all the features as the columns.

Note here that the first column in the X matrix consists of all 1s. This is to incorporate the offset value for the regression line.

Comparison between gradient descent and normal equation:

Gradient Descent	Normal Equation

Needs hyper-parameter tuning for alpha (learning parameter)	No such need
It is an iterative process	It is a non-iterative process
O(kn2) time complexity	O(n3) time complexity due to evaluation of XTX
Prefered when n is extremely large	Becomes quite slow for large values of n

Here, ‘k’ is the maximum number of iterations for gradient descent, and ‘n’ is the total number of data points in the training set.

Clearly, if we have large training data, normal equation is not prefered for use. For small values of ‘n’, normal equation is faster than gradient descent.

What is Machine Learning and Why it matters

11. You run your regression on different subsets of your data, and in each subset, the beta value for a certain variable varies wildly. What could be the issue here?

This case implies that the dataset is heterogeneous. So, to overcome this problem, the dataset should be clustered into different subsets, and then separate models should be built for each cluster. Another way to deal with this problem is to use non-parametric models, such as decision trees, which can deal with heterogeneous data quite efficiently.

12. Your linear regression doesn’t run and communicates that there is an infinite number of best estimates for the regression coefficients. What could be wrong?

This condition arises when there is a perfect correlation (positive or negative) between some variables. In this case, there is no unique value for the coefficients, and hence, the given condition arises.

13. What do you mean by adjusted R2? How is it different from R2?

Adjusted R^2, just like R^2, is a representative of the number of points lying around the regression line. That is, it shows how well the model is fitting the training data. The formula for adjusted R^2 is —

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1} \right]$$

Here, n is the number of data points, and k is the number of features. One drawback of R^2 is that it will always increase with the addition of a new feature, whether the new feature is useful or not. The adjusted R^2 overcomes this drawback. The value of the adjusted R^2 increases only if the newly added feature plays a significant role in the model.

14. How do you interpret the residual vs fitted value curve?

The residual vs fitted value plot is used to see whether the predicted values and residuals have a correlation or not. If the residuals are distributed normally, with a mean around the fitted value and a constant variance, our model is working fine; otherwise, there is some issue with the model. The most common problem that can be found when training the model over a large range of a dataset is heteroscedasticity(this is explained in the answer below). The presence of heteroscedasticity can be easily seen by plotting the residual vs fitted value curve.

15. What is heteroscedasticity? What are the consequences, and how can you overcome it?

A random variable is said to be heteroscedastic when different subpopulations have different variabilities (standard deviation).

The existence of heteroscedasticity gives rise to certain problems in the regression analysis as the assumption says that error terms are uncorrelated and, hence, the variance is constant. The presence of heteroscedasticity can often be seen in the form of a cone-like scatter plot for residual vs fitted values.

One of the basic assumptions of linear regression is that heteroscedasticity is not present in the data. Due to the violation of assumptions, the Ordinary Least Squares (OLS) estimators are not the Best Linear Unbiased Estimators (BLUE). Hence, they do not give the least variance than other Linear Unbiased Estimators (LUEs).

There is no fixed procedure to overcome heteroscedasticity. However, there are some ways that may lead to a reduction of heteroscedasticity. They are —

a. Logarithmising the data: A series that is increasing exponentially often results in increased variability. This can be overcome using the log transformation.
b. Using weighted linear regression: Here, the OLS method is applied to the weighted values of X and Y. One way is to attach weights directly related to the magnitude of the dependent variable.

16. What are parametric models? Give an example.

Parametric models are those with a finite number of parameters. To predict new data, you only need to know the parameters of the model. Examples include linear regression, logistic regression, and linear SVMs.
Non-parametric models are those with an unbounded number of parameters, allowing for more flexibility. To predict new data, you need to know the parameters of the model and the state of the data that has been observed. Examples include decision trees, k-nearest neighbors, and topic models using latent dirichlet analysis.

17. What is the "Curse of Dimensionality?"

The difficulty of searching through a ANSWER space becomes much harder as you have more features (dimensions).
Consider the analogy of looking for a penny in a line vs. a field vs. a building. The more dimensions you have, the higher volume of data you'll need.

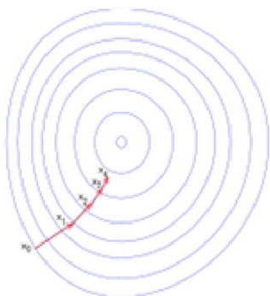
18. Explain the Bias-Variance Tradeoff.

Predictive models have a tradeoff between bias (how well the model fits the data) and variance (how much the model changes based on changes in the inputs).

Simpler models are stable (low variance) but they don't get close to the truth (high bias).

More *complex models* are more prone to being overfit (high variance) but they are expressive enough to get close to the truth (low bias).

The best model for a given problem usually lies somewhere in the middle.



19. Optimization

Algorithms for finding the best parameters for a model.

20- What is the difference between stochastic gradient descent (SGD) and gradient descent (GD)?

Both algorithms are methods for finding a set of parameters that minimize a loss function by evaluating parameters against data and then making adjustments.

In standard gradient descent, you'll evaluate all training samples for each set of parameters. This is akin to taking big, slow steps toward the ANSWER.

In stochastic gradient descent, you'll evaluate only 1 training sample for the set of parameters before updating them. This is akin to taking small, quick steps toward the ANSWER.

21. When would you use GD over SDG, and vice-versa?

GD theoretically minimizes the error function better than SGD. However, SGD converges much faster once the dataset becomes large.

That means GD is preferable for small datasets while SGD is preferable for larger ones.

In practice, however, SGD is used for most applications because it minimizes the error function well enough while being much faster and more memory efficient for large datasets.

22. What is the difference between parametric & non-parametric models?

Parametric model	Non-parametric model
It uses a fixed number of parameters to build the model.	It uses flexible number of parameters to build the model.
Considers strong assumptions about the data.	Considers fewer assumptions about the data.
Computationally faster	Computationally slower
Require lesser data	Require more data
Example – Logistic Regression & Naïve Bayes models	Example – KNN & Decision Tree models

23 What is the difference between Hyperparameters and model parameters?

Model Parameters	Hyperparameters
Model parameters are the features of training data that will learn on its own during training.	Model Hyperparameters are the parameters that determine the entire training process.
For example, • Weights and Biases • Split points in Decision Tree	For example, • Learning Rate • Hidden Layers • Hidden Units
They are internal to the model and their value can be estimated from data.	They are external to the model and their value cannot be estimated from data.

24. What are hyperparameters in Deep Neural Networks?

- Hyperparameters are variables that define the structure of the network. For example, variables such as the learning rate, define how the network is trained.
- They are used to define the number of hidden layers that must be present in a network.
- More hidden units can increase the accuracy of the network, whereas a lesser number of units may cause underfitting.

25. Explain the different algorithms used for hyperparameter optimization.

Grid Search

Grid search trains the network for every combination by using the two set of hyperparameters, learning rate and the number of layers. Then evaluates the model by using Cross Validation techniques.

Random Search

It randomly samples the search space and evaluates sets from a particular probability distribution. For example, instead of checking all 10,000 samples, randomly selected 100 parameters can be checked.

Bayesian Optimization

This includes fine-tuning the hyperparameters by enabling automated model tuning. The model used for approximating the objective function is called surrogate model (Gaussian Process). Bayesian Optimization uses Gaussian Process (GP) function to get posterior functions to make predictions based on prior functions.

Subject: Machine Learning (BE-A&B)

UNIT IV- SYLLABUS: Naïve Bayes and Support Vector Machine
Bayes‟ Theorom, Naïve Bayes‟ Classifiers, Naïve Bayes in Scikit- learn- Bernoulli Naïve Bayes, Multinomial Naïve Bayes, and Gaussian Naïve Bayes.
Support Vector Machine (SVM)- Linear Support Vector Machines, Scikit- learn implementation- Linear Classification, Kernel based classification, Non- linear Examples. Controlled Support Vector Machines, Support Vector Regression.

