

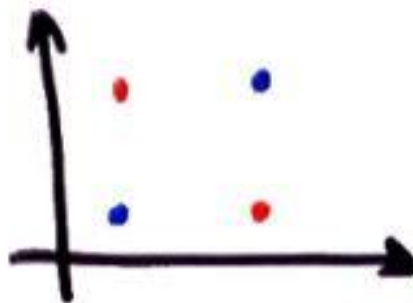
ML MCQs

1. What is Machine learning?
 - a. **The autonomous acquisition of knowledge through the use of computer programs**
 - b. The autonomous acquisition of knowledge through the use of manual programs
 - c. The selective acquisition of knowledge through the use of computer programs
 - d. The selective acquisition of knowledge through the use of manual programs
2. 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**
3. Different learning methods does not include?
 - a. Memorization
 - b. Analogy
 - c. Deduction
 - d. **Introduction**
4. In language understanding, the levels of knowledge that does not include?
 - a. Phonological
 - b. Syntactic
 - c. **Empirical**
 - d. Logical
5. In which algorithm computation time is required more to test unseen samples?
 - a. **K-Nearest Neighbours**
 - b. Decision Tree
 - c. SVM
 - d. Neighbourhood
6. Cluster quality depends on _____ intra-class distance and _____ inter-class distance.
 - a. average, minimum
 - b. **minimum, maximum**
 - c. maximum, minimum
 - d. minimum, average
7. The K-means clustering algorithm is not sensitive to outliers.
 - a. True
 - b. **False**
8. Gini Index would be _____ if dataset is perfectly classified.
 - a. **0**
 - b. 1
 - c. $\frac{1}{2}$
 - d. $\frac{1}{3}$
9. If the samples are an equally divided by target classes, it has entropy of _____.
 - a. 0
 - b. **1**
 - c. $\frac{1}{2}$
 - d. $\frac{1}{3}$

10. In machine learning, most of the applied features need to be identified by an expert before feeding to an algorithm compared to deep learning
- True
 - False**
11. If the data is skewed, _____ is a better measure of central tendency
- Mean
 - Mode
 - Median**
 - None of the above
12. How do you handle missing or corrupted data in a dataset?
- Drop missing rows or columns
 - Replace missing values with mean/median/mode
 - Assign a unique category to missing values
 - All of the above**
13. Consider a linear-regression model, $Y = mX + C$. Values of $X = [2,4,5,6]$ and $Y = [4,6,7,8]$. Calculate MSE loss. Take, $m = 0.5$ and $b = 0.2$
- 15.95**
 - 63.81
 - 95.15
 - 23.04
14. Find the variance for the following set of data representing trees in California (heights in feet): 6, 21, 98, 200, 18, 10
- 5744.36
 - 5447.63
 - 5497.63
 - 5947.36**
15. [True or False] k-NN algorithm does more computation on test time rather than train time.
- TRUE**
 - FALSE
16. What would be the relation between the time taken by 1-NN, 2-NN, 3-NN.
- 1-NN > 2-NN > 3-NN
 - 1-NN < 2-NN < 3-NN
 - 1-NN ~ 2-NN ~ 3-NN**
 - None of these
17. Following are the two statements given for k-NN algorithm, which of the statement(s) is/are true?
- (1) We can choose optimal value of k with the help of cross validation
 - (2) Euclidean distance treats each feature as equally important
- 1
 - 2
 - 1 and 2**
 - None of these
18. Which of the following value of k in the following graph would you give least leave one out cross validation accuracy?



- a. 1
 - b. 2**
 - c. 3
 - d. 5
19. Which of the following is true about model capacity (where model capacity means the ability of neural network to approximate complex functions)?
- a. As number of hidden layers increase, model capacity increases**
 - b. As dropout ratio increases, model capacity increases
 - c. As learning rate increases, model capacity increases
 - d. None of these
20. In a neural network, knowing the weight and bias of each neuron is the most important step. If you can somehow get the correct value of weight and bias for each neuron, you can approximate any function. What would be the best way to approach this?
- a. Assign random values and pray to God they are correct
 - b. Search every possible combination of weights and biases till you get the best value
 - c. Iteratively check that after assigning a value how far you are from the best values, and slightly change the assigned values values to make them better**
 - d. None of these
21. Is the data linearly separable?

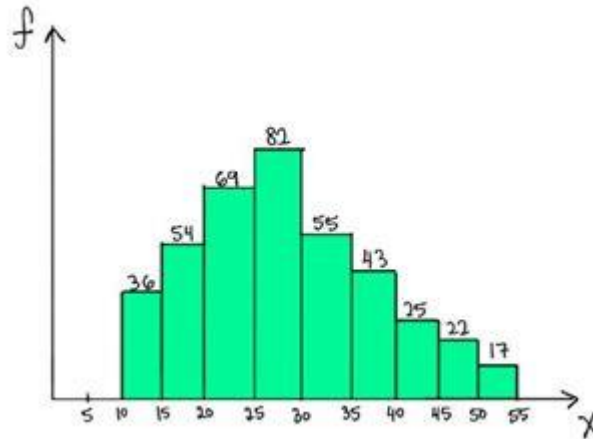


- a. Yes
 - b. No**
22. Which of the following option would you more likely to consider iterating SVM next time?
- a. You want to increase your data points
 - b. You want to decrease your data points
 - c. You will try to calculate more variables**
 - d. You will try to reduce the features
23. True-False: It is possible to design a Linear regression algorithm using a neural network?
- a. True**

