MACHINE LEARNING Assignment-7 Answers

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. 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.

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:** Gini index measures the impurity of the dataset. The formula for Gini index is:

Gini = 1 - Σ(Pi^2)

Gini(A) = 1 - (0.4)^2 - (0.6)^2 = 0.48

Gini(B) = 1 - (0)^2 - (1)^2 = 0

Gini(Dataset) = (0.4) * (0.48) + (0.6) * (0) = 0.192

Entropy: Entropy measures the impurity of the dataset. The formula for entropy is:

Entropy = -Σ(Pi*log2(Pi))

Entropy(A) = -0.4 * log2(0.4) - 0.6 * log2(0.6) = 0.971

Entropy(B) = -0 * log2(0) - 1 * log2(1) = 0

Entropy(Dataset) = (0.4) * (0.971) + (0.6) * (0) = 0.388

The Gini index of the dataset is **0.192** and the entropy of the dataset is **0.388**.

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

Ans. Random Forests have several advantages over Decision Trees:

- Reduced Overfitting: Random Forests reduce overfitting compared to decision trees by combining multiple decision trees and averaging their predictions.
- Better accuracy: Random Forests generally have better accuracy compared to Decision Trees due to the ensemble effect of combining multiple decision trees.
- Robust to Outliers: Random Forests are more robust to outliers than Decision Trees as the outliers tend to be averaged out by the combined predictions of multiple trees.
- No Feature Scaling Required: Random Forests do not require feature scaling as decision trees work on each feature separately.
- Handles Missing Data: Random Forests can handle missing data as they use only a subset of features at each node split.

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

Ans. Scaling is important to ensure that the features are on a similar scale and to prevent features with larger numerical ranges from dominating over features with smaller numerical ranges during model training. Two techniques used for scaling are:

Min-max scaling or normalization: It scales the data to a fixed range of [0,1]. This can be done using the formula: x_scaled = (x - x_min) / (x_max - x_min), where x_min and x_max are the minimum and maximum values of the feature x.

Standardization: It scales the data to have a mean of 0 and standard deviation of 1. This can be done using the formula: x_scaled = (x - mean) / standard deviation, where mean and standard deviation are the mean and standard deviation of the feature x.

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

Ans. Scaling provides the following advantages in optimization using gradient descent algorithm:

- Helps in faster convergence: Scaling can help in faster convergence of gradient descent algorithm as it ensures that the updates made to the parameters are of similar scales.
- Helps in avoiding oscillations: Scaling can help in avoiding oscillations around the minimum as it ensures that the gradient descent algorithm moves smoothly towards the minimum by preventing it from bouncing back and forth between steep and shallow regions.

- Helps in finding the global minimum: Scaling can help in finding the global minimum of the cost function as it ensures that the gradient descent algorithm is not stuck in local minima or saddle points due to differences in scales of the features.
- Helps in regularization: Scaling can help in regularization by preventing some features from dominating others due to differences in their scales.

Techniques used for scaling include Min-Max Scaling, Standard Scaling, and Robust Scaling.

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, accuracy is not a good metric to measure the performance of the model. This is because accuracy is based on the number of correct predictions divided by the total number of predictions, and in the case of an imbalanced dataset, the number of correct predictions for the minority class may be very low. Therefore, a model can have high accuracy even if it fails to detect the minority class properly. In such cases, metrics such as precision, recall, F1 score, and AUC-ROC are better suited to measure the performance of the model. These metrics take into account both true positives and false positives and provide a better evaluation of the model's performance on imbalanced datasets.

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

Ans. F-score, also known as F1-score, is a metric commonly used in binary classification tasks to evaluate the model's precision and recall at the same time. It combines precision and recall into a single score that reflects both aspects of the model's performance.

The F-score is calculated as the harmonic mean of precision and recall:

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

where precision is the number of true positives divided by the number of true positives plus false positives, and recall is the number of true positives divided by the number of true positives plus false negatives.

The F1-score ranges from 0 to 1, where 1 is the best possible score, indicating perfect precision and recall.

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

Ans. In machine learning, fit(), transform(), and fit_transform() are methods used to preprocess data before feeding it to a machine learning model.

- fit() is used to calculate and store the parameters (e.g., mean, standard deviation) of the transformation applied to the data, based on the training data set.
- transform() is used to apply the transformation on a given data set, based on the parameters calculated by the fit() method.
- fit_transform() combines the above two methods and first fits the transformation parameters based on the training data set and then applies the transformation on the same data set.

In summary, fit() is used to learn the transformation parameters from the training data, transform() is used to apply the learned transformation to the data, and fit_transform() is used to do both of these steps at once.

It's important to note that not all preprocessing techniques require fit() method. For example, min-max scaling does not require any parameter calculation, and thus only transform() is used to apply the transformation.