Seat No -

Total number of questions : 60

11930_MACHINE LEARNING

Time : 1hr

Max Marks : 50

N.B

1) All questions are Multiple Choice Questions having single correct option.

2) Attempt any 50 questions out of 60.

3) Use of calculator is allowed.

4) Each question carries 1 Mark.

5) Specially abled students are allowed 20 minutes extra for examination.

6) Do not use pencils to darken answer.

7) Use only black/blue ball point pen to darken the appropriate circle.

8) No change will be allowed once the answer is marked on OMR Sheet.

9) Rough work shall not be done on OMR sheet or on question paper.

10) Darken ONLY ONE CIRCLE for each answer.

Q.no 1. Which of these is not a frequent pattern mining algorithm?

A : Apriori

B : FP growth

C : Decision trees

D : Eclat

Q.no 2. A table with all possible value of a random variable and its corresponding probabilities is called

A : Probability Mass Function

B : Probability Density Function

C : Cumulative distribution function

D : Probability Distribution

Q.no 3. If you use an ensemble of different base models, is it necessary to tune the hyper parameters of all base models to improve the ensemble performance?

A : Yes

B : No

C : Can't Say

D : May be

Q.no 4. Following are the descriptive models

A : Clustering

B : Classification

C : Association rule

D : Classification and Association Rule

Q.no 5. What is the minimum no. of variables/ features required to perform clustering?

A : 0

B : 1

C : 2

D : 3

Q.no 6. The shape of the Normal Curve is ----

A : Spiked

B : Flat

C : Circular

D : Bell shaped

Q.no 7. In random experiment, observations of random variable are classified as -----

A : Events

B : Composition

C : Trials

D : Functions

Q.no 8. Movie Recommendation systems are an example of:

A : Classification and Clustering

B : Clustering and Reinforcement Learning

C : Reinforcement Learning and Regression

D : Regression and classification

Q.no 9. In a Binomial Distribution, the mean and variance are equal.

A : Yes

B : No

C : Can't Say

D : May be

Q.no 10. The weight of persons in a state is a ------

A : Continuous random variable

B : Discrete random variable

C : Irregular random variable

D : Uncertain random variable

Q.no 11. It is not necessary to have a target variable for applying dimensionality reduction algorithms

A : True

B : false

C : Can't Say

D : May be

Q.no 12. The difference between the actual Y value and the predicted Y value found using a regression equation is called the

A : slope

B : residual

C : outlier

D : scatter plot

Q.no 13. Which of the following is the types of supervised learning

A : Classification

B : Clustering

C : Reinforcement Learning

D : k-means

Q.no 14. A perceptron adds up all the weighted inputs it receives, and if it exceeds a certain value, it outputs a 1, otherwise it just outputs a 0.

A : true

B : False

C : Sometimes – it can also output intermediate values as well

D : Can't say

Q.no 15. Mean and variance of Poisson's distribution is the same.

A : Yes

B : No

C : Can't Say

D : May be

Q.no 16. Variance of a constant 'a' is

A : 0

B : a

C : a/2

D : 1

Q.no 17. K means and K-medioids are example of which type of clustering method?

A : partition

B : Hierarchical

C : probabilistic

D : they are not clustering methods

Q.no 18. How many coefficients do you need to estimate in a simple linear regression model

A : 1

B : 2

C : 3

D : 4

Q.no 19. What are closed itemsets?

A : An itemset for which at least one proper super-itemset has same support

B : An itemsetwhose no proper super-itemset has same support

C : An itemset for which at least super-itemset has same confidence

D : An itemsetwhose no proper super-itemset has same confidence

Q.no 20. The classification is considered to be

A : Supervised Learning

B : Unsupervised Learning

C : semi-Supervised Learning

D : Deep Learning

Q.no 21. In the mathematical Equation of Linear Regression Y = Theta_0 + Theta_1 * x + epsilon, (Theta_0,Theta_) refers to _

A : (X-intercept, Slope)

B : (Slope, X-Intercept)

C : (Y-Intercept, Slope)

D : (slope, Y-Intercept)

Q.no 22. What is gini index?

A : It is a type of index structure

B : It is a measure of purity

C : It is an index as well as measure of purity

D : Nothing related to ML

Q.no 23. Which is example of Multi-Class Classification?

A : Plant species classification

B : Social network analysis

C : Medical imaging

D : Email spam detection

Q.no 24. Application of machine learning methods to large databases is called

A : Data Mining.