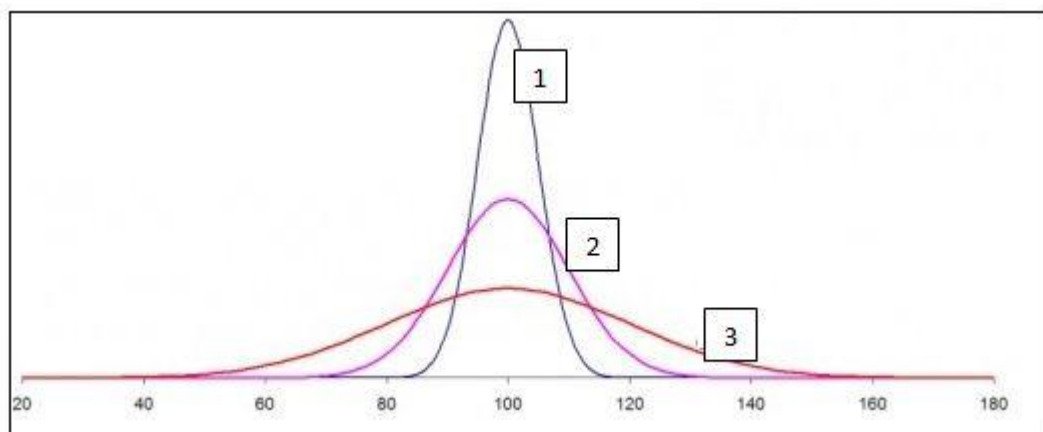
- b. False
- 24. True- False: Overfitting is more likely when you have huge amount of data to train?
 - a. True
 - b. False**
- 25. Choose the option which describes bias in best manner.
 - a. In case of very large x ; bias is low
 - b. In case of very large x ; bias is high**
 - c. We can't say about bias
 - d. None of these
- 26. 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
- 27. an activation function decides whether a neuron should be fired or not [True or False]
 - a. True**
 - b. False
- 28. Gradient Descent is an optimal algorithm to minimize the cost function or to minimize an error.
 - a. True**
 - b. False
- 29. There is feedback in final stage of backpropagation algorithm?
 - a. Yes
 - b. No**
- 30. Which of these measures are used to analyze the central tendency of data?
 - a. Mean and Normal Distribution
 - b. Mean, Median and Mode**
 - c. Mode, Alpha & Range
 - d. Standard Deviation, Range and Mean
 - e. Median, Range and Normal Distribution
- 31. A test is administered annually. The test has a mean score of 150 and a standard deviation of 20. If Ravi's z-score is 1.50, what was his score on the test?
 - a. 180**
 - b. 130
 - c. 30
 - d. 150
 - e. None of the above
- 32. Which of the following measures of central tendency will always change if a single value in the data changes?
 - a. Mean**
 - b. Median
 - c. Mode
 - d. All of above
- 33. If a positively skewed distribution has a median of 50, which of the following statement is true?
 - a. Mean is greater than 50
 - b. Mean is less than 50
 - c. Mode is less than 50

- d. Mode is less than 50
- e. **Both A and C**
- f. Both B and D

34. Which of the following is a possible value for the median of the below distribution?

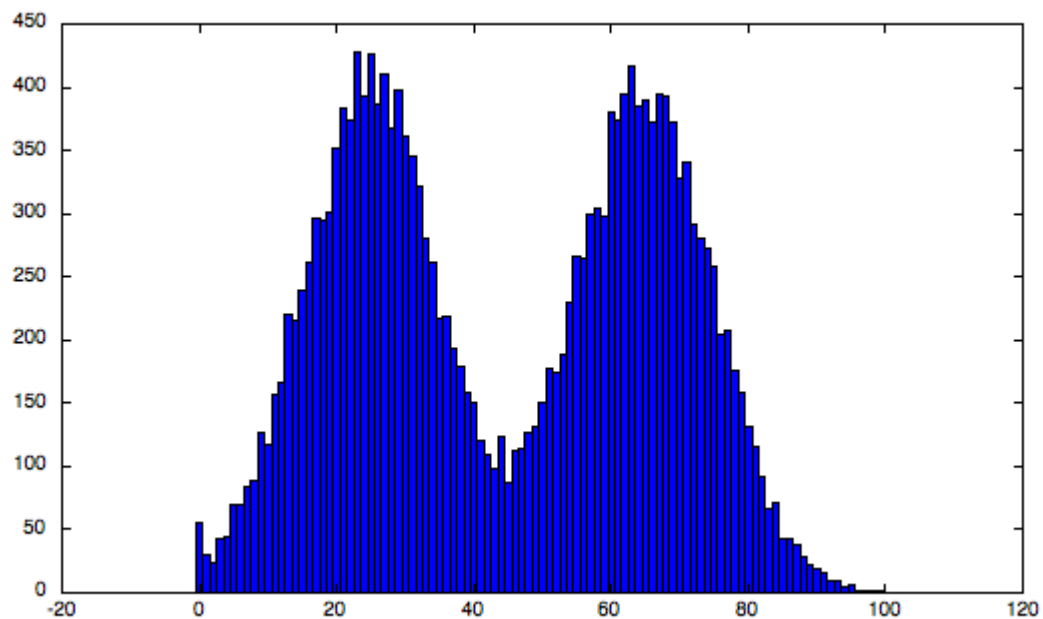


- a. 32
 - b. **26**
 - c. 17
 - d. 40
35. [True or False] Standard deviation can be negative.
- a. True
 - b. **False**
36. Standard deviation is robust to outliers?
- a. True
 - b. **False**
37. For the below normal distribution, which of the following option holds true ?
 σ_1 , σ_2 and σ_3 represent the standard deviations for curves 1, 2 and 3 respectively.