SECTION A: MCQS

1. Naive Bayes classifier
   A. Assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature
   B. Is easy to build and particularly useful for very large data sets.
   C. Is easy to build and particularly useful for small data sets.
   D. Only option 1 & 2
   E. Only option 1 & 3

ANSWER: D

2. Match the following:

1. Bernoulli	a. Can be used for both classification or regression challenges
2. Multinomial	b. Binary distribution, useful when a feature can be present or absent
3. Gaussian.	c. Discrete distribution and is used whenever a feature must be represented by a whole number
4. SVM	d. Continuous distribution characterized by its mean and variance.

ANSWER: 1-b,2-c,3-d,4-a

3. Kernels makes __________ work in _______ by mapping data to __________ where it exhibits _______ patterns.
   A. Nonlinear models, Linear settings, Nonzero dimensions , Nonlinear patterns
   B. Linear models, Nonlinear settings, Higher dimensions , Linear patterns
   C. None of the mentioned
   D. Both of the mentioned

ANSWER: B

4. Tick the odd one
   A. from sklearn.naive_bayes import BernoulliNB
   B. from sklearn.feature_extraction import DictVectorizer
   C. from sklearn.naive_bayes import GaussianNB
   D. Only B
   E. None of the mentioned

ANSWER: E

5. Which of the following are applications of the SVM?
   A. Text and Hypertext Categorization
   B. Image Classification
   C. Clustering of News Articles
   D. All of the above

ANSWER: D

6. The minimum time complexity for training an SVM is O(n2). According to this fact, what sizes of datasets are not best suited for SVM's?

A. Large datasets
B. Small datasets
C. Medium sized datasets
D. Size does not matter

ANSWER: A

7. Steps in Naïve Bayes algorithm are:

a. Use theorem equation to calculate the posterior probability for each class.
b. Convert the data set into a frequency table.
c. Create Likelihood table by finding the probabilities.

Correct sequence of these steps is:

A. a,b,c
B. b,c,a
C. a,c,b
D. b,a,c

ANSWER: B

8. In Naïve Bayes formula ,P(c|x) and P(x|c) represents,

A. P(c|x) is the likelihood which is the probability of predictor given class and P(x|c) is the posterior probability of class
B. P(c|x) is the posterior probability of class and P(x|c) is the likelihood which is the probability of predictor given class.
C. None of the above.
D. Both of the mentioned

ANSWER: B

9. Performance of the SVM depends upon no of training instances, Linear vs. non linear problems and

A. Input scale of features
B. The chosen hyperparameter
C. How you evaluate the model
D. All of the mentioned

ANSWER: D

10. SVM can be widely used in robotics and in computer vision for classifying objects and sensor data .

A. TRUE
B. FALSE

ANSWER: A

11. At a certain university, 4% of men are over 6 feet tall and 1% of women are over 6 feet tall. The total student population is divided in the ratio 3:2 in favour of women. If a student is selected at random from among all those over six feet tall, what is the probability that the student is a woman?

A. 2/5
B. 3/5
C. 3/11
D. 1/100

ANSWER: C

12. Support vector machines, like logistic regression models, give a probabilitydistribution over the possible labels given an input example.

A. TRUE
B. FALSE

ANSWER: B

13. Tick the odd one out with respect to advantages of Naïve bayestheorm

A. Can successfully train on large data set
B. Good for text classification, good for multiclass classification
C. Quick and simple calculation since it is naïve
D. All of the mentioned

ANSWER: A

14. Naive Bayes can be used in

A. Face Recognition, as a classifier to identify the faces or its other features, like nose, mouth, eyes etc
B. Weather Prediction to predict if the weather will be good or bad.
C. Medical Diagnosis to indicate if a patient is at high risk for certain diseases and conditions, such as heart disease, cancer, and other ailments.
D. News Classification to predict whether the news is political, world news, and so on.
E. All of the options mentioned
F. Only option A,B,C

ANSWER: E

15. In Gaussian Naive Bayes, continuous values associated with each feature are assumed to be distributed according to a

A. Gaussian distribution.
B. Normal distribution.
C. Both A & B
D. None of the options mentioned

ANSWER: C

16. The previous probabilities in Bayes theorm that are changed with the help of new available information are classified as

A. Independent probabilities
B. Posterior probabilities
C. Interior probabilities
D. Independent probabilities

ANSWER: B

17. The formula for Bayes' theorem is

A. P(H|E) = P(E|H) * P(H)/ P(E)
B. P(H|E) = P(H)/ P(E)* P(E|H)
C. P(H|E) =P(E)/ P(E|H) * P(H
D. All the options mentioned are correct

ANSWER: A

18. Which of the following loss functions are not convex? (Multiple options may be correct)

A. loss (sometimes referred as mis-classification loss)
B. Hinge loss
C. Logistic loss
D. Squared error loss

ANSWER: A

19. Which of the following properties is false in the case of a Bayesian Network?
   A. The edges are directed
   B. Contains cycles
   C. Represents conditional independence relations among random variables
   D. All of the above

ANSWER: B

20. Which of the following is/are not true regarding an SVM?
   A. For two dimensional data points, the separating hyperplane learnt by a linear SVM will be a straight line.
   B. In theory, a Gaussian kernel SVM can model any complex separating hyperplane.
   C. For every kernel function used in a SVM, one can obtain a equivalent closed form basis expansion.
   D. Over fitting in an SVM is a function of number of support vectors.

ANSWER: C

21. Select the correct statement Support Vector Machine
   A. SVM is a supervised machine learning algorithm which can be used for both classification and regression challenges.
   B. SVM is a supervised machine learning algorithm which can be used for both classification andclustering challenges.
   C. Both of the options mentioned
   D. None of the options mentioned

ANSWER: A

22. Hyperplane is
   A. Hyperplanes are decision boundaries that help classify the data points.
   B. A hyperplane in an n-dimensional Euclidean space is a flat, n-1 dimensional subset of that space that divides the space into two disconnected parts.
   C. The most common example of hyperplanes in practice is with support vector machines
   D. Option A & C only
   E. All of the options mentioned

ANSWER: E

23. The Naive Bayes algorithm is called "naive" because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features.

   A. TRUE
   B. FALSE

ANSWER: A

24. The components of the above statement are[TICK THE ODD ONE OUT]
   A. P(A|B): Probability (conditional probability) of occurrence of event A given the event B is true
   B. P(A) and P(B): Probabilities of the occurrence of event A and B respectively
   C. P(B|A): Probability of the occurrence of event B given the event A is false

ANSWER: C

25. Steps followed in Naïve Bayes theorem are (arrange the correct sequence)
   a. Draw the likelihood table for the features against the classes
   b. Create a frequency table for all the features against the different classes.
   c. Calculate maxiP(Ci|x1,x2,…,xn)
   d. Calculate the conditional probabilities for all the classes,

ANSWER: b-a-d-c

1.Support vector machine (SVM) is a ________ classifier?
A.Discriminative
B.Generative
ANSWER: A

2.SVM can be used to solve __________ problems.

A. Classification
B. Regression
C. Clustering
D. Both Classification and Regression

ANSWER: D

3.SVM is a __________ learning algorithm

A. Supervised
B. Unsupervised

ANSWER: A

4.SVM is termed as _______ classifier

A. Minimum margin
B. Maximum margin

ANSWER: B

5.The training examples closest to the separating hyperplane are called as ______

A. Training vectors
B. Test vectors
C. Support vectors

ANSWER: C

6.Which of the following is a type of SVM?

