"Techno-Social Excellence"
Marathwada Mitramandal's
INSTITUTE OF TECHNOLOGY (MMIT)
Lohgaon, Pune-411 047

"Towards ubiquitous Computing Technology"

Department of Computer Engineering

[Total No. of Questions:-25] **[Total No. of pages- 07]**
[BE]- 410250

B. E. (Computer) [Second Semester]
Online Pre-Examination on MCQs, April 2020
Machine Learning
(2015 Pattern)

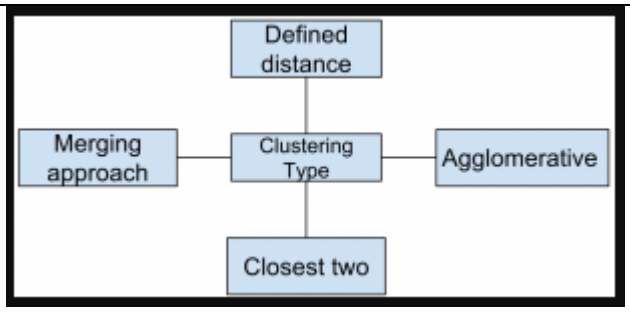
Time: 1 Hr **Maximum Marks: 50**
Date: 27/04/2020

Instructions:- (i) Choose appropriate option for your Answer.
(ii) Each question carries 2 Marks
(iii) Assume suitable data, if necessary

Q1.	Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____			
	A. independent probabilities	B. posterior probabilities	C. interior probabilities	D. dependent probabilities
Q2.	The method in which the previously calculated probabilities are revised with new probabilities is classified as			
	A. updating theorem	B. revised theorem	C. Bayes theorem	D. dependency theorem
Q3.	The previous probabilities in Bayes Theorem that are changed with the help of new available information are classified as			
	A.independent probabilities	B. posterior probabilities	C. interior probabilities	D. dependent probabilities
Q4.	The model which assumes that all our features are binary such that they take only two values is			
	A. Multinomial Naïve Bayes	B. Gaussian Naïve Bayes	C. Bernoulli Naïve Bayes	D.none
Q5.	Any linear model can be turned into a non-linear model by applying the **kernel** trick to the model			

	A.true	B.false	C.	D.
Q6.	**The effectiveness of an SVM depends upon:**			
	Selection of A Kernel	B) Kernel Parameters	C) Soft Margin Parameter C	D) All of the above
Q7.	In Classification Model, Which Technique can help you to choose a threshold that balance sensitivity and specificity			
	Confusion Matrix	b) ROC curve	c) MAPE	d) None of the Above
Q8.	**In Decision Tree, by comparing the impurity across all possible splits in all possible Predictors, the next split is choosen. How we can measure the Impurity ?**			
	UC b) **Entropy, Ginni** c) **ROC** d) **MAPE**	**UC** b) **Entropy, Ginni** c) **ROC** d) **MAPE**	**UC** b) **Entropy, Ginni** c) **ROC** d) **MAPE**	**UC** b) **Entropy, Ginni** c) **ROC** d) **MAPE**
Q9.	**How we can avoid the overfitting in Decision Tree**			
	A. CHAID(Stopping the Tree Growth	B Pruning the Full Grown Tree .	C.Both of above	D. None of the Above
Q10.	**Predictive Errors are due to**			
	A.	B.	C.Both of above	D. None of the Above

Q11.	**. Random Forest Modeling (Ensemble Modeling) uses .**			
	A. **Bagging(BootStrap Samples)** B] Boosting c)Both of above d) None of the Above	B.	C.	D.
Q12.	**Support Vector Machines are based on the concept of decision planes that define decision boundaries. Which technique used to identify the right hyperplances**			
	A. **Kernels**	B. **CHAID**	Both of above	none
Q13.	**Which one is NOT used as a Kernel in Support Vector Machine Modeling**			
	A. **Radial Basis Function(RBF)**	B. **Polynomial**	C. **Sigmoid**	D. **AUC**
Q14.	**To Identify the similar(neighbouring) records, K-Nearest Neighbous uses which distance metrics**			
	A. **Euclidean Distance**	B. **Perpendicular Distance**	C. **Diagonal Distance**	D. **None of the Above**
Q15.				