- a. $\sigma_1 > \sigma_2 > \sigma_3$
- b. **$\sigma_1 < \sigma_2 < \sigma_3$**
- c. $\sigma_1 = \sigma_2 = \sigma_3$
- d. None

38. What happens when we introduce more variables to a linear regression model?
- The r squared value may increase or remain constant, the adjusted r squared may increase or decrease.**
 - The r squared may increase or decrease while the adjusted r squared always increases.
 - Both r square and adjusted r square always increase on the introduction of new variables in the model.
 - Both might increase or decrease depending on the variables introduced.
39. Which of the following is true about below given histogram?



- Above histogram is unimodal
 - Above histogram is bimodal**
 - Given above is not a histogram
 - None of the above
40. 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
41. Which of the following is true about Residuals ?
- Lower is better**
 - Higher is better
 - A or B depend on the situation
 - None of these
42. Choose the option which describes bias in best manner.
- In case of very large x ; bias is low
 - In case of very large x ; bias is high**
 - We can't say about bias

- d. None of these
- 43. 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
- 44. If the demand is 100 during October 2016, 200 in November 2016, 300 in December 2016, 400 in January 2017. What is the 3-month simple moving average for February 2017?
 - a. **300**
 - b. 350
 - c. 400
 - d. Need more information
- 45. 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**
- 46. To find the minimum or the maximum of a function, we set the gradient to zero because
 - a. **The value of the gradient at extrema of a function is always zero**
 - b. Depends on the type of problem
 - c. Both A and B
 - d. None of the above
- 47. The most widely used metrics and tools to assess a classification model are
 - a. Confusion matrix
 - b. Cost-sensitive accuracy
 - c. Area under the ROC curve
 - d. **All of the above**
- 48. 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
- 49. 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**
- 50. 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**
 - d. None of the above
- 51. 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 options

52. Which of the following statement(s) is / are true for Gradient Decent (GD) and Stochastic Gradient Decent (SGD)?

Statement 1: In GD and SGD, you update a set of parameters in an iterative manner to minimize the error function.

Statement 2: In SGD, you have to run through all the samples in your training set for a single update of a parameter in each iteration.

Statement 3: In GD, you either use the entire data or a subset of training data to update a parameter in each iteration.

- a. **Only 1**
- b. Only 2
- c. Only 3
- d. All the statements

53. "Convolutional Neural Networks can perform various types of transformation (rotations or scaling) in an input". Is the statement correct True or False?

- a. True
- b. **False**

54. Which of the following techniques perform similar operations as dropout in a neural network?

- a. **Bagging**
- b. Boosting
- c. Stacking
- d. None of these

55. If you increase the number of hidden layers in a Multi Layer Perceptron, the classification error of test data always decreases. True or False?

- a. True
- b. **False**

56. What is the sequence of the following tasks in a perceptron?

- 1) Initialize weights of perceptron randomly
- 2) Go to the next batch of dataset
- 3) If the prediction does not match the output, change the weights
- 4) For a sample input, compute an output

- a. 1, 2, 3, 4
- b. 4, 3, 2, 1
- c. 3, 1, 2, 4
- d. **1, 4, 3, 2**

57. Can a neural network model the function ($y=1/x$)?

- a. **Yes**
- b. No

58. In which neural net architecture, does weight sharing occur?

- a. Convolutional neural Network
- b. Recurrent Neural Network
- c. Fully Connected Neural Network
- d. **Both A and B**

59. The number of neurons in the output layer should match the number of classes (Where the number of classes is greater than 2) in a supervised learning task. True or False?

- a. True
 - b. False**
60. In a neural network, which of the following techniques is used to deal with overfitting?
- a. Dropout
 - b. Regularization
 - c. Batch Normalization
 - d. All of these**
61. What if we use a learning rate that's too large?
- a. Network will converge
 - b. Network will not converge**
 - c. Can't Say
 - d. None of the above
62. When pooling layer is added in a convolutional neural network, translation in-variance is preserved. True or False?
- a. True**
 - b. False
63. Which gradient technique is more advantageous when the data is too big to handle in RAM simultaneously?
- a. Full Batch Gradient Descent
 - b. Stochastic Gradient Descent**
64. What are the factors to select the depth of neural network?
- 1) Type of neural network (eg. MLP, CNN etc)
 - 2) Input data
 - 3) Computation power, i.e. Hardware capabilities and software capabilities
 - 4) Learning Rate
 - 5) The output function to map
- a. 1, 2, 4, 5
 - b. 2, 3, 4, 5
 - c. 1, 3, 4, 5
 - d. All of these**
65. Increase in size of a convolutional kernel would necessarily increase the performance of a convolutional network.
- a. True
 - b. False**
66. Gradient of a continuous and differentiable function
- a. is zero at a minimum
 - b. is zero at a saddle point
 - c. decreases as you get closer to the minimum
 - d. is non-zero at a maximum
 - e. Options a, b, c**
67. K-fold cross-validation is
- a. linear in K**
 - b. quadratic in K
 - c. cubic in K
 - d. exponential in K
68. What strategies can help reduce overfitting in decision trees?

- a. Pruning
 - b. Enforce a minimum number of samples in leaf nodes
 - c. Enforce a maximum depth for the tree
 - d. All of the above**
69. Neural networks
- a. optimize a convex cost function
 - b. can be used for regression as well as classification
 - c. can be used in an ensemble
 - d. option B & C both**
70. Which of the following are true about generative models?
- a. They model the joint distribution $P(\text{class} = C \text{ AND sample} = x)$
 - b. They can be used for classification
 - c. Linear discriminant analysis is a generative model
 - d. All of the above**
71. Which of the following methods can't achieve zero training error on any linearly separable dataset?
- a. Decision tree
 - b. Hard-margin SVM
 - c. K-nearest neighbours**
 - d. Perceptron
72. Which of the following are true about subset selection?