B : Artificial Intelligence

C : Big Data Computing

D : Internet of Things

Q.no 25. Like the probabilistic view, the _____ view allows us to associate a probability of membership with each classification.

A : exampler

B : deductive

C : classical

D : inductive

Q.no 26. In terms of the bias-variance decomposition, a 1-nearest neighbor classifierhas than a 3-nearest neighbor classifierer.

A : higher variance

B : higher bias

C : lower variance

D : Higher Bias & Lower variance

Q.no 27. 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 inside the sample on test data

C : To find the maxima or minima at the local point

D : Normalize the data

Q.no 28. The VC dimension of hypothesis space H1 is larger than the VC dimension of hypothesis space H2. Which of the following can be inferred from this?

A : The number of examples required for learning a hypothesis in H1 is larger than the number of examples required for H2.

B : The number of examples required for learning a hypothesis in H1 is smaller than the number of examples required for H2.

C : No relation to number of samples required for PAC learning

D : The number of examples required for learning a hypothesis in H1 is smaller than the number of examples required for H1.

Q.no 29. Support Vector Machine is ----

A : Logical Model

B : Proababilistic Model

C : Geometric Model

D : Neural Network model

Q.no 30. Lasso can be interpreted as least-squares linear regression where

A : weights are regularized with the L1 norm

B : the weights have a Gaussian prior

C : weights are regularized with the L2 norm

D : the solution algorithm is simpler

Q.no 31. Having multiple perceptrons can actually solve the XOR problem satisfactorily: this is because each perceptron can partition off a linear part of the space itself, and they can then combine their results.

A : True – this works always, and these multiple perceptrons learn to classify even complex problems

B : False – perceptrons are mathematically incapable of solving linearly inseparable functions, no matter what you do

C : True – perceptrons can do this but are unable to learn to do it – they have to be explicitly hand-coded

D : False – just having a single perceptron is enough

Q.no 32. The Classification predictions can be evaluated using _____

A : accuracy

B : Error rate

C : root mean squared error

D : Value of K

Q.no 33. Which of the following is a good test dataset characteristic? (A) Large enough to yield meaningful results (B) Is representative of the dataset as a whole

A : A only

B : B only

C : Both A and B

D : None of A and B

Q.no 34. 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

Q.no 35. The Both PCA and Lasso can be used for feature selection. Which of the following statements are true?

A : Lasso selects a subset (not necessarily a strict subset) of the original features

B : PCA and Lasso both allow you to specify how many features are chosen

C : PCA produces features that are non linear combinations of the original features

D : PCA and Lasso are the same if you use the kernel trick

Q.no 36. Supervised learning and unsupervised clustering both require which is correct according to the statement.

A : output attribute.

B : hidden attribute.

C : input attribute.

D : categorical attribute

Q.no 37. A person trained to interact with a human expert in order to capture their knowledge.

A : knowledge programmer

B : knowledge developer
r

C : knowledge engineer

D : knowledge extractor

Q.no 38. E(X) = n*p*q is for which distribution?

A : Bernoulli's

B : Binomial

C : Poisson's

D : Normal

Q.no 39. What is back propagation?

A : It is another name given to the curvy function in the perceptron

B : It is the transmission of error back through the network to adjust the inputs

C : It is the transmission of error back through the network to allow weights to be adjusted so that the network can learn

D : None of the mentioned

Q.no 40. Linear Regression has dependent variables that have _____

A : K-Values

B : continuous values.

C : N-Values

D : Square Values

Q.no 41. If the probability that a bomb dropped from a place will strike the target is 60% and if 10 bombs are dropped, find mean and variance?

A : 0.4, 0.25

B : 6,24

C : 0.4, 0.17

D : 0.6, 0.17

Q.no 42. Which of the following, specifies the prior probability of each utterance?

A : Sound Model

B : Language Model

C : Visual Model

D : System Model

Q.no 43. Among the following, is viewed as problem of probabilistic inference?

A : Speech recognition

B : Speaking

C : Hearing

D : Utterance

Q.no 44. What is Morphological Segmentation?

A : Does Discourse Analysis

B : Separate words into individual morphemes and identify the class of the morphemes

C : Is an extension of propositional logic

D : Automatic Text Summarization

Q.no 45. Wrapper methods are hyper-parameter selection methods that

A : Should be used whenever possible because they are computationally efficient

B : Should be avoided unless there are no other options because they are always prone to overfitting.