16Which of the following clustering type has characteristic shown in the below figure?

A. Partitional b) Hierarchical c) Naive bayes d) None of the mentioned	B.	C.	D.

Explanation:Hierarchical clustering groups data over a variety of scales by creating a cluster tree or dendrogram.

17. 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

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

18. 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

Explanation: Hierarchical clustering is an agglomerative approach.

19. 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

Explanation: K-means clustering follows partitioning approach.

20 Point out the wrong statement. **[1 M]**
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
Explanation: k-nearest neighbor has nothing to do with k-means.

21. Which of the following function is used for k-means clustering?
a) k-means
b) k-mean
c) heatmap
d) none of the mentioned

Explanation: K-means requires a number of clusters.

22.Which of the following clustering requires merging approach?
a) Partitional
b) Hierarchical
c) Naive Bayes
d) None of the mentioned

Explanation: Hierarchical clustering requires a defined distance as well.

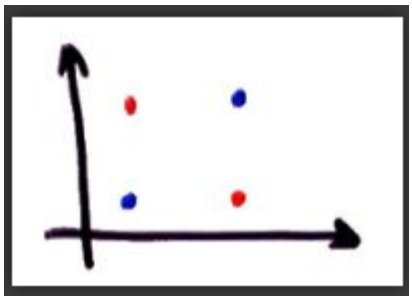
23.K-means is not deterministic and it also consists of number of iterations.
a) True
b) False

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

24. a system that is capable of predicting the future preference of a set of items for a user, and recommend the top items. **[2 M]**
a) **Recommendation Systems**
b) collaborative filtering
Content based Systems
d)all of the above
25 A **content based** recommender works with data that the user provides, either explicitly (rating) or implicitly (clicking on a link). [2 M]

a)true

b)false
26.Is the data linearly separable? [2M]



A) Yes
B) No

Explanation :If you can draw a line or plane between the data points, it is said to be linearly separable.

27.Which of the following are universal approximators? [2M]

A) Kernel SVM
B) Neural Networks
C) Boosted Decision Trees
D) All of the above

All of the above methods can approximate any function.

28. **In which of the following applications can we use deep learning to solve the problem?**

A) Protein structure prediction
B) Prediction of chemical reactions
C) Detection of exotic particles
D) All of these

Solution: D

We can use neural network to approximate any function so it can theoretically be used to solve any problem.

29. **The number of nodes in the input layer is 10 and the hidden layer is 5. The maximum number of connections from the input layer to the hidden layer are**

A) 50
B) Less than 50
C) More than 50
D) It is an arbitrary value

Solution: A

Since MLP is a fully connected directed graph, the number of connections are a multiple of number of nodes in input layer and hidden layer.

30. Decision Tree is a display of an algorithm. [2M]

a) True
b) False

31. 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. [2M]
a) Decision tree
b) Graphs
c) Trees
d) Neural Networks

32. Choose from the following that are Decision Tree nodes?
a) Decision Nodes
b) End Nodes
c) Chance Nodes
d) All of the mentioned

33. **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

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.

34. **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

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.

35. **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.

36. **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.

37. **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.

38. **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.

39. **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

Solution: D

If you search any point on X1 you won't find any point that gives 100% accuracy.

