MACHINE LEARNING WORKSHEET – 7

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. 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?

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.

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.

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

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.

ANS. To calculate the Gini index and entropy of the dataset, we need to know the probabilities of each class. In this case, the probability of class A is 0.4 and the probability of class B is 0.6.

Gini Index:

The Gini index measures the impurity of a set of examples. It is defined as follows:

Gini Index = 1 - (p(A)^2 + p(B)^2)

where p(A) is the probability of class A and p(B) is the probability of class B

Substituting the values, we get:

Gini Index = 1- (0.4^2 + 0.6^2)

= 1 - (0.16 + 0.36)

= 1 - 0.52

= 0.48

Therefore, the Gini index of the dataset is 0.48

Entropy:

Entropy is another measure of impurity that is commonly used in decision tree algorithms it is defined as follows:

Entropy= - p(A)*log2(p(A)) - p(8)*log2(p(8))

Substituting the values, we get

Entropy = - 0.4 *log2(0.4) - 0.6 *log2(0.6)

= -0.4 * (-1.321)-0.6 * (-0.736)

= 0.5296

Therefore the entropy of the dataset is 0.5295

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

ANS. 1. Reduction of overfitting: Random Forests reduce overfitting by averaging the results of multiple trees and by selecting a random subset of features for each tree.

2. Better performance Random Forests tend to perform better than Decision Trees on large datasets with high dimensionality.

3. Outliers have less impact Outliers and noise in the data have less impact on Random Forests than on Decision Trees, since Random Forests are built from multiple trees that each only see a subset of the data.

4. Feature importance: Random Forests provide a measure of feature importance, which can be useful for feature selection and understanding the underlying data.

5. Robustness to missing values: Random Forests are robust to missing values, since they can use the available values for each feature to make a prediction, instead of discarding the entire sample

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

ANS. Scaling is required to bring all numerical features to a common scale, as features on different scales may have a different impact on the model's performance. This can result in features with larger scales having a higher impact on the model, even if they are not necessarily more important than features with smaller scales. Scaling the features ensures that each feature has an equal impact on the model's performance.

Two commonly used techniques for scaling are:

1. Standardization: In this technique, the features are transformed to have a mean of 0 and a standard deviation of 1. This technique works well when the data follows a normal distribution.

2. Min-max scaling: In this technique, the features are scaled to a fixed range, typically between 0 and 1. This technique works well when the data is not normally distributed or when the range of the features is known.

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

ANS. Scaling provides several advantages in optimization using gradient descent algorithm so,

1. It speeds up the convergence of the algorithm by allowing it to take larger steps in the direction of the gradient. This can be particularly important when dealing with high- dimensional data.

2. It helps to avoid oscillations and overshooting the optimal solution. When the data is not scaled, the algorithm may keep bouncing back and forth across the optimal solution, making it harder to converge.

3. Scaling can improve the numerical stability of the optimization algorithm. This is particularly important when the input data has different scales or ranges, which can cause numerical issues such as overflow or underflow.

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. In case of a highly imbalanced dataset for a classification problem, accuracy is not a good metric to measure the performance of the model. This is because in an imbalanced dataset, the majority class may dominate the prediction, leading to high accuracy despite poor performance on the minority class.

In such cases, other evaluation metrics such as precision, recall, F1 score or Area Under the Receiver Operating Characteristic curve (AUC-ROC) should be used to evaluate the performance of the model. These metrics take into account the true positives, true negatives, false positives and false negatives of each class and provide a more balanced view of the model's performance.

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

ANS. F-score (also known as F1-score) is a metric used to evaluate the performance of a classification model. It combines precision and recall into a single value.

The formula for F1-score is:

F1-score = 2* (precision recall) / (precision + recall)

ANS. where precision = true positives/ (true positives + false positives)

and recall true positives/ (true positives + false negatives)

In other words, the F1-score is the harmonic mean of precision and recall. It ranges from 0 to 1, with a higher value indicating better performance of the model.

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

ANS. fit() method is used to fit the model to the training data. It calculates the coefficients of the model that best fit the data.

transform() method is used to transform the data to a new form. It applies some transformations to the data using the coefficients calculated by the fit() method.

fit_transform() method is a combination of the above two methods. It first fits the model to the data and then applies the transformation to the data.