Statement 1: Subset selection can reduce overfitting

Statement 2: Finding the true best subset takes exponential time

Statement 3: Subset selection can substantially decrease the bias of support vector machines

Statement 4: Ridge regression frequently eliminates some of the features

- a. Statement 1 is correct
 - b. Statement 1 & 2 are correct**
 - c. All the statements are correct
 - d. None of the statement is correct
73. In neural networks, nonlinear activation functions such as sigmoid, tanh, and ReLU
- a. speed up the gradient calculation in backpropagation, as compared to linear units
 - b. are applied only to the output units
 - c. help to learn nonlinear decision boundaries**
 - d. always output values between 0 and 1
74. Which of the following can help to reduce overfitting in an SVM classifier?
- a. Use of slack variables**
 - b. Normalizing the data
 - c. High-degree polynomial features
 - d. Setting a very low learning rate
75. Which of the following are true about bagging?

Statement 1: In bagging, we choose random subsamples of the input points with replacement

Statement 2: Bagging is ineffective with logistic regression, because all of the learners learn exactly the same decision boundary

Statement 3: The main purpose of bagging is to decrease the bias of learning algorithms.

Statement 4: If we use decision trees that have one sample point per leaf, bagging never gives lower training error than one ordinary decision tree

- a. **Statement 1 & 4 are correct**
 - b. Statement 1 & 2 are correct
 - c. Statement 1, 2 & 4 are correct
 - d. All the statements are correct
76. Back propagation is a learning technique that adjusts weights in the neural network by propagating weight changes.
- a. Forward from source to sink
 - b. **Backward from sink to source**
 - c. Forward from source to hidden nodes
 - d. Backward from sink to hidden nodes
77. An artificial neuron receives n inputs $x_1, x_2, x_3, \dots, x_n$ with weights w_1, w_2, \dots, w_n attached to the input links. The weighted sum _____ is computed to be passed on to a non-linear filter Φ called activation function to release the output.
- a. $\sum w_i$
 - b. $\sum x_i$
 - c. $\sum w_i + \sum x_i$
 - d. **$\sum w_i * x_i$**
78. A neuron with 3 inputs has the weight vector $[0.2 \ -0.1 \ 0.1]^T$ and a bias $\theta = 0$. If the input vector is $X = [0.2 \ 0.4 \ 0.2]^T$ then the total input to the neuron is:
- a. 0.20
 - b. 1.0
 - c. **0.02**
 - d. -1.0
79. In Delta Rule for error minimization
- a. weights are adjusted w.r.to change in the output
 - b. **weights are adjusted w.r.to difference between desired output and actual output**
 - c. weights are adjusted w.r.to difference between input and output
 - d. none of the above
80. What is the objective of backpropagation algorithm?
- a. to develop learning algorithm for multilayer feedforward neural network
 - b. to develop learning algorithm for single layer feedforward neural network
 - c. **to develop learning algorithm for multilayer feedforward neural network, so that network can be trained to capture the mapping implicitly**
 - d. none of the mentioned
81. The backpropagation law is also known as generalized delta rule, is it true?
- a. **Yes**
 - b. No
82. What is true regarding backpropagation rule?
- a. it is also called generalized delta rule
 - b. error in output is propagated backwards only to determine weight updates
 - c. there is no feedback of signal at any stage
 - d. **all of the mentioned**
83. There is feedback in final stage of backpropagation algorithm?
- a. Yes
 - b. **No**

84. What is true regarding backpropagation rule?
- it is a feedback neural network
 - actual output is determined by computing the outputs of units for each hidden layer**
 - hidden layer output is not all important, they are only meant for supporting input and output layers
 - none of the mentioned
85. What are general limitations of back propagation rule?
- local minima problem
 - slow convergence
 - scaling
 - all of the mentioned**
86. What are the general tasks that are performed with backpropagation algorithm?
- pattern mapping
 - function approximation
 - prediction
 - all of the mentioned**
87. Does backpropagation learning is based on gradient descent along error surface?
- Yes**
 - No
 - cannot be said
 - it depends on gradient descent but not error surface
88. How can learning process be stopped in backpropagation rule?
- there is convergence involved
 - no heuristic criteria exist
 - on basis of average gradient value**
 - none of the mentioned
89. How can false minima be reduced in case of error in recall in feedback neural networks?
- by providing additional units
 - by using probabilistic update**
 - can be either probabilistic update or using additional units
 - none of the mentioned
90. What is objective of linear autoassociative feedforward networks?
- to associate a given pattern with itself**
 - to associate a given pattern with others
 - to associate output with input
 - none of the mentioned
91. The data scientists at "BigMart Inc" have collected 2013 sales data for 1559 products across 10 stores in different cities. Also, certain attributes of each product based on these attributes and store have been defined. The aim is to build a predictive model and find out the sales of each product at a particular store during a defined period.

Which learning problem does this belong to?

- Supervised learning**
 - Unsupervised learning
 - Reinforcement learning
 - None
92. Which methodology does Decision Tree (ID3) take to decide on first split?
- Greedy approach**

- b. Look-ahead approach
- c. Brute force approach
- d. None of these

93. There are 24 predictors in a dataset. You build 2 models on the dataset:

- 1) Bagged decision trees and
- 2) Random forest

Let the number of predictors used at a single split in bagged decision tree is A and Random Forest is B.

Which of the following statement is correct?

- a. **A \geq B**
- b. A < B
- c. A \gg B
- d. Cannot be said since different iterations use different numbers of predictors

94. Why do we prefer information gain over accuracy when splitting?

- a. Decision Tree is prone to overfit and accuracy doesn't help to generalize
- b. Information gain is more stable as compared to accuracy
- c. Information gain chooses more impactful features closer to root
- d. **All of these**

95. Random forests (While solving a regression problem) have the higher variance of predicted result in comparison to Boosted Trees (Assumption: both Random Forest and Boosted Tree are fully optimized).