C : Are useful mainly when the learning machines are "black boxes"

D : Should be avoided altogether.

Q.no 46. Find the mean and variance of X? Where x={0, 1, 2, 3, 4} and f(x) = { 1/9, 2/9, 3/9, 2/9, 1/9}

A : 2 ,4 / 3

B : 3,4/3

C : 2,2/3

D : 3,3/3

Q.no 47. 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 2

B : 1 and 3

C : 2 and 3

D : 1,2 and 3

Q.no 48. What is the expectation of X? Where x={0, 1, 2, 3} and f(x) = { 1/6, 2/6, 2/6, 1/6}

A : 0.5

B : 1.5

C : 2.5

D : 3.5

Q.no 49. What is a heuristic function?

A : A function to solve mathematical problems

B : A function which takes parameters of type string and returns an integer value

C : A function whose return type is nothing

D : A function that maps from problem state descriptions to measures of desirability

Q.no 50. If P(x) = 0.5 and x = 4, then E(x) = ?

A : 1

B : 0.5

C : 4

D : 2

Q.no 51. Different learning methods does not include?

A : Memorization

B : Analogy

C : Deduction

D : Introduction

Q.no 52. Knowledge may be

A : Declarative and procedural

B : Declarative and non-procedural

C : Procedural and non-procedural

D : Declarative, procedural and non-procedural

Q.no 53. 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

Q.no 54. Which of the following is a reasonable way to select the number of principal components "k"?

A : Choose k to be the smallest value so that at least 99% of the varinace is retained. - answer

B : Choose k to be 99% of m (k = 0.99*m, rounded to the nearest integer).

C : Choose k to be the largest value so that 99% of the variance is retained.

D : Use the elbow method

Q.no 55. In a Binomial Distribution, if 'n' is the number of trials and 'p' is the probability of success, then the mean value is given by

A : n*p

B : n

C : p

D : n*p*(1-p)

Q.no 56. Suppose your model is demonstrating high variance across the different training sets. Which of the following is NOT valid way to try and reduce the variance?

A : Increase the amount of traning data in each traning set

B : Improve the optimization algorithm being used for error minimization.

C : Decrease the model complexity

D : Reduce the noise in the training data

Q.no 57. Which of the following is direct application of frequent itemset mining?

A : Social Network Analysis

B : Market Basket Analysis

C : Outlier Detection

D : Intrusion Detection

Q.no 58. You are given sesimic data and you want to predict next earthquake , this is an example of

A : Supervised learning

B : Reinforcement learning

C : Unsupervised learning

D : Dimensionality reduction

Q.no 59. 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 : 1,2 and 3

Q.no 60. What does FP growth algorithm do?

A : It mines all frequent patterns through pruning rules with lesser support

B : It mines all frequent patterns through pruning rules with higher support

C : It mines all frequent patterns by constructing a FP tree

D : It mines all frequent patterns without any pruning rules with higher support

Q.no 1. Which of the following is true?

A : Both apriori and FP-Growth uses horizontal data format

B : Both apriori and FP-Growth uses vertical data format

C : Apriori uses horizontal and FP-Growth uses vertical data format

D : Apriori uses vertical and FP-Growth uses horizontal data format

Q.no 2. In the regression equation Y = 75.65 + 0.50*X, the intercept is

A : 0.5

B : 75.65

C : 1

D : indeterminable

Q.no 3. 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

Q.no 4. In Standard normal distribution, the value of median is ----

A : 2

B : 1

C : 0

D : Not Fixed

Q.no 5. In boosting, individual base learners can be parallel.

A : Yes

B : No

C : Can't Say

D : May be

Q.no 6. The apriori algorithm works in a ----- and ------ fashion?

A : top-down and depth-first

B : top-down and breath-first

C : bottom-up and depth-first

D : bottom-up and breath-first

Q.no 7. If machine learning model output involves target variable then that model is called as

A : Descriptive model

B : Predictive Model

C : Reinforcement Learning

D : All of the above

Q.no 8. VC dimension stands for

A : Vepin Chervo

B : vapnik chervonenkis

C : Vebinic cheronic

D : Velocory Cherumal

Q.no 9. Impact of high variance on the training set ?

