

# MACHINE LEARNING

## ASSIGNMENT - 7

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

**ANS: D**

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

**ANS: D**

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

**ANS: B**

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

**ANS: A**

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

**ANS: A**

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

**ANS: C**

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.

**ANS: B**

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  
**ANS: A**

**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:**

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

**ANS:**

Decision Tree	Random Forest
A decision tree is a tree-like model of decisions along with possible outcomes in a diagram.	A classification algorithm consisting of many decision trees combined to get a more accurate result as compared to a single tree.
There is always a scope for overfitting, caused due to the presence of variance.	Random forest algorithm avoids and prevents overfitting by using multiple trees.
The results are not accurate.	This gives accurate and precise results.
Decision trees require low computation, thus reducing time to implement and carrying low accuracy.	This consumes more computation. The process of generation and analyzing is time-consuming.
It is easy to visualize. The only task is to fit the decision tree model.	This has complex visualization as it determines the pattern behind the data.

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 rescale the data and it’s used when we want features to be compared on the same scale for our algorithm. And, when all features are in the same scale, it also helps algorithms to understand the relative relationship better.  
Scaling Techniques

Types

Primary Scaling Techniques

Nominal Scale

Ordinal Scale

Interval Scale

Ratio Scale

Other Scaling Techniques

Comparative Scales

Non-Comparative Scales

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

**ANS:** Optimization refers to the task of minimizing/maximizing an objective function. Gradient Descent is the most common optimization algorithm in machine learning and deep learning. It is a first-order optimization algorithm. This means it only takes into account the first derivative when performing the updates on the parameters. On each iteration, we update the parameters in the opposite direction of the gradient of the objective function  $J(w)$  w.r.t the parameters where the gradient gives the direction of the steepest ascent. The size of the step we take on each iteration to reach the local minimum is determined by the learning rate  $\alpha$ . Therefore, we follow the direction of the slope downhill until we reach a local minimum.

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:** 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.

**ANS:** 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. 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.

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

**ANS:** The `fit(data)` method is used to compute the mean and std dev for a given feature to be used further for scaling. The `transform(data)` method is used to perform scaling using mean and std dev calculated using the `.fit()` method. The `fit_transform()` method does both fits and transform.