A. Maximum margin classifier
B. Soft margin classifier
C. Support vector regression
D. All of the above

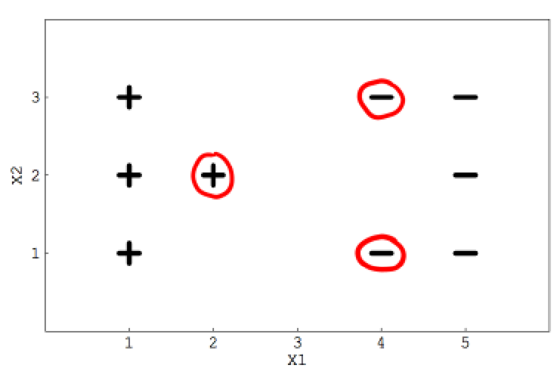
ANSWER: D

7.The goal of the SVM is to _________

A. Find the optimal separating hyperplane which minimizes the margin of training data
B. Find the optimal separating hyperplane which maximizes the margin of training data

ANSWER: B

Context for 8-9
Suppose you are using a Linear SVM classifier with 2 class classification problem. Now you have been given the following data in which some points are circled red that are representing support vectors.



8 If you remove the following any one red points from the data. Does the decision boundary will change?
A) Yes
B) No
ANSWER: A

9 [True or False] If you remove the non-red circled points from the data, the decision boundary will change?
A) True
B) False
ANSWER: B

10 What do you mean by generalization error in terms of the SVM?
A) How far the hyperplane is from the support vectors
B) How accurately the SVM can predict outcomes for unseen data
C) The threshold amount of error in an SVM
ANSWER: B

11 The cost parameter in the SVM means:
A) The number of cross-validations to be made
B) The kernel to be used
C) The tradeoff between misclassification and simplicity of the model
D) None of the above
ANSWER: C

12.Suppose you are building a SVM model on data X. The data X can be error prone which means that you should not trust any specific data point too much. Now think that you want to build a SVM model which has quadratic kernel function of polynomial degree 2 that uses Slack variable C as one of it's hyper parameter. Based upon that give the answer for following question.
What would happen when you use very large value of C(C->infinity)?
Note: For small C was also classifying all data points correctly

A) We can still classify data correctly for given setting of hyper parameter C
B) We can not classify data correctly for given setting of hyper parameter C
C) Can't Say
D) None of these
ANSWER: A

13 What would happen when you use very small C (C~0)?
A) Misclassification would happen
B) Data will be correctly classified
C) Can't say
D) None of these
ANSWER: A

14. If I am using all features of my dataset and I achieve 100% accuracy on my training set, but ~70% on validation set, what should I look out for?
A) Underfitting
B) Nothing, the model is perfect
C) Overfitting
ANSWER: C

15 Which of the following are real world applications of the SVM?
A) Text and Hypertext Categorization
B) Image Classification
C) Clustering of News Articles
D) All of the above
ANSWER: D

Question Context: 16 – 18
Suppose you have trained an SVM with linear decision boundary after training SVM, you correctly infer that your SVM model is under fitting.

16) Which of the following option would you more likely to consider iterating SVM next time?
A) You want to increase your data points
B) You want to decrease your data points
C) You will try to calculate more variables
D) You will try to reduce the features
ANSWER: C

17) Suppose you gave the correct answer in previous question. What do you think that is actually happening?
1. We are lowering the bias
2. We are lowering the variance
3. We are increasing the bias
4. We are increasing the variance

A) 1 and 2
B) 2 and 3
C) 1 and 4
D) 2 and 4
ANSWER: C

18 In above question suppose you want to change one of it's(SVM) hyperparameter so that effect would be same as previous questions i.e model will not under fit?
A) We will increase the parameter C
B) We will decrease the parameter C
C) Changing in C don't effect
D) None of these
ANSWER: A

19.We usually use feature normalization before using the Gaussian kernel in SVM. What is true about feature normalization?
1. We do feature normalization so that new feature will dominate other
2. Some times, feature normalization is not feasible in case of categorical variables
3. Feature normalization always helps when we use Gaussian kernel in SVM
A) 1
B) 1 and 2
C) 1 and 3
D) 2 and 3
ANSWER: B

Question Context: 20-22
Suppose you are dealing with 4 class classification problem and you want to train a SVM model on the data for that you are using One-vs-all method. Now answer the below questions?
20) How many times we need to train our SVM model in such case?
A) 1
B) 2
C) 3
D) 4
ANSWER: D

21 Suppose you have same distribution of classes in the data. Now, say for training 1 time in one vs all setting the SVM is taking 10 second. How many seconds would it require to train one-vs-all method end to end?
A) 20
B) 40

C) 60
D) 80
ANSWER: B

22. Suppose your problem has changed now. Now, data has only 2 classes. What would you think how many times we need to train SVM in such case?
A) 1
B) 2
C) 3
D) 4
ANSWER: A

Question context: 23 – 24
Suppose you are using SVM with linear kernel of polynomial degree 2, Now think that you have applied this on data and found that it perfectly fit the data that means, Training and testing accuracy is 100%.

23 Now, think that you increase the complexity(or degree of polynomial of this kernel). What would you think will happen?
A) Increasing the complexity will overfit the data
B) Increasing the complexity will underfit the data
C) Nothing will happen since your model was already 100% accurate
D) None of these
ANSWER: A

24 In the previous question after increasing the complexity you found that training accuracy was still 100%. According to you what is the reason behind that?
1. Since data is fixed and we are fitting more polynomial term or parameters so the algorithm starts memorizing everything in the data
2. Since data is fixed and SVM doesn't need to search in big hypothesis space
A) 1
B) 2
C) 1 and 2
D) None of these
ANSWER: C

25) What is/are true about kernel in SVM?
1. Kernel function map low dimensional data to high dimensional space
2. It's a similarity function
A) 1
B) 2
C) 1 and 2
D) None of these
ANSWER: C

Short Answer Questions

1.What are the low and high Bias Machine Learning algorithms?

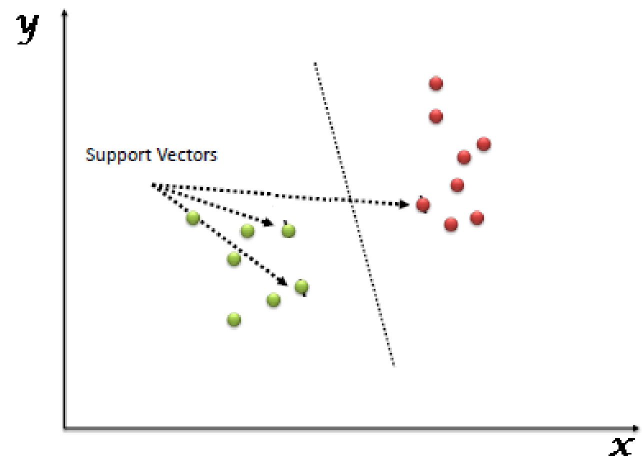
- Low bias Machine learning algorithms include Decision Trees, SVM, and k-NN.
- High bias Machine learning algorithms include Linear Regression and Logistic Regression.

2.Mention the advantages of performing the dimensionality reduction before fitting the SVM?

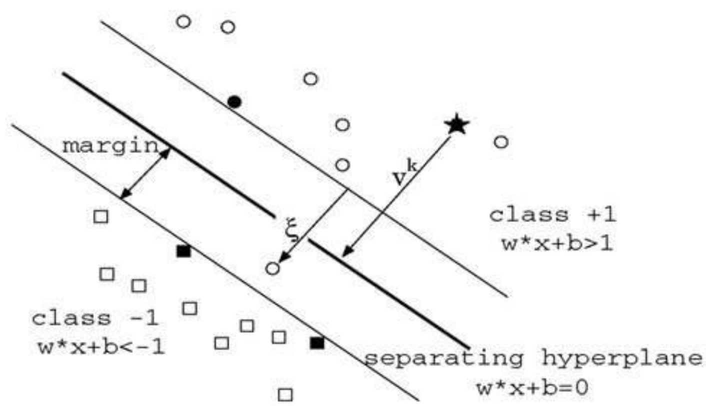
The Support Vector Machine Learning Algorithm mostly performs better in the reduced space. This is beneficial to perform the dimensionality reduction before fitting the SVM when the number of features is large while compared to many observations.

