**MACHINE LEARNING – WORKSHEET 4**

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

1. Which of the following in sklearn library is used for hyper parameter tuning?

A) GridSearchCV() B) RandomizedCV()

C) K-fold Cross Validation D) None of the above

**Answers:--** C) K-fold Cross Validation

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

**Answers:--** 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

**Answers:--** 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

**Answers:--** 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

**Answers:--**

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.

**Answers:--** 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.

**Answers:--** Bias will decrease, Variance increase

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

Train accuracy=0.95

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

**Answers:--** 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.

**Answers:--Gini Index=A/(A+B)**

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

**Answers:--**

1) Easy to interpret. Straightforward graphical visualization is intuitive makes it easy for anyone to understand

2) Useful to identify variables most helpful in prediction. It splits based on the attribute significance.

3) Handles both continuous and categorical targets attributes.

4) Performs well on large datasets

5) Not influenced by outliers and missing values to an extent.

6) Non parametric method- does make an assumption about the form of the mapping function. As they do not make any assumption about mapping function they are free to learn any functional form present in the data

7) Trees can handle qualitative input variables without the need to create dummy variables

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

**Answers:--** Machine learning algorithm just sees number — if there is a vast difference in the range say few ranging in thousands and few ranging in the tens, and it makes the underlying assumption that higher ranging numbers have superiority of some sort. So these more significant number starts playing a more decisive role while training the model.

The machine learning algorithm works on numbers and does not know what that number represents. A weight of 10 grams and a price of 10 dollars represents completely two different things — which is a no brainer for humans, but for a model as a feature, it treats both as same.

**The most common techniques of feature scaling are Normalization and Standardization.**

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

**Answers:--** We can speed up**gradient descent** by scaling because θ descends quickly on small ranges and slowly on large ranges, and oscillates inefficiently down to the optimum when the variables are very uneven.

We can use fixed learning rate during training without worrying about learning rate decay.

It has straight trajectory towards the minimum and it is guaranteed to converge in theory to the global minimum if the loss function is convex and to a local minimum if the loss function is not convex.

It has unbiased estimate of gradients. The more the examples, the lower the standard error.

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?

**Answers:--** In case of a highly imbalanced dataset for a classification problem , accuracy is not good metric to measure the performance of the model.

Classification accuracy is a metric that summarizes the performance of a classification model as the number of correct predictions divided by the total number of predictions.

It is easy to calculate and intuitive to understand, making it the most common metric used for evaluating classifier models. This intuition breaks down when the distribution of examples to classes is severely skewed.

Intuitions developed by practitioners on balanced datasets, such as 99 percent representing a skillful model, can be incorrect and dangerously misleading on imbalanced classification predictive modeling problems.

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

**Answers:--** The F measure (F1 score or F score) is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

**F-Measure = (2 \* Precision \* Recall) / (Precision + Recall)**

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

**Answers:--**The **fit** method, when applied to the training dataset,learns the model parameters (for example, mean and standard deviation). We then need to apply the **transform** method on the training dataset to get the transformed (scaled) training dataset. We could also perform both of this steps in one step by applying **fit\_transform** on the training dataset.

In summary, **fit** performs the training, **transform** changes the data in the pipeline in order to pass it on to the next stage in the pipeline, and **fit\_transform** does both the fitting and the transforming in one possibly optimized step.