1. **How do we perform Bayesian classification when some features are missing?**

2. (A) We assuming the missing values as the mean of all values.

3. (B) We ignore the missing features.

4. (C) We integrate the posteriors probabilities over the missing features.

5. (D) Drop the features completely.

6. **Answer:** *Option-C.*

2. Which of the following statement is False in the case of the KNN Algorithm?

(A) For a very large value of K, points from other classes may be included in the neighborhood.

(B) For the very small value of K, the algorithm is very sensitive to noise.

(C) KNN is used only for classification problem statements.

(D) KNN is a lazy learner.

Answer: *Option-C*

Explanation: We can use KNN for both regression and classification problem statements. In classification, we use the majority class based on the value of K while in regression we take an average of all points and then give the predictions.

3. Which of the following statement is TRUE?

(A) Outliers should be identified and removed always from a dataset.

(B) Outliers can never be present in the testing dataset.

(C) Outliers is a data point that is significantly close to other data points.

(D) The nature of our business problem determines how outliers are used.

Answer: *Option-D*

Explanation: The nature of a business problem often determines the use of outliers e.g, in case of problems where class imbalance condition exists like Credit Card Fraud detection, where the records for fraud class are very less with respect to no fraud class.

4. The following data is used to apply a linear regression algorithm with least squares regression line Y=a_1X. Then, the approximate value of a_1 is given by:(X-Independent variable, Y-Dependent variable)

(A) 27.876   (B) 32.650   (C) 40.541   (D) 28.956

X	1	20	30	40
Y	1	400	800	1300

Answer: *Option-D*

Explanation: Hint: Use the ordinary least square method.

5. The robotic arm will be able to paint every corner in the automotive parts while minimizing the quantity of paint wasted in the process. Which learning technique is used in this problem?

(A) Supervised Learning.

(B) Unsupervised Learning.

(C) Reinforcement Learning.

(D) Both (A) and (B).

Answer: *Option-C*

Explanation: Here robot is learning from the environment, by taking the rewards for positive actions and penalties for negative actions.

6. Which one of the following statements is TRUE for a Decision Tree?

(A) Decision tree is only suitable for the classification problem statement.

(B) In a decision tree, the entropy of a node decreases as we go down a decision tree.

(C) In a decision tree, entropy determines purity.

(D) Decision tree can only be used for only numeric valued and continuous attributes.

Answer: *Option-B*

Explanation: Entropy helps to determine the impurity of a node and as we go down the decision tree, entropy decreases.

7. How do you choose the right node while constructing a decision tree?

(A) An attribute having high entropy

(B) An attribute having high entropy and information gain

(C) An attribute having the lowest information gain.

(D) An attribute having the highest information gain.

Answer: *Option-D*

Explanation: We select first those attributes which are having maximum information gain.

8. What kind of distance metric(s) are suitable for categorical variables to find the closest neighbors?

(A) Euclidean distance.

(B) Manhattan distance.

(C) Minkowski distance.

(D) Hamming distance.

Answer: *Option-D*

Explanation: Hamming distance is a metric for comparing two binary data strings i.e, suitable for categorical variables.

9. In the Naive Bayes algorithm, suppose that prior for class w1 is greater than class w2, would the decision boundary shift towards the region R1(region for deciding w1) or towards region R2(region for deciding w2)?

(A) towards region R1.

(B) towards region R2.

(C) No shift in decision boundary.

(D) It depends on the exact value of priors.

Answer: *Option-B*

Explanation: Upon shifting the decision boundary towards region R2, we preserve the prior probabilities proportion since prior for w1 is greater than w2.

10. Which of the following statements is FALSE about Ridge and Lasso Regression?

(A) These are types of regularization methods to solve the overfitting problem.

(B) Lasso Regression is a type of regularization method.

(C) Ridge regression shrinks the coefficient to a lower value.

(D) Ridge regression lowers some coefficients to a zero value.

Answer: *Option-D*

Explanation: Ridge regression never drops any feature instead it shrinks the coefficients. However, Lasso regression drops some features by making the coefficient of that feature zero. Therefore, the latter one is used as a Feature Selection Technique.

11. Which of the following is FALSE about Correlation and Covariance?

(A) A zero correlation does not necessarily imply independence between variables.

(B) Correlation and covariance values are the same.

(C) The covariance and correlation are always the same sign.

(D) Correlation is the standardized version of Covariance.

Answer: *Option-B*

Explanation: Correlation is defined as covariance divided by standard deviations and therefore, is the standardized version of covariance.

12. In Regression modeling we develop a mathematical equation that describes how, (Predictor-Independent variable, Response-Dependent variable)

(A) one predictor and one or more response variables are related.

(B) several predictors and several response variables response are related.

(C) one response and one or more predictors are related.

(D) All of these are correct.

Answer: *Option-C*

Explanation: In the regression problem statement, we have several independent variables but only one dependent variable.

13. True or False: In a naive Bayes algorithm, when an attribute value in the testing record has no example in the training set, then the entire posterior probability will be zero.

(A) True   (B) False   (C) Can't determined   (D) None of these.

Answer: *Option-A*

Explanation: Since for a particular value in the attribute, the probability will be zero due to the absence of an example present in the training dataset. This usually leads to the problem of zero probability in the Naive Bayes algorithm. For further reference refer to the given article Link**.**

14. Which of the following is NOT True about Ensemble Techniques?

(A) Bagging decreases the variance of the classifier.

(B) Boosting helps to decrease the bias of the classifier.

