**Machine Learning**

1. Which of the following in sk-learn library is used for hyper parameter tuning?

Ans: (D) All of the above

2. In which of the below ensemble techniques trees are trained in parallel?

Ans: (A) Random forest

3. In machine learning, if in the below line of code:

sklearn.svm.SVC (C=1.0, kernel='rbf', degree=3)

we increasing the C hyper parameter, what will happen?

Ans: (A) The regularization will increase

4. Check the below line of code and answer the following questions:

sklearn.tree.DecisionTreeClassifier(\*criterion='gini',splitter='best',max\_depth=None, min\_samples\_split=2)

Which of the following is true regarding max\_depth hyper parameter?

Ans: (A) It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown.

5. Which of the following is true regarding Random Forests?

Ans: (A) It's an ensemble of weak learners.

(C) In case of classification problem, the prediction is made by taking mode of the class labels predicted by the component trees.

6. What can be the disadvantage if the learning rate is very high in gradient descent?

Ans: (A) Gradient Descent algorithm can diverge from the optimal solution.

(B) Gradient Descent algorithm can keep oscillating around the optimal solution and may not settle.

7. As the model complexity increases, what will happen?

Ans: (B) Bias will decrease, Variance increase

8. Suppose I have a linear regression model which is performing as follows:

Train accuracy=0.95 and Test accuracy=0.75

Which of the following is true regarding the model?

Ans: (B) Model is overfitting

9. Suppose we have a dataset which have two classes A and B. The percentage of class A is 40% and percentage of class B is 60%. Calculate the Gini index and entropy of the dataset.

Ans: Gini index = 1- (p(A)2 + p(B)2) = 1-((0.4)2+(0.6)2) = 0.48

Entropy = -(p(A) log2(p(A)) + p(B) log2(p(B))) = -(0.4\*log2(0.4) + 0.6\*log2(0.6))) = 0.97

10. What are the advantages of Random Forests over Decision Tree?

Ans: The main advantage of random forest over Decision Trees is that random forest alleviates the problem of over-fitting. The decision trees are highly likely to overfit the training data if we do not regularize them by restricting the max depth of the tree, max number of leaf nodes etc. Random forest is an ensemble technique in which each tree is trained on a bootstrapped sample which is taken from the training dataset and the prediction is made by taking average of the prediction made by each tree. So, random forests are very less likely to overfit and they generalize well on unseen data as compared to decision trees.

11. What is the need of scaling all numerical features in a dataset? Name any two techniques used for scaling.

Ans: Most of the times, a dataset will contain features highly varying in magnitudes, units and range. But since, most of the machine learning algorithms use Euclidean distance between two data points in their computations, this is a problem because, the results would vary greatly between different units, 5 km and 5000 m. The features with high magnitudes will weight a lot more in the distance calculations than features with low magnitudes. To avoid this effect, we use scaling which transforms all features to the same level of magnitudes. The two major scaling techniques used now a days are standardization and normalization.

12. Write down some advantages which scaling provides in optimization using gradient descent algorithm.

Ans: The advantage which scaling provides in optimization using gradient descent algorithm is that if the data is scaled, the gradient descent algorithm reached the optimal solution quickly as compared to the data which is not scaled.

13. In case of a highly imbalanced dataset for a classification problem, is accuracy a good metric to measure the performance of the model. If not, why?

Ans: Consider the case of a highly imbalanced dataset in which we have 90% data points belonging to label 1 and only 10% belonging to label 0. In that case even if we naively assign label 1 to all the data points we will get 90% accuracy because 90% data-points are of label 1 so, accuracy here will not be a good metric to evaluate the performance of a model. Even a very naive model like giving label 1 without even looking at any feature will give you 90% accuracy so, the accuracy will not be a good metric to use here.