3. Explain SVM algorithm in detail.

SVM stands for support vector machine, it is a supervised machine learning algorithm which can be used for both ***Regression*** **and Classification**. If you have n features in your training data set, SVM tries to plot it in n-dimensional space with the value of each feature being the value of a particular coordinate. SVM uses hyperplanes to separate out different classes based on the provided kernel function.



4. What are the support vectors in SVM?



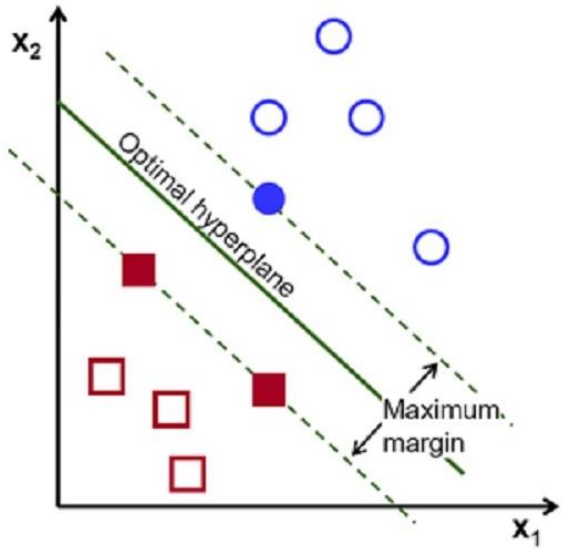
In the diagram, we see that the thinner lines mark the distance from the classifier to the closest data points called the support vectors (darkened data points). The distance between the two thin lines is called the margin.

5. What are the different kernels in SVM?

There are four types of kernels in SVM.

1. Linear Kernel
2. Polynomial kernel
3. Radial basis kernel
4. Sigmoid kernel

6.Why SVM is an example of a large margin classifier?



- SVM is a type of classifier which classifies positive and negative examples, here blue and red data points
- As shown in the image, the largest margin is found in order to avoid overfitting ie,.. the optimal hyperplane is at the maximum distance from the positive and negative examples(Equal distant from the boundary lines).
- To satisfy this constraint, and also to classify the data points accurately, the margin is maximised, that is why this is called the large margin classifier.

7. ***What is the role of C in SVM?***
Ans. The C parameter tells the SVM optimization how much you want to avoid misclassifying each training example. For large values of C, the optimization will choose a smaller-margin hyperplane if that hyperplane does a better job of getting all the training points classified correctly. Conversely, a very small value of C will cause the optimizer to look for a larger-margin separating hyperplane, even if that hyperplane misclassifies more points. For very tiny values of C, you should get misclassified examples, often even if your training data is linearly separable.

8. **What is generalization error in terms of the SVM?**
Ans. Generalisation error in statistics is generally the out-of-sample error which is the measure of how accurately a model can predict values for previously unseen data

9. ***What is a kernel in SVM? Why do we use kernels in SVM?***
Ans. SVM algorithms use a set of mathematical functions that are defined as the kernel. The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types. For example *linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid.* Introduce Kernel functions for sequence data, graphs, text, images, as well as vectors. The most used type of kernel function is RBF. Because it has localized and finite response along the entire x-axis. The kernel functions return the inner product between two points in a suitable feature space. Thus by defining a notion of similarity, with little computational cost even in very high-dimensional spaces.

10. **Suppose you are using RBF kernel in SVM with high Gamma value. What does this signify?**
Ans. The gamma parameter in SVM tuning signifies the influence of points either near or far away from the hyperplane.
For a low gamma, the model will be too constrained and include all points of the training dataset, without really capturing the shape.
For a higher gamma, the model will capture the shape of the dataset well.

Subject: Machine Learning (BE-A&B)

UNIT IV- SYLLABUS: Decision Trees and Ensemble Learning
Decision Trees- Impurity measures, Feature Importance. Decision Tree Classification with Scikitlearn, Ensemble Learning-Random Forest, AdaBoost, Gradient Tree Boosting, Voting Classifier. Introduction to Meta Classifier: Concepts of Weak and eager learner, Ensemble methods, Bagging, Boosting, Random Forests.
Clustering Fundamentals- Basics, K-means: Finding optimal number of clusters, DBSCAN, Spectral Clustering. Evaluation methods based on Ground Truth- Homogeneity, Completeness, Adjusted Rand Index.

SECTION A: MCQS

1. Given that we can select the same feature multiple times during the recursive partitioning of the input space, is it always possible to achieve 100% accuracy on the training data (given that we allow for trees to grow to their maximum size) when building decision trees?
   A. Yes
   B. No

ANSWER: B

2. Which of the following statements are true with respect to the application of Cost-Complexity Pruning and Reduced Error Pruning with Cross-Validation?
   A. In Reduced Error Pruning, the pruned tree error can never be less than the original tree
   B. on the training dataset.
   C. In Cost Complexity Pruning, the pruned tree error can never be less than the original
   D. tree on the validation dataset.
   E. In Reduced Error Pruning, the pruned tree error can never be less than the original tree
   F. on the validation dataset.
   G. both (b) and (c)

ANSWER: A

3. Suppose on performing reduced error pruning, we collapsed a node and observed an improvement in the prediction accuracy on the validation set. Which among the following statements re possible in light of the performance improvement observed? (multiple options may be correct)
   A. The collapsed node helped overcome the effect of one or more noise affected data points
   B. in the training set
   C. The validation set had one or more noise affected data points in the region corresponding
   D. to the collapsed node
   E. The validation set did not have any data points along at least one of the collapsed branches
   F. The validation set did have data points adversely affected by the collapsed node

ANSWER: A

4. Which of these classifiers do not require any additional modifications to their original descriptions (as seen in the lectures) to use them when we have more than 2 classes? (multiple options may be correct)
   A. decision trees
   B. logistic regression
   C. support vector machines
   D. k nearest neighbors

ANSWER: A & D

5. In a random forest model let m << p be the number of randomly selected features that are used to identify the best split at any node of a tree. Which of the following are true? (p is the original number of features) (Multiple options may be correct)
   A. increasing m reduces the correlation between any two trees in the forest
   B. decreasing m reduces the correlation between any two trees in the forest
   C. increasing m increases the performance of individual trees in the forest
   D. decreasing m increases the performance of individual trees in the forest

ANSWER: B & C

6. In AdaBoost, we re-weight points giving points misclassified in previous iterations more weight. Suppose we introduced a limit or cap on the weight that any point can take (for example, say we introduce a restriction that prevents any point's weight from exceeding a value of 10). Which among the following would be an effect of such a modification? (Multiple options may be correct)
   A. We may observe the performance of the classifier reduce as the number of stages increase
   B. It makes the final classifier robust to outliers
   C. It may result in lower overall performance

ANSWER: B & C

7. Which of the following method(s) is not inherently sequential?
   A. Gradient Boosting
   B. Committee machines
   C. AdaBoost

ANSWER: B

8. Boosting techniques typically give very high accuracy classifiers by sequentially training a collection of similar low-accuracy classifiers. Which of the following statements are true with respect to Boosting? (multiple options may be correct)
   A. LogitBoost (like AdaBoost, but with Logistic Loss instead of Exponential Loss) is less
   B. susceptible to overfitting than AdaBoost.
   C. Boosting techniques tend to have low bias and high variance
   D. Boosting techniques tend to have low variance and high bias
   E. For basic linear regression classifiers, there is no effect of using Gradient Boosting.

