FLIP ROBO

MACHINE LEARNING

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

Answer - D) All of the above

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

Answer - 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?

Answer - B) The regularization will decrease

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?

Answer - C) both A & B

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

Answer - A) It's an ensemble of weak learners.

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

Answer - C) Both of them

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

Answer - B) model is overfitting

Q9 to Q15 are subjective answer type questions, Answer them briefly.

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.

Answer -

$$Gini = 1 - \sum_j p_j^2$$

Gini = 1 - (0.40^2 + 0.60^2)
= 1 - (0.16+0.36)
= 1 – (0.52)
= 0.48

$$Entropy = -\sum_j p_j \log_2 p_j$$

Entropy = - [0.4 * log2(0.4) + 0.6 * log2(0.6)]
= - [0.4 * -1.32192809489 + 0.6 * -0.736965594166]
= 0.97

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

Answer - Random forests overcome several problems with decision trees, including:

Reduction in overfitting: by averaging several trees, there is a significantly lower risk of overfitting.

Less variance: By using multiple trees, you reduce the chance of stumbling across a classifier that doesn't perform well because of the relationship between the train and test data.
As a consequence, in almost all cases, random forests are more accurate than decision trees.

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

Answer - Since the features have different scales, there is a chance that higher weightage is given to features with higher magnitude. This will impact the performance of the machine learning algorithm and obviously, We do not want our algorithm to be biased towards one feature. We scale our data before employing a distance based algorithm so that all the features contribute equally to the result.
Normalization is a scaling technique in which values are shifted and re scaled so that they end up ranging between 0 and 1. It is also known as Min-Max scaling.

Standardization is another scaling technique where the values are centred around the mean with a unit standard deviation. This means that the mean of the attribute becomes zero and the resultant distribution has a unit standard deviation.

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

Answer - Machine learning algorithms that use gradient descent as an optimization technique, such as linear regression, logistic regression, neural networks, and so on, require data to be scaled. The presence of feature value X in the formula influences the gradient descent step size. The difference in feature ranges will result in different step sizes for each feature. We scale the data

before feeding it to the model to ensure that the gradient descent moves smoothly towards the minima and that the gradient descent steps are updated at the same rate for all features.

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?

Answer - Accuracy can still be a useful metric when the class distribution is slightly skewed. When there is a significant skew in the class distributions, accuracy can become an unreliable measure of model performance. The reason for this unreliability is centred on the average machine learning practitioner and classification accuracy intuitions. Typically, classification predictive modelling is applied to small datasets with equal or nearly equal class distributions. As a result, most practitioners develop the intuition that high accuracy scores (or low error rate scores) are good, and values above 90% are excellent. On an imbalanced classification problem, achieving 90 percent or even 99 percent classification accuracy may be trivial. This means that intuitions for classification accuracy based on balanced class distributions will be applied and will be incorrect, leading practitioners to believe that a model has good or even excellent performance when, in fact, it does not.

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

Answer - The F-score, also known as the F1-score, is a model's accuracy on a dataset. It is used to evaluate binary classification systems that categorise examples as 'Positive' or 'Negative'. The F-score is a method of combining the model's precision and recall; it is defined as the harmonic mean of the model's precision and recall. The F-score is widely used for assessing information retrieval systems such as search engines, as well as many types of machine learning models, particularly in natural language processing. It is possible to change the F-score to prioritise precision over recall, or vice versa. The F0.5- and F2-scores, as well as the standard F1-score, are common adjusted F-scores. The standard F1-score formula is the harmonic mean of precision and recall. The F-score of a perfect model is 1.

$$F_1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

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

Answer - Fit() - In the fit() method, we use the necessary formula and perform the calculation on the feature values of the input data before fitting this calculation to the transformer. To use the fit() method, we must use. fit() in front of the transformer object.

Transform() – We probably transform the data in the transform() method, where we apply the calculations that we calculated in fit() to each data point in feature F. Because we transform the fit calculations, we must use.transform() in front of a fit object. When we create an object of the fit method, we simply place it in front of the, as shown in the preceding section. The transform method uses those calculations to transform the scale of the data points, and the output is always in the form of a sparse matrix or array.

Fit_transform() - In essence, this fit transform() method combines the fit method and the transform method, making it equivalent to fit (). transform(). This approach converts the data

points and fits and transforms the input data all at once. When both are required, using fit and transform separately will reduce the model's effectiveness. Instead, use fit transform(), which will perform both tasks.