14. What is “f-score" metric? Write its mathematical formula.

Ans: f-score metric is the metric which to measure the performance of a classification model. F-score (F1 score) is the harmonic mean of precision and recall:

15. What is the difference between fit(), transform() and fit\_transform()?

Ans: fit() : fit() is used for training. It calculates the parameters(weights) to be applied on different features in order to transform them as required by the model.

transform() : transform uses the parameters(weights) learned by fit() and applies it to respective features of dataset in order to change them as per requirement of the model.

fit\_transform() : fit\_transform() first fits() the data and then transforms the same dataset on which fit is performed. Basically, it does both the tasks on the same dataset.

fit() is generally used on training data in supervised learning where there is no need to change the training data, fit\_transform is used on training data in unsupervised learning where training data must also be modified and transform() is used only on test data because we need to change the test data as per the parameters(weights) learned by training.

**Statistics**

1. A die is thrown 1402 times. The frequencies for the outcomes 1, 2, 3, 4, 5 and 6 are given in the following table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome | 1 | 2 | 3 | 4 | 5 | 6 |
| Frequency | 400 | 300 | 157 | 180 | 175 | 190 |

Ans: B) 0.135

2. A telephone directory page has 400 telephone numbers. The frequency distribution of their unit place digit (for example, in the number 25827689, the unit place digit is 9 is given in table below:

First row refers to the digits

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Second row to their frequencies. 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 44 | 52 | 44 | 44 | 40 | 20 | 28 | 56 | 32 | 40 |

Ans: (D) 0.53

3. A tyre manufacturing company which keeps a record of the distance covered before a tyre needed to be replaced. The table below shows the results of 1100 cases.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Distance  (miles) | <4000 | 4000-9000 | 9001-14000 | >14000 |
| Frequency | 20 | 260 | 375 | 445 |

If we buy a new tyre of this company, what is the probability that the tyre will last more than 9000 miles?

Ans: (C) 0.745

4. Please refer to the case and table given in the question No. 3 and determine what is the probability that if we buy a new tyre then it will last in the interval [4000-14000] miles?

Ans: (B) 0.577

5. We have a box containing cards numbered from 0 to 9. We draw a card randomly from the box. If it is told to you that the card drawn is greater than 4 what is the probability that the card is odd?

Ans: (C) 0.6

6. We have a box containing cards numbered from 1 to 8. We draw a card randomly from the box. If it is told to you that the card drawn is less than 4 what is the probability that the card is even?

Ans: (A) 0.33

7. A die is thrown twice and the sum of the numbers appearing is observed to be 7. What is the conditional probability that the number 6 has appeared at least on one of the die?

Ans: (C) 0.33

8. Consider the experiment of tossing a coin. If the coin shows tail, toss it again but if it shows head, then throw a die. Find the conditional probability of the event that ‘the die shows a number greater than 4’ given that ‘there is at least one Head’.

Ans: (B) 0.22

9. There are three persons Evan, Ross and Michelle. These people lined up randomly for a picture. What is the probability of Ross being at one of the ends of the line?

Ans: (A) 0.66

10. Let us make an assumption that each born child is equally likely to be a boy or a girl. Now suppose, if a family has two children, what is the conditional probability that both are girls given that at least one of them is a girl?

Ans: (A) 0.33

11. Consider the same case as in the question no. 10. It is given that elder child is a boy. What is the conditional probability that both children are boys?

Ans: (C) 0.5

12. We toss a coin. If we get head, we toss a coin again and if we get tail we throw a die. What is the probability of getting a number greater than 4 on die?

Ans: (A) 0.166

13. We toss a coin. If we get head, we toss a coin again and if we get tail we throw a die. What is the probability of getting an odd number on die?

Ans: (D) 0.25

14. Suppose we throw two dice together. What is the conditional probability of getting sum of two numbers found on the two die after throwing is less than 4, provided that the two numbers found on the two die are different?

Ans: (D) 0.06

15. A box contains three coins: two regular coins and one fake two-headed coin, you pick a coin at random and toss it. What is the probability that it lands heads up?

Ans: (B) 2/3