WORKSHEET- 7 MACHINE LEARNING

In Q1 to Q8, only one option is correct, choose the correct option:

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

a) GridSearchCV()

b) RandomizedCV()

c) K-fold Cross Validation

d) All of the above

Answer : d) All of the above

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

a) Random forest

b) Adaboost

c) Gradient Boosting

d) All of the above

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?

a) The regularization will increase

b) The regularization will decrease

c) No effect on regularization

d) kernel will be changed to linear

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?

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

b) It denotes the number of children a node can have.

c) both A & B

d) None of the above

Answer : c) both A & B

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

a) It's an ensemble of weak learners.

b) The component trees are trained in series

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

d) None of the above

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?

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.

c) Both of them

d) None of them

Answer : c) Both of them

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

a) Bias will increase, Variance decrease

b) Bias will decrease, Variance increase

c) Both bias and variance increase

d) Both bias and variance decrease.

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?

a) model is underfitting

b) model is overfitting

c) model is performing good

d) None of the above

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 : Scaling means nothing but transforming your data so that it fits within a specific scale or range, like 0-1. scaling of the data makes it easy for a model to learn and understand the problem. It is important to scale the features so that they are all on the same scale, as this helps the model to assign equal importance to all features and make predictions without bias.

Scaling technique:

WORKSHEET-7

1.standard scaling (or z-scaling or Standardization):

It is calculated by subtracting the mean and dividing by the standard deviation to shift the distribution to have a mean of zero and a standard deviation of one.

Z-score is a variation of scaling that represents the number of standard deviations away from the mean. You would use z-score to ensure your feature distributions have mean = 0 and std = 1. It's useful when there are a few outliers

2. min-max scaling :

Min-Max scaling is a normalization technique that enables us to scale data in a dataset to a specific range using each feature's minimum and maximum value.

Unlike standard scaling, where data are scaled based on the standard normal distribution(with *mean = 0* and *standard deviation = 1*), the min-max scaler uses each column's minimum and maximum value to scale the data series.

It is calculated by subtracting by the minimum value and dividing by the difference between the maximum and minimum values

WORKSHEET-7

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

Answer : Machine learning algorithms like linear regression, logistic regression, neural network, etc. that use gradient descent as an optimization technique require data to be scaled. The presence of feature value X in the formula will affect the step size of the gradient descent. The difference in ranges of features will cause different step sizes for each feature. To ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features, we scale the data before feeding it to the model.

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 : When the class distribution is slightly skewed, accuracy can still be a useful metric. When the skew in the class distributions are severe, accuracy can become an unreliable measure of model performance. The reason for this unreliability is centered around the average machine learning practitioner and the intuitions for classification accuracy. Typically, classification predictive modeling is practiced with small datasets where the class distribution is equal or very close to equal. Therefore, most practitioners develop an intuition that large accuracy score (or conversely small error rate scores) are good, and values above 90 percent are great. Achieving 90 percent classification accuracy, or even 99 percent classification accuracy, may be trivial on an imbalanced classification problem.This means that intuitions for classification accuracy developed on balanced class distributions will be applied and will be wrong, misleading the

practitioner into thinking that a model has good or even excellent performance when it, in fact, does not.

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

Answer : The F-score, also called the F1-score, is a measure of a model’s accuracy on a dataset. It is used to evaluate binary classification systems, which classify examples into ‘Positive’ or ‘Negative’. The F-score is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model’s precision and recall. The F-score is commonly used for evaluating information retrieval systems such as search engines, and also for many kinds of machine learning models, in particular in natural language processing. It is possible to adjust the F-score to give more importance to precision over recall, or vice-versa. Common adjusted F-scores are the F0.5-score and the F2-score, as well as the standard F1-score.The formula for the standard F1-score is the harmonic mean of the precision and recall. A perfect model has an F-score of 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, where we use the required formula and perform the calculation on the feature values of input data and fit this calculation to the transformer. For applying the fit() method we have to use .fit() in front of the transformer object.

Transform() – For changing the data we probably do transform, in the transform() method, where we apply the calculations that we have calculated in fit() to every data point in feature F. We have to use .transform() in front of a fit object because we transform the fit calculations.

We use the example that is used above section when we create an object of the fit method then we just put it in front of the .transform and transform method uses those calculations to transform the scale of the data points, and the output will we get is always in the form of sparse matrix or array.

Fit_transform() - This fit_transform() method is basically the combination of fit method and transform method, it is equivalent to fit().transform(). This method performs fit and transform on the input data at a single time and converts the data points. If we use fit and transform separate when we need both then it will decrease the efficiency of the model so we use fit_transform() which will do both the work.