ANSWER: A,B ,D

9. In a tournament classifier with N classes. What is the complexity of the number of classifiers we require?
   A. O(N)
   B. O(N2)
   C. O(N.log(N))
   D. O(log(N))

ANSWER: A

10. Which of the following statements are true about ensemble classifiers? (multiple options may be correct)
   A. The different learners in boosting based ensembles can be trained in parallel
   B. The different learners in bagging based ensembles can be trained in parallel
   C. Boosting based algorithms which iteratively re-weight training points, such as AdaBoost,
   D. are more sensitive to noise than bagging based methods.
   E. Boosting methods generally use strong learners as individual classifiers
   F. Boosting methods generally use weak learners as individual classifiers.
   G. An individual classifier in a boosting based ensemble is trained with every point in the
   H. training set.

ANSWER: B.C.E.F

11. In which approach do the classification models train on data sets whose distribution are modified in comparison to the distribution of the original training data set
    A. bagging
    B. boosting
    C. both
    D. neither

ANSWER: C

(Explanation: In bagging, each classifier is trained on a data set generated by sampling from the original training data set with replacement, resulting in stochastic modification of the original data distribution. Similarly, in boosting, each classifier is trained on a data set generated by modifying the weights of the data instances based on the performance of the previous classifier,resulting again in modification of the original data distribution.)

12. Which of the following is/are false about bagging?
    A. Bagging reduces variance of the classifier
    B. Bagging increases the variance of the classifier
    C. Bagging can help make robust classifiers from unstable classifiers
    D. Bagging results in increased bias

ANSWER : B.D

(Explanation : In bagging we combine the outputs of multiple classifiers trained on different samples of the training data. This helps in reducing overall variance. Due to the reduction in variance, normally unstable classifiers can be made robust with the help of bagging.)

13. Considering the AdaBoost algorithm, which among the following statements is false?
    A. In each stage, we try to train a classifier which makes accurate predictions on any subset
    B. of the data points where the subset size is at least half the size of the data set
    C. In each stage, we try to train a classifier which makes accurate predictions on a subset of
    D. the data points where the subset contains more of the data points which were miscalssified
    E. in earlier stages
    F. The weight assigned to an individual classifier depends upon the number of data points
    G. correctly classified by the classifier
    H. The weight assigned to an individual classifier depends upon the weighted sum error of
    I. misclassified points for that classifier

ANSWER: A,C

14. Which of the following measure best analyze the performance of a classifier?
    A. Precision
    B. Recall
    C. Accuracy
    D. Time complexity
    E. Depends on the application

ANSWER: E

(Explanation Different applications might need to optimize different performance measures.Applications of machine learning span over playing games to very critical domains(such as health and security). Measures like accuracy for instance cannot be reliable when we have a dataset with significant class imbalance. So there cannot be a single measure to analyze the effectiveness of a classifier in all environments.)

15. Which of the following is required by K-means clustering?
    A. defined distance metric
    B. number of clusters
    C. initial guess as to cluster centroids
    D. all of the mentioned

ANSWER: D

16. Point out the wrong statement.
A. k-means clustering is a method of vector quantization
B. k-means clustering aims to partition n observations into k clusters
C. k-nearest neighbor is same as k-means
D. none of the mentioned

ANSWER: C

17. K-means is not deterministic and it also consists of number of iterations.
A. True
B. False

ANSWER: A

18. Which points are eliminated by the DBSCAN algorithm?
A. Core points
B. Border points
C. Noise points

ANSWER: C

19. How is the density of point p at the density based clustering defined?
A. MinPts minus number of data points in an epsilon-neighbourhood
B. Number of data points in an epsilon-neighbourhood of p
C. Reciprocal value of the distance from p to the nearest neighbour

ANSWER: B

20 Which of the following tasks can be best solved using Clustering.
A. Predicting the amount of rainfall based on various cues
B. Detecting fraudulent credit card transactions
C. Training a robot to solve a maze

ANSWER: B

21. Which of the following properties are characteristic of decision trees?
A. High bias
B. High variance
C. Lack of smoothness of prediction surfaces
D. Unbounded parameter set

ANSWER : B.C.D

22. Which among the following prevents overfitting when we perform bagging?
A. The use of sampling with replacement as the sampling technique
B. The use of weak classifiers
C. The use of classification algorithms which are not prone to overfitting
D. The practice of validation performed on every classifier trained

ANSWER: B

23. Consider an alternative way of learning a Random Forest where instead of randomly sampling the attributes at each node, we sample a subset of attributes for each tree and build the tree on these features. Would you prefer this method over the original or not, and why?

A. Yes, because it reduces the correlation between the resultant trees
B. Yes, because it reduces the time taken to build the trees due to the decrease in the attributes considered
C. No, because many of the trees will be bad classifiers due to the absence of critical features considered in the construction of some of the trees

ANSWER: C

24. In case of limited training data, which technique, bagging or stacking, would be preferred, and why?

A. Bagging, because we can combine as many classifier as we want by training each on a different sample of the training data
B. Bagging, because we use the same classification algorithms on all samples of the training data
C. Stacking, because each classifier is trained on all of the available data
D. Stacking, because we can use different classification algorithms on the training data

ANSWER: C

25. Is AdaBoost sensitive to outliers?

A. Yes
B. No

ANSWER: A

26. Which of the following is/are true about bagging trees?

1. In bagging trees, individual trees are independent of each other
2. Bagging is the method for improving the performance by aggregating the results of weak learners

A) 1
B) 2
C) 1 and 2
D) None of these

Answer: C

27. Which of the following is/are true about Random Forest and Gradient Boosting ensemble methods?

1. Both methods can be used for classification task
2. Random Forest is use for classification whereas Gradient Boosting is use for regression task
3. Random Forest is use for regression whereas Gradient Boosting is use for Classification task
4. Both methods can be used for regression task

A) 1
B) 2
C) 3
D) 4
E) 1 and 4

Answer: E

28. In Random forest you can generate hundreds of trees (say T1, T2 …..Tn) and then aggregate the results of these tree. Which of the following is true about individual(Tk) tree in Random Forest?

1. Individual tree is built on a subset of the features
2. Individual tree is built on all the features
3. Individual tree is built on a subset of observations
4. Individual tree is built on full set of observations

A) 1 and 3
B) 1 and 4
C) 2 and 3
D) 2 and 4

Answer: A

29. Which of the following is true about "max_depth" hyperparameter in Gradient Boosting?

1. Lower is better parameter in case of same validation accuracy
2. Higher is better parameter in case of same validation accuracy
3. Increase the value of max_depth may overfit the data
4. Increase the value of max_depth may underfit the data

A) 1 and 3
B) 1 and 4
C) 2 and 3
D) 2 and 4

Answer: A

30. Which of the following algorithm doesn't uses learning Rate as of one of its hyperparameter?

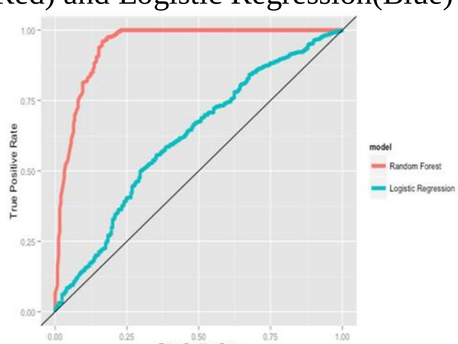
1. Gradient Boosting
2. Extra Trees
3. AdaBoost
4. Random Forest

A) 1 and 3
B) 1 and 4
C) 2 and 3
D) 2 and 4

Answer: D

31. Which of the following algorithm would you take into the consideration in your final model building on the basis of performance?

Suppose you have given the following graph which shows the ROC curve for two different classification algorithms such as Random Forest(Red) and Logistic Regression(Blue)



A) Random Forest
B) Logistic Regression
C) Both of the above
D) None of these

Answer: A

32. Which of the following is true about training and testing error in such case?
Suppose you want to apply AdaBoost algorithm on Data D which has T observations. You set half the data for training and half for testing initially. Now you want to increase the number of data points for training T1, T2 … Tn where T1 < T2…. Tn-1 < Tn.