- a. True
- b. False
- c. **Cannot be determined**

96. Assume everything else remains same, which of the following is the right statement about the predictions from decision tree in comparison with predictions from Random Forest?

- a. Lower Variance, Lower Bias
- b. Lower Variance, Higher Bias
- c. Higher Variance, Higher Bias
- d. **Lower Bias, Higher Variance**

97. Which of the following tree based algorithm uses some parallel (full or partial) implementation?

- a. Random Forest
- b. Gradient Boosted Trees
- c. XGBOOST
- d. **Both A and C**
- e. A, B and C

98. Which of the following is not possible in a boosting algorithm?

- a. **Increase in training error.**
- b. Decrease in training error
- c. Increase in testing error
- d. Decrease in testing error
- e. Any of the above

99. Let's say we have m numbers of estimators (trees) in a boosted tree. Now, how many intermediate trees will work on modified version (OR weighted) of data set?
- 1
 - $M - 1$**
 - M
 - Can't say
 - None of the above
 -
100. Generally, in terms of prediction performance which of the following arrangements are correct:
- Bagging>Boosting>Random Forest>Single Tree
 - Boosting>Random Forest>Single Tree>Bagging
 - Boosting>Random Forest>Bagging>Single Tree**
 - Boosting >Bagging>Random Forest>Single Tree
101. In which of the following application(s), a tree based algorithm can be applied successfully?
- Recognizing moving hand gestures in real time
 - Predicting next move in a chess game
 - Predicting sales values of a company based on their past sales
 - A and B
 - A, B, and C**
102. Which splitting algorithm is better with categorical variable having high cardinality?
- Information Gain
 - Gain Ratio**
 - Change in Variance
 - None of these
103. Suppose we have missing values in our data. Which of the following method(s) can help us to deal with missing values while building a decision tree?
- Let it be. Decision Trees are not affected by missing values
 - Fill dummy value in place of missing, such as -1
 - Impute missing value with mean/median
 - All of these**
104. While creating a Decision Tree, can we reuse a feature to split a node?
- Yes**
 - No
105. Decision Trees are not affected by multicollinearity in features:
- True**
 - False
106. In Random Forest, which of the following is randomly selected?
- Number of decision trees
 - features to be taken into account when building a tree
 - samples to be given to train individual tree in a forest
 - B and C**

- e. A, B and C
107. Which of the following are the disadvantage of Decision Tree algorithm?
- a. Decision tree is not easy to interpret
 - b. Decision tree is not a very stable algorithm
 - c. Decision Tree will over fit the data easily if it perfectly memorizes it
 - d. **Both B and C**
108. What can be the maximum depth of decision tree (where k is the number of features and N is the number of samples)? Our constraint is that we are considering a binary decision tree with no duplicate rows in sample (Splitting criterion is not fixed).
- a. N
 - b. $N - k - 1$
 - c. **$N - 1$**
 - d. $k - 1$
109. Below is a list of parameters of Decision Tree. In which of the following cases higher is better?
- a. Number of samples used for split
 - b. Depth of tree
 - c. Samples for leaf
 - d. **Can't Say**
110. 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
111. 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
112. 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
113. The minimum time complexity for training an SVM is $O(n^2)$. 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
114. The effectiveness of an SVM depends upon:
- a. Selection of Kernel
 - b. Kernel Parameters

- c. Soft Margin Parameter C
 - d. **All of the above**
115. Support vectors are the data points that lie closest to the decision surface.
- a. **True**
 - b. False
116. 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**
117. 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
118. 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
119. What would happen when you use very small C ($C \sim 0$)?
- a. **Misclassification would happen**
 - b. Data will be correctly classified
 - c. Can't say
 - d. None of these
120. 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**
121. 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. How many times we need to train our SVM model in such case?
- a. 1
 - b. 2
 - c. 3
 - d. **4**
122. 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
- 1
 - 2
 - 1 and 2**
 - None of above
123. k-NN algorithm does more computation on test time rather than train time.
- True**
 - False
124. Which of the following distance metric can not be used in k-NN?
- Manhattan
 - Minkowski
 - Tanimoto
 - Jaccard
 - Mahalanobis
 - All can be used**
125. Which of the following option is true about k-NN algorithm?
- It can be used for classification
 - It can be used for regression
 - It can be used in both classification and regression**
126. Which of the following statement is true about k-NN algorithm?
- 1) k-NN performs much better if all of the data have the same scale
 - 2) k-NN works well with a small number of input variables (p), but struggles when the number of inputs is very large
 - 3) k-NN makes no assumptions about the functional form of the problem being solved
- 1 and 2
 - 1 and 3
 - Only 1
 - All of the above**
127. Which of the following machine learning algorithm can be used for imputing missing values of both categorical and continuous variables?
- K-NN**
 - Linear Regression
 - Logistic Regression
128. Which of the following will be Euclidean Distance between the two data point A(1,3) and B(2,3)?
- 1**
 - 2
 - 4
 - 8

129. Which of the following will be true about k in k -NN in terms of Bias?
- When you increase the k the bias will be increases**
 - When you decrease the k the bias will be increases
 - Can't say
 - None of these
130. Which of the following will be true about k in k -NN in terms of variance?
- When you increase the k the variance will increases
 - When you decrease the k the variance will increases**
 - Can't say
 - None of these
131. When you find noise in data which of the following option would you consider in k -NN?
- I will increase the value of k**
 - I will decrease the value of k
 - Noise can not be dependent on value of k
 - None of these
132. In k -NN it is very likely to overfit due to the curse of dimensionality. Which of the following option would you consider to handle such problem?
- 1) Dimensionality Reduction
 - 2) Feature selection
- 1
 - 2
 - 1 and 2**
 - None of above
133. You have given the following 2 statements, find which of these option is/are true in case of k -NN?
- 1) In case of very large value of k , we may include points from other classes into the neighborhood.
 - 2) In case of too small value of k the algorithm is very sensitive to noise
- 1
 - 2
 - 1 and 2**
 - None of these
134. Which of the following algorithm is not an example of an ensemble method?
- Extra Tree Regressor
 - Random Forest
 - Gradient Boosting
 - Decision Tree**

135. 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**
136. 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
137. 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**
138. True or False: Ensemble learning can only be applied to supervised learning methods.
- a. True
 - b. False**
139. 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

140. 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?