40. **What will be the minimum accuracy you can get?**

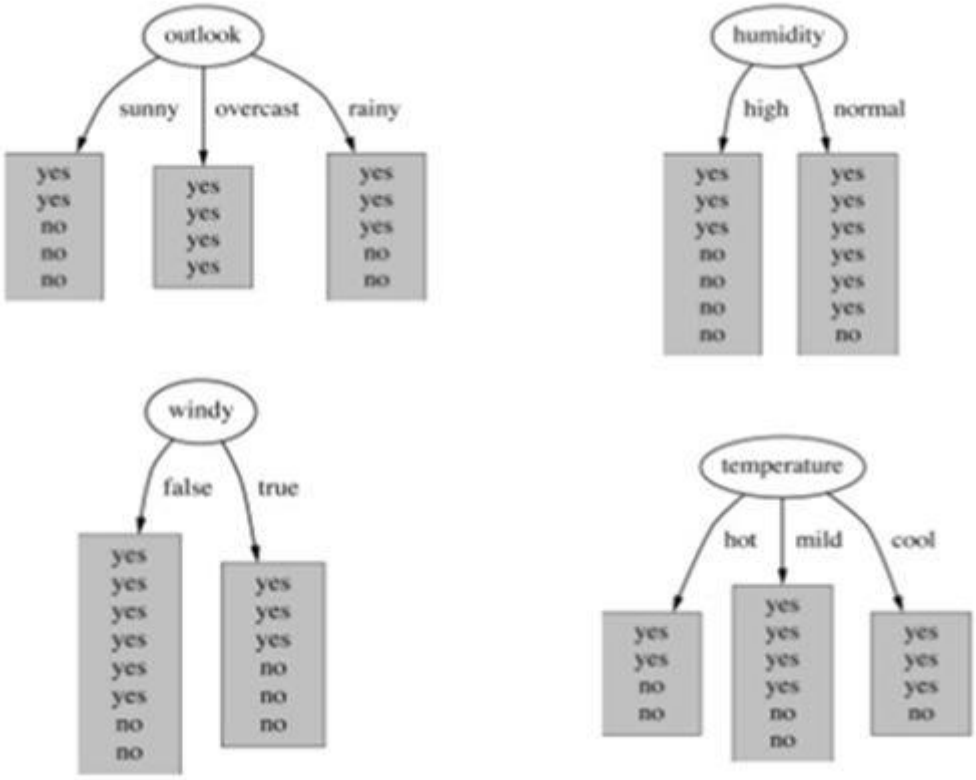
A) Always greater than 70%
B) Always greater than and equal to 70%
C) It can be less than 70%
D) None of these

Solution: 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

Solution: A

Information gain increases with the average purity of subsets. So option A would be the right answer.

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

Solution: C

Both are true and self explanatory

43) 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.

44) 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 variance

Solution: A

Answer is self explanatory

Context 22-23

Suppose, you are building a Gradient Boosting model on data, which has millions of observations and 1000's of features. Before building the model you want to consider the difference parameter setting for time measurement.

45) Consider the hyperparameter "number of trees" and arrange the options in terms of time taken by each hyperparameter for building the Gradient Boosting model?

Note: remaining hyperparameters are same

1. Number of trees = 100

2. Number of trees = 500
3. Number of trees = 1000

A) 1~2~3
B) 1<2<3

C) 1>2>3
D) None of these

Solution: B

The time taken by building 1000 trees is maximum and time taken by building the 100 trees is minimum which is given in solution B

46) Now, Consider the learning rate hyperparameter and arrange the options in terms of time taken by each hyperparameter for building the Gradient boosting model?

Note: Remaining hyperparameters are same

1. learning rate = 1
2. learning rate = 2
3. learning rate = 3

A) 1~2~3
B) 1<2<3

C) 1>2>3
D) None of these

Solution: A

Since learning rate doesn't affect time so all learning rates would take equal time.

47) 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

Solution: C

Learning rate should be low but it should not be very low otherwise algorithm will take so long to finish the training because you need to increase the number trees.

48) [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

Solution: A

49) When you use the boosting algorithm you always consider the weak learners. Which of the following is the main reason for having weak learners?

1. To prevent overfitting
2. To prevent under fitting

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

Solution: A

To prevent overfitting, since the complexity of the overall learner increases at each step. Starting with weak learners implies the final classifier will be less likely to overfit.

50) 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

51) How to select best hyperparameters 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.

52) 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.

53) 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.

https://forms.gle/AF4uUJXKiH6sceyQ9-yam test link