A) The difference between training error and test error increases as number of observations increases
B) The difference between training error and test error decreases as number of observations increases
C) The difference between training error and test error will not change
D) None of These

Answer: B

33. In random forest or gradient boosting algorithms, features can be of any type. For example, it can be a continuous feature or a categorical feature. Which of the following option is true when you consider these types of features?

A) Only Random forest algorithm handles real valued attributes by discretizing them
B) Only Gradient boosting algorithm handles real valued attributes by discretizing them
C) Both algorithms can handle real valued attributes by discretizing them
D) None of these

Answer: C

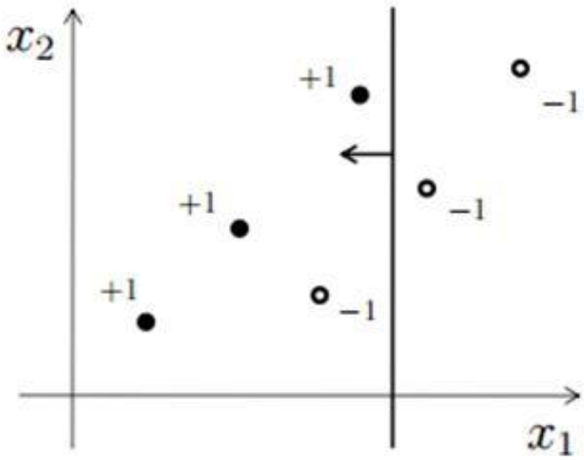
34. Suppose you are using a bagging based algorithm say a RandomForest in model building. Which of the following can be true?

1. Number of tree should be as large as possible
2. You will have interpretability after using RandomForest

A) 1
B) 2
C) 1 and 2
D) None of these

Answer: A

Context 11-14
Consider the following figure for answering the next few questions. In the figure, X1 and X2 are the two features and the data point is represented by dots (-1 is negative class and +1 is a positive class). And you first split the data based on feature X1(say splitting point is x11) which is shown in the figure using vertical line. Every value less than x11 will be predicted as positive class and greater than x will be predicted as negative class.



35. How many data points are misclassified in above image?
    - A) 1
    - B) 2
    - C) 3
    - D) 4

Answer: A

36. Which of the following splitting point on feature x1 will classify the data correctly?
    - A) Greater than x11
    - B) Less than x11
    - C) Equal to x11
    - D) None of above

Answer: D

37. If you consider only feature X2 for splitting. Can you now perfectly separate the positive class from negative class for any one split on X2?
    - A) Yes
    - B) No

Answer: B

38. Now consider only one splitting on both (one on X1 and one on X2) feature. You can split both features at any point. Would you be able to classify all data points correctly?
    - A) TRUE
    - B) FALSE

Answer: B

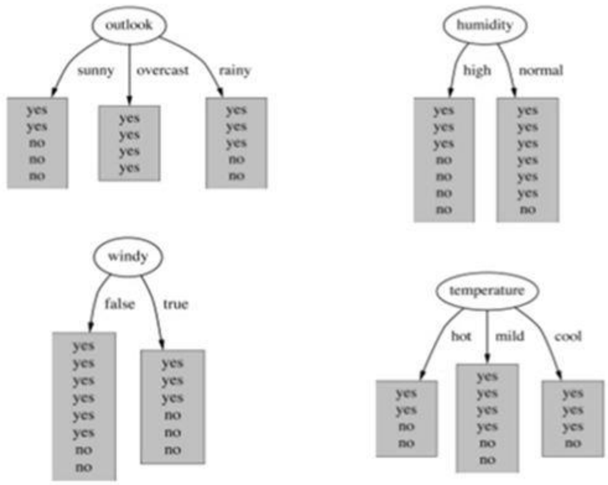
Context 15-16
Suppose, you are working on a binary classification problem with 3 input features. And you chose to apply a bagging algorithm(X) on this data. You chose max_features = 2 and the n_estimators =3. Now, Think that each estimators have 70% accuracy.
Note: Algorithm X is aggregating the results of individual estimators based on maximum voting
39. What will be the maximum accuracy you can get?
A) 70%
B) 80%
C) 90%
D) 100%
Answer: D
Refer below table for models M1, M2 and M3.

Actual predictions	M1	M2	M3	Output
1	1	0	1	1
1	1	0	1	1
1	1	0	1	1
1	0	1	1	1
1	0	1	1	1
1	0	1	1	1
1	1	1	1	1
1	1	1	0	1
1	1	1	0	1
1	1	1	0	1

40. What will be the minimum accuracy you can get?
A) Always greater than 70%
B) Always greater than and equal to 70%
C) It can be less than 70%
D) None of these
Answer: C
Refer below table for models M1, M2 and M3.

Actual predictions	M1	M2	M3	Output
1	1	0	0	0
1	1	1	1	1
1	1	0	0	0
1	0	1	0	0
1	0	1	1	1
1	0	0	1	0
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

41. Suppose you are building random forest model, which split a node on the attribute, that has highest information gain. In the below image, select the attribute which has the highest information gain?



A) Outlook
B) Humidity
C) Windy
D) Temperature
Answer: A

42. Which of the following is true about the Gradient Boosting trees?

1. In each stage, introduce a new regression tree to compensate the shortcomings of existing model
2. We can use gradient decent method for minimize the loss function

A) 1
B) 2
C) 1 and 2
D) None of these
Answer: C

43. Which of the following is true when you choose fraction of observations for building the base learners in tree based algorithm?

A) Decrease the fraction of samples to build a base learners will result in decrease in variance
B) Decrease the fraction of samples to build a base learners will result in increase in variance
C) Increase the fraction of samples to build a base learners will result in decrease in variance
D) Increase the fraction of samples to build a base learners will result in Increase in varianc

Answer: A

44. In greadient boosting it is important use learning rate to get optimum output. Which of the following is true abut choosing the learning rate?

A) Learning rate should be as high as possible
B) Learning Rate should be as low as possible
C) Learning Rate should be low but it should not be very low
D) Learning rate should be high but it should not be very high

Answer: C

45. [True or False] Cross validation can be used to select the number of iterations in boosting; this procedure may help reduce overfitting.

A) TRUE

B) FALSE

Answer: A

46. Which one of these is not a tree based learner?

A. CART
B. ID3
C. Bayesian classifier
D. Random Forest

ANSWER: C

47. Which one of these is a tree based learner?

A. Rule based
B. Bayesian Belief Network
C. Bayesian classifier
D. Random Forest

ANSWER: d

48. What is the approach of basic algorithm for decision tree induction?

A. Greedy
B. Top Down
C. Procedural
D. Step by Step

ANSWER : A

49. In the root node how many classes will be there?

A. 3
B. 2
C. 4
D. 14

ANSWER : B

50. Which among the following is/are some of the assumptions made by the k-means algorithm (assuming Euclidean distance measure)?

A. Clusters are spherical in shape
B. Clusters are of similar sizes
C. Data points in one cluster are well separated from data points of other clusters
D. There is no wide variation in density among the data points

ANSWER : A & B

SECTION A: Short questions and answers

1. What is Clustering?

Clustering is dividing data points into homogeneous classes or clusters:

- Points in the same group are as similar as possible
- Points in different group are as dissimilar as possible

When a collection of objects is given, we put objects into group based on similarity.

2. Application of Clustering:

Clustering is used in almost all the fields. You can infer some ideas from Example 1 to come up with lot of clustering applications that you would have come across.

Listed here are few more applications, which would add to what you have learnt.

- Clustering helps marketers improve their customer base and work on the target areas. It helps group people (according to different criteria's such as willingness, purchasing power etc.) based on their similarity in many ways related to the product under consideration.
- Clustering helps in identification of groups of houses on the basis of their value, type and geographical locations.
- Clustering is used to study earth-quake. Based on the areas hit by an earthquake in a region, clustering can help analyse the next probable location where earthquake can occur.