Note: We are working on a regression problem

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**

141. Which of the following is true about weighted majority votes?

1. We want to give higher weights to better performing models
2. Inferior models can overrule the best model if collective weighted votes for inferior models is higher than best model
3. Voting is special case of weighted voting

- a. 1 and 3
- b. 2 and 3
- c. 1 and 2
- d. 1, 2 and 3**
- e. None of above

142. Which of the following is true about bagging?

1. Bagging can be parallel
2. The aim of bagging is to reduce bias not variance
3. Bagging helps in reducing overfitting

- a. 1 and 2
- b. 2 and 3
- c. 1 and 3**
- d. All of these

143. True or False: In boosting, individual base learners can be parallel.

- a. True
- b. False**

144. 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. Both**
 - d. None of above
145. In machine learning, an algorithm (or learning algorithm) is said to be unstable if a small change in training data cause the large change in the learned classifiers.
True or False: Bagging of unstable classifiers is a good idea.
- a. True**
 - b. False
146. Can decision trees be used for performing clustering?
- a. True**
 - b. False
147. Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:
- 1 - Capping and flouring of variables
 - 2 - Removal of outliers
- a. 1 only**
 - b. 2 only
 - c. 1 and 2
 - d. None of these
148. What is the minimum no. of variables/ features required to perform clustering?
- a. 0
 - b. 1**
 - c. 2
 - d. 3
149. For two runs of K-Mean clustering is it expected to get same clustering results?
- a. Yes
 - b. No**
150. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means
- a. Yes**
 - b. No
 - c. Can't Say
 - d. None of these

151. Which of the following can act as possible termination conditions in K-Means?

- 1 - For a fixed number of iterations.
- 2 - Assignment of observations to clusters does not change between iterations. Except for cases with a bad local minimum.
- 3 - Centroids do not change between successive iterations.
- 4 - Terminate when RSS falls below a threshold

- a. 1, 3 and 4
- b. 1, 2 and 3
- c. 1, 2 and 4
- d. **All of the above**

152. 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

153. 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

154. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):

- 1 - Creating different models for different cluster groups.
- 2 - Creating an input feature for cluster ids as an ordinal variable.
- 3 - Creating an input feature for cluster centroids as a continuous variable.
- 4 - Creating an input feature for cluster size as a continuous variable.

- a. 1 only
- b. 1 and 2
- c. 1 and 4
- d. 3 only
- e. 2 and 4
- f. **All of the above**

155. In which of the following cases will K-Means clustering fail to give good results?
- 1 - Data points with outliers
 - 2 - Data points with different densities
 - 3 - Data points with round shapes
 - 4 - Data points with non-convex shapes
- a. 1 and 2
 - b. 2 and 3
 - c. 2 and 4
 - d. 1, 2 and 4**
 - e. 1,2,3 and 4
156. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?
- 1) Single-link
 - 2) Complete-link
 - 3) Average-link
- a. 1 and 2
 - b. 1 and 3
 - c. 2 and 3
 - d. 1, 2 and 3**
157. Feature scaling is an important step before applying K-Mean algorithm. What is reason behind this?
- a. In distance calculation it will give the same weights for all features**
 - b. You always get the same clusters. If you use or don't use feature scaling
 - c. In Manhattan distance it is an important step but in Euclidian it is not
 - d. None of these
158. Which of the following method is used for finding optimal of cluster in K-Mean algorithm?
- a. Elbow method
 - b. Manhattan method
 - c. Ecludian mehthod
 - d. All of the above
 - e. None of these
159. What is true about K-Mean Clustering?
- 1 - K-means is extremely sensitive to cluster center initializations
 - 2 - Bad initialization can lead to Poor convergence speed
 - 3 - Bad initialization can lead to bad overall clustering
- a. 1 and 3
 - b. 1 and 2
 - c. 2 and 3

d. 1, 2 and 3

160. What is true about K-Mean Clustering?

- 1 - K-means is extremely sensitive to cluster center initializations
 - 2 - Bad initialization can lead to Poor convergence speed
 - 3 - Bad initialization can lead to bad overall clustering
- a. 1 and 3
 - b. 1 and 2
 - c. 2 and 3
 - d. 1, 2 and 3**

161. Imagine, you have 1000 input features and 1 target feature in a machine learning problem. You have to select 100 most important features based on the relationship between input features and the target features.

Do you think, this is an example of dimensionality reduction?

- a. Yes**
- b. No

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

- a. True**
- b. False

163. Dimensionality reduction algorithms are one of the possible ways to reduce the computation time required to build a model.

- a. True**
- b. False

164. Which of the following algorithms cannot be used for reducing the dimensionality of data?

- a. t-SNE
- b. PCA
- c. LDA
- d. None of these**

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

- 1 - PCA is an unsupervised method
 - 2 - It searches for the directions that data have the largest variance
 - 3 - Maximum number of principal components \leq number of features
 - 4 - All principal components are orthogonal to each other
- a. 1 and 2
 - b. 1 and 3
 - c. 2 and 3
 - d. 1, 2 and 3**

- e. 1,2 and 4
- f. **All of the above**

166. Suppose we are using dimensionality reduction as pre-processing technique, i.e, instead of using all the features, we reduce the data to k dimensions with PCA. And then use these PCA projections as our features. Which of the following statement is correct?