A : overfitting

B : underfitting

C : both underfitting & overfitting

D : Depents upon the dataset

Q.no 10. 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

Q.no 11. A random variable that assumes a finite or a countably infinite number of values is called -----

A : Continuous random variable

B : Discrete random variable

C : Irregular random variable

D : Uncertain random variable

Q.no 12. Ensembles will yield bad results when there is significant diversity among the models.

A : Yes

B : No

C : Can't Say

D : May be

Q.no 13. Mutually Exclusive events _____

A : Contain all sample points

B : Contain all common sample points

C : Does not contain any sample point

D : Does not contain any common sample point

Q.no 14. What are closed frequent itemsets?

A : A closed itemset

B : A frequent itemset

C : An itemset which is both closed and frequent

D : An empty set

Q.no 15. In simple term, machine learning is
(A) Training based on historical data
(B) Prediction to answer a query

A : A only

B : B only

C : Both A and B

D : None of A and B

Q.no 16. Normal Distribution is symmetric is about ----

A : Variance

B : Standard Deviation

C : Mean

D : Covariance

Q.no 17. Linear Regression is a _____ machine learning algorithm.

A : Supervised

B : Unsupervised

C : Semi-Supervised

D : Can't say

Q.no 18. If machine learning model output doesnot involves target variable then that model is called as

A : Descriptive model

B : Predictive Model

C : Reinforcement Learning

D : All of the above

Q.no 19. Out of the following values, which one is not possible in probability?

A : 1

B : 0

C : 0.5

D : -0.5

Q.no 20. What are maximal frequent itemsets?

A : A frequent itemset whose no super-itemset is frequent

B : A frequent itemset whose super-itemset is also frequent

C : A non-frequent itemset whose super-itemset is frequent

D : A non-frequent itemset whose super-itemset is also non-frequent

Q.no 21. Which of the following indicates the fundamental of least squares?

A : arithmetic mean should be maximized

B : arithmetic mean should be zero

C : arithmetic mean should be neutralized

D : arithmetic mean should be minimized

Q.no 22. 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 : It is not used anywhere

Q.no 23. You trained a binary classifier model which gives very high accuracy on the training data, but much lower accuracy on validation data. Which is false.

A : This is an instance of overfitting

B : This is an instance of underfitting

C : The training was not well regularized

D : The training and testing examples are sampled from different distributions

Q.no 24. Which is example of Binary classification?

A : Social network analysis

B : Medical imaging

C : Email spam detection

D : Image segmentation

Q.no 25. Which of the following sentence is FALSE regarding regression?

A : It relates inputs to outputs.

B : It is used for prediction.

C : It may be used for interpretation.

D : It discovers causal relationships

Q.no 26. 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

Q.no 27. How do we know if we are underfitting?

A : If by increasing capacity we decrease generalization error

B : If the error representing the training set is relatively small

C : If by increasing capacity we increase generalization error

D : generalization error is large

Q.no 28. Chance Nodes are represented by _____

A : disk

B : square

C : circle

D : triangle

Q.no 29. For a given data having 100 examples,if squared SE1,SE2,and SE3 are 13.33,3.33,and 4.00respectively,calculate Mean Squared Error(MSE)?

A : 0.663

B : 0.2066

C : 0.345

D : 0.567

Q.no 30. Some telecommunication company wants to segment their customers into distinct groups ,this is an example of

A : Supervised learning

B : Reinforcement learning

C : Unsupervised learning

D : Data extraction

Q.no 31. The error function most suited for gradient descent using logistic regression is

A : The entropy function.

B : The squared error.

C : The cross-entropy function.

D : The number of mistakes.

Q.no 32. Which is the Decision tree creation algorithm?

A : K-Means

B : ID3

C : K--medoid

D : Naive Bayes

Q.no 33. What are support vectors?

A : All the examples that have a non-zero weight αk in a SVM

B : The only examples necessary to compute f(x) in an SVM.

C : All of the above

D : None of the above

Q.no 34. Which algorithm that can be used for binary classification?

A : K-Means

B : Naive Bayes

C : K--medoid

D : Hierarchical

Q.no 35. Neural Networks are complex _____ with many parameters.

A : Linear Functions

B : Nonlinear Functions

C : Discrete Functions

D : Exponential Functions

Q.no 36. What are the reasons for overfitting?

A : Small number of features

B : Training set is too large

C : Noisy data & large number of features

D : Testing data is to small

Q.no 37. Which of the following classifications would best suit the student performance classification systems?