3.What is K-means Clustering?

K-means (Macqueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining.

4.K-means Clustering Method?

If k is given, the K-means algorithm can be executed in the following steps:

A. Partition of objects into k non-empty subsets
B. Identifying the cluster centroids (mean point) of the current partition.
C. Assigning each point to a specific cluster
D. Compute the distances from each point and allot points to the cluster where the distance from the centroid is minimum.
E. After re-allotting the points, find the centroid of the new cluster formed.

5.what are the Clustering Algorithms?

A Clustering Algorithm tries to analyse natural groups of data on the basis of some similarity. It locates the centroid of the group of data points. To carry out effective clustering, the algorithm evaluates the distance between each point from the centroid of the cluster.

The goal of clustering is to determine the intrinsic grouping in a set of unlabelled data.

6.How to deal with K-means for Big Data accepting inputs with more than 22 attributes?

Is it really necessary to have such high dimensional space? If you think about it, the result of k-means will be a set of k vectors of 22 means (in the 22 variables). I don't think it will be easy to interpret this, and certainly it will be impossible to visualize. It may be better to reduce the number of dimensions first by PCA (principle components). Do you have any idea about the number of clusters (k) — this would seem to be something worth exploring rather than pushing the size of the attributes.

If you MUST have large dimensions, you could code your own k-means algorithm without this limitation — the k-means idea is quite simple and you could write the code in MATLAB instead of relying on canned code.

7.Whether clustering can be applied for a 3D data set for classification?

The k-means algorithms do not depend upon the dimensions of data. All that you need is to find a suitable metric to measure the similarity. Only problem will be the time taken as the more the dimensions the more is the computation time for similarity. Since this is to be done repeatedly, it requires a lot of time.

8. How to Finding Cluster Centers that Minimize Distortion?

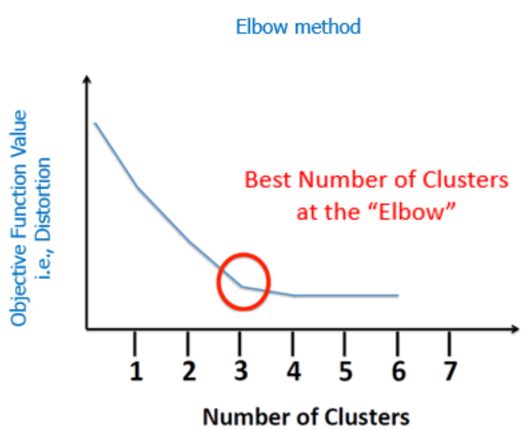
Solution can be found by setting the partial derivative of Distortion w.r.t. each cluster center to zero.

$$\frac{\partial \textbf{\textit{Distortion}}}{\partial \mu_j} = \frac{\partial}{\partial \mu_j} \sum_{i \in \text{OwnedBy}(\mu_j)} (x_i - \mu_j)^2 = -2 \sum_{i \in \text{OwnedBy}(\mu_j)} (x_i - \mu_j) = 0 \ \textbf{\textit{(for minimum)}}$$

$$\Rightarrow \mu_j = \frac{1}{|\text{OwnedBy}(\mu_j)|} \sum_{i \in \text{OwnedBy}(\mu_j)} x_i$$

For any k clusters, the value of k should be such that even if we increase the value of k from after several levels of clustering the distortion remains constant. The achieved point is called the "Elbow".

This is the ideal value of k, for the clusters created.



9.When would you use k means cluster and when would you use hierarchical cluster?

Hierarchical Clustering and k-means clustering complement each other. In hierarchical clustering, the researcher is not aware of the number of clusters to be made whereas in k-means clustering, the number of clusters to be made are specified before-hand.

Advice- If unaware about the number of clusters to be formed, use hierarchical clustering to determine the number and then use k-means clustering to make more stable clusters as hierarchical clustering is a single-pass exercise whereas k-means is an iterative process.

10.K-Means has an assumption “each cluster has a roughly equal number of observations”?

There is certainly no assumption in standard K-means algorithms that assumes an equal number of points in each cluster. However, certain standard algorithms do have a tendency towards equalising the *spatial* variance of clusters, which can result in a (rough) tendency towards equality of cluster sizes in cases where there is overlap between the clusters. For example, one standard method is to estimate the clusters by minimising the within-cluster sum-of-squares (WCSS). In cases where there are several overlapping clusters, this method has a tendency to allocate points in a way that (roughly) equalises the spacial variance of the clusters, which may result in (rough) equalisation of the number of points in each cluster. Alternative methods that use parametric forms to allow greater freedom of variance in each cluster will lack this tendency.

Subject: Machine Learning (BE-A&B)

UNIT IV- SYLLABUS: Clustering Techniques
Hierarchical Clustering, Expectation maximization clustering, Agglomerative Clustering- Dendrograms, Agglomerative clustering in Scikit- learn, Connectivity Constraints.
Introduction to Recommendation Systems- Naïve User based systems, Content based Systems, Model free collaborative filtering-singular value decomposition, alternating least squares. Fundamentals of Deep Networks-Defining Deep learning, common architectural principles of deep networks, building blocks of deep networks.

SECTION A: MCQS

1. In hierarchical clustering, is it possible for a point to be closer to points in other clusters than to points in its own cluster? If so, in which approach will this tend to be observed?
   A. Single-link
   B. Complete-link
   C. Centroid-based
   D. None of the above

ANSWER: B AND C

2. What assumption does the CURE clustering algorithm make with regards to the shape of the clusters?
   A. No assumption
   B. Spherical
   C. Elliptical

ANSWER: A

3. What would, in general, can be the effect of increasing MinPts in DBSCAN while retaining the same Eps parameter? (Note that more than one statement may be correct)
   A. Increase in the sizes of individual clusters
   B. Decrease in the sizes of individual clusters
   C. Increase in the number of clusters
   D. Decrease in the number of clusters

ANSWER: B AND C

4. Considering single-link and complete-link hierarchical clustering, is it possible for a point to be closer to points in other clusters than to points in its own cluster? If so, in which approach will this tend to be observed?
   A. No
   B. Yes, single-link clustering
   C. Yes, complete-link clustering
   D. Yes, both single-link and complete-link clustering

ANSWER: D

5. A graph is said to be k-connected if there does not exist a set of k-1 vertices whose removal disconnects the graph. If we define clusters as comprising of k-connected components of the thresholded graphs, does this result in a well-defined clustering algorithm?
   A. Yes
   B. No

ANSWER: A

Prepared by: Mrs. Deepali Gohil& Mrs. Nilam Patil (Subject Teachers)

6. A set of nodes forms a p-cluster, if at least p percentage of the edges from the nodes in the set go to another node in the set. If we define clusters as comprising of p-clusters of the thresholded graphs, does this result in a well-defined clustering algorithm?

A. Yes
B. No

ANSWER: B

7. In the CURE clustering algorithm, representative points of a cluster are moved a fraction of the distance between their original location and the centroid of the cluster. Would it make more sense to move them all a fixed distance towards the centroid instead? Why or why not?

A. Yes, because this approach will ensure that the original cluster shape is preserved.
B. No, because this approach will not be as effective against outliers as the original approach.

ANSWER: B

8. Suppose while performing DBSCAN we randomly choose a point which has less than MinPts number of points in its neighbourhood. Which among the following is true for such a point?