- a. Higher 'k' means more regularization
- b. **Higher 'k' means less regularization**
- c. Can't say

167. Which of the following statement is true for a t-SNE cost function?

- a. It is asymmetric in nature.
- b. **It is symmetric in nature.**
- c. It is same as the cost function for SNE.

168. Which of the following statement is correct for t-SNE and PCA?

- a. t-SNE is linear whereas PCA is non-linear
- b. t-SNE and PCA both are linear
- c. t-SNE and PCA both are nonlinear
- d. **t-SNE is nonlinear whereas PCA is linear**

169. 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

170. What will happen when eigenvalues are roughly equal?

- a. PCA will perform outstandingly
- b. **PCA will perform badly**
- c. Can't Say
- d. None of above

171. PCA works better if there is?

- 1 - A linear structure in the data
 - 2 - If the data lies on a curved surface and not on a flat surface
 - 3 - 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

172. What happens when you get features in lower dimensions using PCA?

- 1 - The features will still have interpretability
- 2 - The features will lose interpretability
- 3 - The features must carry all information present in data

4 - 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**

173. Which of the following option(s) is / are true?

- 1 - You need to initialize parameters in PCA
 - 2 - You don't need to initialize parameters in PCA
 - 3 - PCA can be trapped into local minima problem
 - 4 - 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**

174. Which of the following options are correct, when you are applying PCA on a image dataset?

- 1 - It can be used to effectively detect deformable objects.
 - 2 - It is invariant to affine transforms.
 - 3 - It can be used for lossy image compression.
 - 4 - It is not invariant to shadows.
- a. 1 and 2
 - b. 2 and 3
 - c. **3 and 4**
 - d. 1 and 4

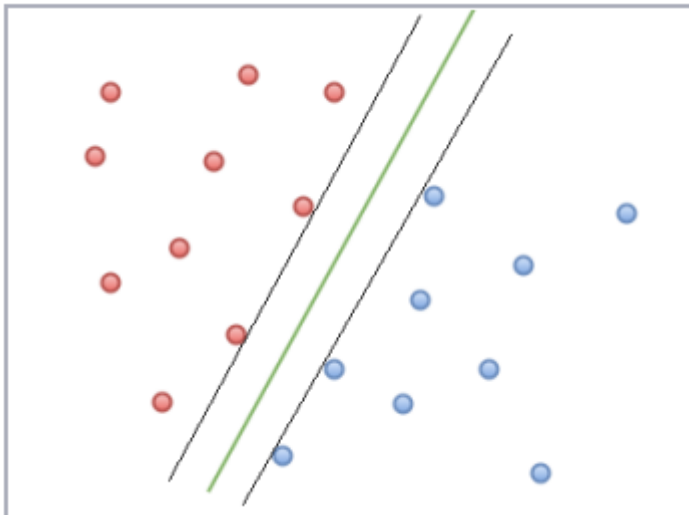
175. A threshold to determine whether an itemset is to be considered for generating Association Rules is called

- a. Support
- b. Confidence
- c. **Minimum Support**
- d. Lift

176. Given the following List of Transactions with Support = 20% and Confidence = 50%, is the rule $A5 \rightarrow (A1, A2)$ a strong rule?

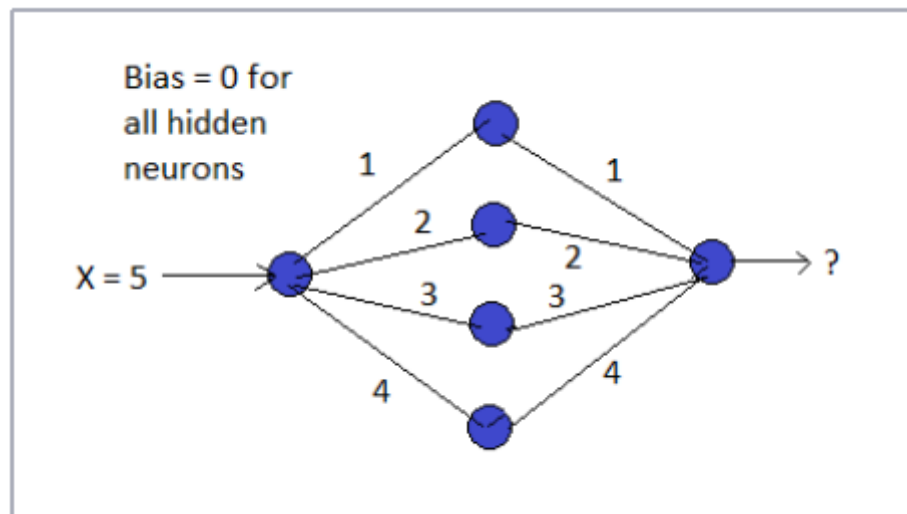
Transaction	Items
1	A1, A2, A5
2	A2, A4
3	A2, A3
4	A1, A2, A4
5	A1, A3
6	A2, A3
7	A1, A2, A3, A5
8	A1, A2, A3
9	A1, A3

- a. **Yes**
 - b. No
 - c. Not Enough information provided to answer the question
177. Which of the following is not a step in performing PCA?
- a. Compute Co-variance Matrix
 - b. Calculate Eigenvectors and Eigenvalues
 - c. **Remove Eigenvectors with highest Eigenvalues**
 - d. Arrange the Eigenvectors in descending order
178. Which of the following is/are Dimensionality Reduction Techniques?
- a. **Principal Component Analysis**
 - b. **Linear Discriminant Analysis**
 - c. Support Vector Machines
 - d. Association Rule Mining
179. SVM can be used to classify linearly inseparable data by
- a. **Transforming it to a higher dimensional space with a kernel**
 - b. Reducing the dimensionality of the data
 - c. Not Possible. You can't classify Linearly Inseparable data using SVM.
 - d. Depends whether it is a binary classification problem or a multi-class classification problem
180. What is the number of Support Vectors in the Given Image?