A : if then analysis

B : Market Basket Analysis

C : Regression analysis

D : Cluster analysis

Q.no 38. Decision Nodes are represented by _____

A : disk

B : square

C : circle

D : triangle

Q.no 39. What are two steps of tree pruning work?

A : Pessimistic pruning and post pruning

B : Postpruning and Prepruning

C : Cost complexity pruning and time complexity pruning

D : Pessimistic pruning and Optimistic pruning

Q.no 40. If X and Y in a regression model are totally unrelated,

A : the correlation coefficient would be -1

B : the coefficient of determination would be 0

C : the coefficient of determination would be 1

D : the SSE would be 0

Q.no 41. Type of matrix decomposition model is

A : Descriptive model

B : Predictive Model

C : Logical model

D : Geometrical model

Q.no 42. In a Binomial Distribution, if p, q and n are probability of success, failure and number of trials respectively then variance is given by

A : n*p

B : n*p*q

C : p

D : n*p*(1-p)

Q.no 43. What do you mean by support(A)?

A : Total number of transactions containing A

B : Total number of transactions not containing A

C : Number of transactions containing A / Total number of transactions

D : Number of transactions not containing A / Total number of transactions

Q.no 44. What is the relation between candidate and frequent itemsets?

A : A candidate itemset is always a frequent itemset

B : A frequent itemset must be a candidate itemset

C : No relation betweencandidate and frequent

D : candidate and frequent are same

Q.no 45. Point out the wrong statement.

A : Regression through the origin yields an equivalent slope if you center the data first

B : Normalizing variables results in the slope being the correlation

C : Least squares is not an estimation tool

D : None of the mentioned

Q.no 46. Find lambda in Poisson's distribution if the probabilities of getting a head in biased coin toss as 3/4 and 6 coins are tossed.

A : 3.5

B : 4.5

C : 5.5

D : 6.5

Q.no 47. In the example of predicting number of babies based on stork's population ,Number of babies is

A : outcome

B : feature

C : observation

D : attribute

Q.no 48. What does Apriori algorithm do?

A : It mines all frequent patterns through pruning rules with lesser support

B : It mines all frequent patterns through pruning rules with higher support

C : It mines all frequent patterns through pruning rules without support

D : It mines all frequent patterns without any pruning rules with higher support

Q.no 49. A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college. Here feature type is ----

A : nominal

B : ordinal

C : categorical

D : boolean

Q.no 50. Which of the folllowing is an example of feature extraction?

A : Construction bag of words from an email

B : Applying PCA to project high dimensional data

C : Removing stop words

D : Forward selection

Q.no 51. Let S1 and S2 be the set of support vectors and w1 and w2 be the learnt weight vectors for a linearly
separable problem using hard and soft margin linear SVMs respectively. Which of the following are correct?

A : S1 is subset of S2

B : S1 may not be a subset of S2

C : w1 = w2

D : All of the above

Q.no 52. Indicate which one is a method of density estimation

A : Histogram based

B : Branch and bound procedure

C : Neighborhood distance

D : Elbow method

Q.no 53. 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

Q.no 54. Which of the following methods can not achieve zero training error on any linearly separable dataset?

A : Decision tree

B : 15-nearest neighbors

C : Hard-margin SVM

D : Perceptron

Q.no 55. For a Poisson Distribution, if mean(m) = 1, then P(1) is?

A : Indeterminate

B : e

C : e/2

D : 1/e

Q.no 56. Suppose, you have 2000 different models with their predictions and want to ensemble predictions of best x models. Now, which of the following can be a possible method to select the best x models for an ensemble?

A : Step wise forward selection

B : Step wise backward elimination

C : Step wise forward and backward elimination

D : Gradiant Descent

Q.no 57. When do you consider an association rule interesting?

A : If it only satisfies min_support

B : If it only satisfies min_confidence

C : If it satisfies both min_support and min_confidence

D : There are no measures to check so

Q.no 58. 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

Q.no 59. If E(x) = 2 and E(z) = 4, then E(z – x) =?

A : 2

B : 6

C : 0

D : can't say

Q.no 60. What is not true about FP growth algorithms?

A : It mines frequent itemsets without candidate generation

B : There are chances that FP trees may not fit in the memory

C : FP trees are very expensive to build

D : It expands the original database to build FP trees.