A. It is treated as noise, and not considered further in the algorithm
B. It becomes part of its own cluster
C. Depending upon other points, it may later turn out to be a core point
D. Depending upon other points, it may be density connected to other points

ANSWER: D

9. Which of the following statements are true about similarity graph based representations which are used for spectral clustering? (Note that more than one statements may be correct)

A. One can give a tighter upper bound than O(n) (where n is the number of data points) on the maximum degree of the vertex corresponding to a point in its kNN based similarity graph representation
B. One can give a tighter upper bound than O(n) (where n is the number of data points) on the maximum degree of the vertex corresponding to a point in its epsilon neighborhood based similarity graph representation
C. If a is in the k nearest neighbors of b, then b is in the k nearest neighbors of a
D. If a is in the epsilon neighborhood of b, then b is in the epsilon neighborhood of a

ANSWER: A AND D

10. Which of the following is untrue regarding Expectation Maximization algorithm?

A. An initial guess is made as to the location and size of the site of interest in each of the sequences, and these parts of the sequence are aligned
B. The alignment provides an estimate of the base or amino acid composition of each column in the site
C. The column-by-column composition of the site already available is used to estimate the probability of finding the site at any position in each of the sequences
D. The row-by-column composition of the site already available is used to estimate the probability

ANSWER:D

11. Out of the two repeated steps in EM algorithm, the step 2 is _______

A. the maximization step
B. the minimization step
C. the optimization step
D. the normalization step

Prepared by: Mrs. Deepali Gohil& Mrs. Nilam Patil (Subject Teachers)

ANSWER:A

12. Point out the correct statement.
    A. The choice of an appropriate metric will influence the shape of the clusters
    B. Hierarchical clustering is also called HCA
    C. In general, the merges and splits are determined in a greedy manner
    D. All of the mentioned

ANSWER: D

13. Which of the following is finally produced by Hierarchical Clustering?
    A. final estimate of cluster centroids
    B. tree showing how close things are to each other
    C. assignment of each point to clusters
    D. all of the mentioned

ANSWER: B

14. Which of the following is required by K-means clustering?
    A. defined distance metric
    B. number of clusters
    C. initial guess as to cluster centroids
    D. all of the mentioned

ANSWER: D

15. Hierarchical clustering should be primarily used for exploration.
    A. True
    B. False

ANSWER: A

16. ___________clustering approach initially assumes that each data instance represents a single cluster
    A. Hierarchical Clustering
    B. Expectation maximization clustering
    C. Agglomerative Clustering
    D. K means clustering

ANSWER: C

17. Sentiment Analysis is an example of:
    a. Regression
    b. Classification
    c. Clustering
    d. Reinforcement Learning

    A. a Only
    B. a and b
    C. a and c
    D. a, b and c
    E. a, b and d
    ANSWER: E

Prepared by: Mrs. Deepali Gohil& Mrs. Nilam Patil (Subject Teachers)

18. Can decision trees be used for performing clustering?

A. True
B. False

ANSWER: A

19. After performing K-Means Clustering analysis on a dataset, you observed the following dendrogram. Which of the following conclusion can be drawn from the dendrogram?

A. There were 28 data points in clustering analysis
B. The best no. of clusters for the analyzed data points is 4
C. The proximity function used is Average-link clustering
D. The above dendrogram interpretation is not possible for K-Means clustering analysis

ANSWER: D

20. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithm for the same dataset?

A. Proximity function used
B. of data points used
C. of variables used
D. B and c only
E. All of the above

ANSWER:E

21. What is the most appropriate no. of clusters for the data points represented by the following dendrogram:

A. 2
B. 4
C. 6
D. 8

ANSWER: B

22. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?

a. Single-link
b. Complete-link
c. Average-link
d. Options:

A. a and b
B. a and c
C. b and c
D. a,b and c

ANSWER: D

Prepared by: Mrs. Deepali Gohil& Mrs. Nilam Patil (Subject Teachers)

23. Which of the following is/are valid iterative strategy for treating missing values before clustering analysis?

A. Imputation with mean
B. Nearest Neighbor assignment
C. Imputation with Expectation Maximization algorithm
D. All of the above

ANSWER:C

24. If you are using Multinomial mixture models with the expectation-maximization algorithm for clustering a set of data points into two clusters, which of the assumptions are important:

A. All the data points follow two Gaussian distribution
B. All the data points follow n Gaussian distribution (n >2)
C. All the data points follow two multinomial distribution
D. All the data points follow n multinomial distribution (n >2)

ANSWER: C

25. Chameleon is

A. Density basedalgorithm
B. Partitioning basedalgorithm
C. Model basedalgorithm
D. Hierarchical clustering algorithm

ANSWER: D

26. Which of the following clustering algorithms suffers from the problem of convergence at local optima?

1. K- Means clustering algorithm
2. Agglomerative clustering algorithm
3. Expectation-Maximization clustering algorithm
4. Diverse clustering algorithm

A. 1 only
B. 2 and 3
C. 2 and 4
D. 1 and 3
E. 1,2 and 4
F. All of the above
ANSWER: D.

27. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithm for the same dataset?

A. Proximity function used
B. of data points used
C. of variables used
D. B and c only
E. All of the above
ANSWER: E

28. Collaborative Filtering and Content Based Models are the two popular recommendation engines, what role does NLP play in building such algorithms.

A. Feature Extraction from text
B. Measuring Feature Similarity

C Engineering Features for vector space learning model
D. All of these
ANSWER: D

29..The singular value decomposition of a real matrix is unique.
A.True
B.False
ANSWER: B

30. All but one of these techniques can be used for building a content filtering profile for a user. Which of these techniques is NOT used for building a content filtering profile?

A. Provide an interface where users can specify and edit their own vector.
B. Build an attribute preference vector from explicit user ratings.
C. Build an attribute preference vector based on the most popular items in the catalog.
D. Build an attribute preference vector based on user

ANSWER: C

31. Each of the following statements describes Entrée Style recommenders except one.
Which of these statements DOES NOT describe the Entrée Style Recommenders?

A. They don't use individual users' ratings of the items anywhere in the recommendation process.
B. They build a model of user preferences that can be used to provide personalized recommendations.
C. They require a substantial collection of information about the items being recommended.
D. They provide an interface that allows the user to refine recommendations by requesting items that differ in a certain way from the current recommendation.

ANSWER: B

32.When is "term-frequency" most useful as part of a content-filtering recommender?

A. When certain items are much more popular than other items.
B. When the attributes of the items can apply in different degrees to different items.
C. When users are unlikely to have experienced many of the items in the system.
D. When certain terms aren't very useful because they apply to too many different items.

ANSWER: B

33.TRUE/FALSE things that are *similar* about model based clustering and kmeans clustering .

(i) both methods are used to assign data to clusters
(ii) both methods work well with spherical clusters

ANSWER:TRUE

34.TRUE/FALSE things that are *different* about model based clustering and kmeans clustering .
(i) kmeans has no model, mclust uses a model
(ii) kmeans has no objective criterion for choosing number of clusters; mclust uses the objective BIC criterion.
ANSWER:TRUE

35. In model-based clustering, when do observations come from the same true cluster?
A. When they come from the same distribution
B. When they have the highest posterior probability of belonging to the same cluster
C. When they are close to each other in terms of Mahalanobis distance
D. When they are close to each other in terms of Euclidean distance
ANSWER:A

Prepared by: Mrs. Deepali Gohil& Mrs. Nilam Patil (Subject Teachers)

36.Which of the followings is responsible for indexing content in a manner that permits fast retrieval through multiple search mechanisms?

A. legacy integration
B. system architecture
C. Scalability
D. Collaborative filtering

Answers: D

37.Which of the following is an example of active learning?
A. News Recommender system
B. Dust cleaning machine
C Automated vehicle
D. None of the mentioned
ANSWER:A

38. Autonomous Question/Answering systems are _______________
A. Expert Systems
B. Rule Based Expert Systems
C Decision Tree Based Systems
D. All of the mentioned
ANSWER: D

39.You have collected a data of about 10,000 rows of tweet text and no other information. You want to create a tweet classification model that categorizes each of the tweets in three buckets – positive, negative and neutral. Which of the following models can perform tweet classification with regards to context mentioned above?
A. Naive Bayes
B. SVM
C None of the above
ANSWER: C

40. Classes of Collaborative Filtering includes
A. User Based
B: Item Based
C.Both User Based and Item Based
D.None of them
ANSWER: C