- a. 19
- b. 14
- c. **5**
- d. 1

181. The number of parameters which are trainable in a Feed Forward Neural Network (the connection weights and the bias terms) with 1 input layer with 3 neurons, 2 hidden layers each with 4 hidden units and an output layer with 1 unit is
- 41**
 - 40
 - 36
 - 32
182. A feed-forward Neural Network has weights between the input layer and the only hidden layer and between the hidden layer and output layer as 1,2,3,4 and all the bias terms of all layers are 0. The activation function is linear and the input is 5. What will be the output?



- 50
 - 150**
 - 25
 - 100
183. What is the formula for output of a single neuron in a Neural Network?
- Activation(Z) where $Z = W * b + X$, X = Input, W = connection weights, b = bias term
 - $Z = \text{Activation}(X)$ followed by $ZX + b$, where X = Input, W = connection weights, b = bias term
 - Activation(Z) where $Z = WX + b$, X = Input, W = connection weights, b = bias term**
 - None of the above
184. If you initialize all the weight and bias terms in a Neural Network to be 0, then there is no point in having multiple neurons in the hidden layer since all the neurons will be doing the same computation.
- True**
 - False
185. Among the following which is not a "Hyperparameter" in a Neural Network?
- Number of Hidden Layers
 - Number of hidden neurons in the hidden layers
 - Learning Rate

d. Training Error

186. The deeper layers of a Convolutional Neural Network typically learn more complex patterns of the input image than the earlier layers.

- a. True
- b. False

187. Which statement(s) is/are correct with respect to Gradient Descent Algorithm?

- a. **Gradient descent is an optimization algorithm used to find the values of parameters of a function that minimizes a cost function.**
- b. Gradient descent is an optimization algorithm used to update the values of hyperparameters of a function whose goal is to find saddle points in highdimensional space.
- c. **The learning rate determines how fast the Gradient Descent Algorithm converges or if it converges at all.**
- d. Calculating the Partial Derivative is an optional step

188. For the given confusion matrix what are the values of Precision for class A and and Recall for class B respectively?

		Predicted	
		A	B
Actual	A	50	10
	B	5	35

- a. **0.90, 0.875**
- b. 0.77, 0.83
- c. 0.23, 0.17
- d. 0.1, 0.125

189. What is the correct formula for F1 Score?

- a. $(\text{Precision} + \text{Recall}) / (\text{Precision} \times \text{Recall})$
- b. $2 \times (\text{Precision} + \text{Recall}) / (\text{Precision} \times \text{Recall})$
- c. **$2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$**
- d. $(\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$

190. Neural Networks are trained in the following order:

- a. Loss Function Calculation > Forward Propagation of Errors > Parameter Initialization > Gradient Descent(Differentiation) > Back Propagation > Parameter Update
- b. **Parameter Initialization > Forward Propogation > Loss Function > Differentiation > Back Propogation of Errors > Parameter Update**
- c. Parameter Initialization > Backward Propagation > Loss Function > Differentiation > Forward Propagation > Parameter Update

- d. Parameter Initialization > Loss Function > Forward Propagation > Differentiation > Back Propagation of Errors > Parameter Re-initialization

191. Which of the following is not a clustering method for Unsupervised Learning?

- a. K-Means
- b. Hierarchical Clustering
- c. Density Based Clustering
- d. Agglomerative Clustering
- e. Randomized Clustering

192. Which one of the following is not a type of learning?

- a. Unsupervised Learning
- b. Superimposed Learning**
- c. Transfer Learning
- d. Reinforcement Learning
- e. Supervised Learning

193. Which of the following is not a suitable model for the given data?

			Target
Acceleration to Horsepower Ratio	Horsepower	Car Length	Mileage
0.41	4000	12.5	7.4
0.9	3500	14	8.2
1.44	2300	15	11.6
2.05	2900	10.3	15
0.5	3600	11	18.8

- a. Logistic Regression**
- b. Linear Regression
- c. Support Vector Machines
- d. K Nearest Neighbors

194. Backpropagation can be used for Unsupervised Learning

- a. True
- b. False**

195. Which metric is foolproof while evaluating any classification task?

- a. Accuracy
- b. F1 Score**
- c. Mean Squared Error
- d. Mean Absolute Error

196. Which of the following is not an Optimizer Function?

- a. Adagrad
- b. RMSProp

- c. Categorical Cross Entropy
 - d. Adam
 - e. Gradient Descent
197. Probing the search space with the hope of finding other promising solutions that are yet to be found. This is also known as?
- a. Agent-Reward Tradeoff
 - b. Reinforcement Learning
 - c. Exploitation
 - d. Exploration**
198. Q-Learning is a model-free Reinforcement Learning Algorithm
- a. True**
 - b. False
199. What property should a feedback network have, to make it useful for storing information?
- a. accretive behaviour**
 - b. interpolative behaviour
 - c. both accretive and interpolative behaviour
 - d. none of the mentioned
200. Linear neurons can be useful for application such as interpolation, is it true?
- a. Yes**
 - b. No
201. What is the objective of a pattern storage task in a network?
- a. to store a given set of patterns
 - b. to recall a give set of patterns
 - c. both to store and recall**
 - d. none of the mentioned
202. What are the issues on which biological networks proves to be superior than AI networks?
- a. robustness & fault tolerance
 - b. flexibility
 - c. collective computation
 - d. all of the mentioned**

Prepared by:

Priyal Vaghela (CSE Dept., DEPSTAR)

Phenil Buch (CSE Dept., DEPSTAR)