(C) Bagging combines the predictions from different models and then finally gives the results.

(D) Bagging and Boosting are the only available ensemble techniques.

Answer: *Option-D*

Explanation: Apart from bagging and boosting there are other various types of ensemble techniques such as Stacking, Extra trees classifier, Voting classifier, etc.

15. **Which of the following statement is TRUE about the Bayes classifier?**

(A) Bayes classifier works on the Bayes theorem of probability.

(B) Bayes classifier is an unsupervised learning algorithm.

(C) Bayes classifier is also known as maximum apriori classifier.

(D) It assumes the independence between the independent variables or features.

Answer: *Option-A*

Explanation: Bayes classifier internally uses the concept of Bayes theorem for doing the predictions for unseen data points.

SECTION-2:

16. Which of the following SGD variants is based on both momentum and adaptive learning?

(A) RMSprop.

(B) Adagrad.

(C) Adam.

(D) Nesterov.

Answer: *Option-C*

Explanation: Adam, being a popular deep learning optimizer is based on both momentum and adaptive learning.

17. Which of the following activation function output is zero centered?

(A) Hyperbolic Tangent.

(B) Sigmoid.

(C) Softmax.

(D) Rectified Linear unit(ReLU).

Answer: *Option-A*

Explanation: Hyperbolic Tangent activation function gives output in the range [-1,1], which is symmetric about zero.

18. Which of the following is FALSE about Radial Basis Function Neural Network?

(A) It resembles Recurrent Neural Networks(RNNs) which have feedback loops.

(B) It uses radial basis function as activation function.

(C) While outputting, it considers the distance of a point with respect to the center.

(D) The output given by the Radial basis function is always an absolute value.

Answer: *Option-A*

Explanation: Radial basis functions do not resemble RNN but are used as an artificial neural network, which takes a distance of all the points from the center rather than the weighted sum.

19. In which of the following situations, you should NOT prefer Keras over TensorFlow?

(A) When you want to quickly build a prototype using neural networks.

(B) When you want to implement simple neural networks in your initial learning phase.

(C) When you are doing critical and intensive research in any field.

(D) When you want to create simple tutorials for your students and friends.

Answer: *Option-C*

Explanation: Keras is not preferred since it is built on the top of Tensorflow which provides both high-level and low-level APIs.

20. Which of the following is FALSE about Deep Learning and Machine Learning algorithms?

(A) Deep Learning algorithms work efficiently on a high amount of data.

(B) Feature Extraction needs to be done manually in both ML and DL algorithms.

(C) Deep Learning algorithms are best suited for unstructured data.

(D) Deep Learning algorithms require high computational power.

Answer: *Option-B*

Explanation: Usually, in deep learning algorithms, feature extraction happens automatically in hidden layers.

21. Which of the following is FALSE for neural networks?

(A) Artificial neurons are similar in operation to biological neurons.

(B) Training time for a neural network depends on network size.

(C) Neural networks can be simulated on conventional computers.

(D) The basic unit of neural networks are neurons.

Answer: *Option-A*

Explanation: Artificial neuron is not similar in working as compared to biological neuron since artificial neuron first takes a weighted sum of all inputs along with bias followed by applying an activation function to gives the final result whereas the working of biological neuron involves axon, synapses, etc.

22. Which of the following logic function cannot be implemented by a perceptron having 2 inputs?

(A) AND.　　　(B) OR.　　　(C) NOR.　　　(D) XOR.

Answer: *Option-D*

Explanation: Perceptron always gives a linear decision boundary, however for the Implementation of the XOR function we need a non-linear decision boundary.

23. Inappropriate selection of learning rate value in gradient descent gives rise to:

(A) Local Minima.

(B) Oscillations.

(C) Slow convergence.

(D) All of the above.

Answer: *Option-D*

Explanation: Learning rate decides how much fast or slow our optimizer is able to achieve the global minimum so, by choosing an inappropriate value of learning rate we may not reach global minimum instead stuck at a local minimum, oscillate around the minimum and because of which the convergence times increases.

SECTION-3:

24. What will be the output of the following code?

(A) 25   (B) 26   (C) 6   (D) None of these

Answer: *Option-B*

Explanation: where( ) function gives an array of indices where the value of the particular index is zero in n_array.

25. **What will be the output of the following code?**

(A) 0   (B) 1   (C) 2   (D) Code is not executable.

Answer: *Option-C*

Explanation: Cross-product of two vectors are not commutative.

26. **What will be the output of the following code?**

(A) 0   (B) 1   (C) 2   (D)Answer not fixed due to randomness.

Answer: *Option-C*

Explanation: random function returns samples from the "standard normal" distribution.

27. **What will be the output of the following code?**

(A) 2   (B) 3   (C) 4   (D) 5

Answer: *Option-D*

Explanation: argsort( ) function first sorts the array in ascending order and then gives the output as an index of those sorted array elements in the initial array.

28. What will be the output of the following code?

(A) 1   (B) 2   (C) 0   (D) 3

Answer: *Option-A*

Explanation: ndim function returns the dimension.

29. **What will be the output of the following code?**

(A) 2   (B) 4   (C) 6   (D) 8

Answer: *Option-B*

Explanation: arange( ) function gives a 1-d array with values from 0 to 5 and reshape function resize our array to 2-d. Accordingly, trace gives the sum of diagonal elements of the result matrix.

30. What will be the output of the following code?

(A) INT   (B) FLOAT   (C) STR   (D) BOOL.

Answer: *Option-B*