Machine learning worksheet7.

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

2. In which of the below ensemble techniques trees are trained in parallel?
Ans : All of the above

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 :

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 : both A & B

5. Which of the following is true regarding Random Forests?
Ans : 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 :Gradient Descent algorithm can diverge from the optimal solution.

7. As the model complexity increases, what will happen?
Ans :Both bias and variance decrease.

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 :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 :

10. What are the advantages of Random Forests over Decision Tree?
Ans : Decision Tree
A decision tree is a tree-like model of decisions along with possible outcomes in a diagram.
There is always a scope for overfitting, caused due to the presence of variance.
The results are not accurate.
Decision trees require low computation, thus reducing time to implement and carrying low accuracy.
It is easy to visualize. The only task is to fit the decision tree model.

Random Forest
A classification algorithm consisting of many decision trees combined to get a more accurate result as compared to a single tree.
Random forest algorithm avoids and prevents overfitting by using multiple trees.
This gives accurate and precise results.
This consumes more computation. The process of generation and analyzing is time-consuming.
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 :

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 f(x) parameterized by x. In machine/deep learning terminology, it's the task of minimizing the cost/loss function J(w) parameterized by the model's parameters w ∈ R^d. Optimization algorithms (in case of minimization) have one of the following goals:

Find the global minimum of the objective function. This is feasible if the objective function is convex, i.e. any local minimum is a global minimum.
Find the lowest possible value of the objective function within its neighborhood. That's usually the case if the objective function is not convex as the case in most deep learning problems.
There are three kinds of optimization algorithms:

Optimization algorithm that is not iterative and simply solves for one point.
Optimization algorithm that is iterative in nature and converges to acceptable solution regardless of the parameters initialization such as gradient descent applied to logistic regression.
Optimization algorithm that is iterative in nature and applied to a set of problems that have non-convex cost functions such as neural networks. Therefore, parameters' initialization plays a critical role in speeding up convergence and achieving lower error rates.

Advantages :
1)We can use fixed learning rate during training without worrying about learning rate decay.
2)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.
3)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?

Ans :

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’.
An F-score is the harmonic mean of a system’s precision and recall values. It can be calculated by the following formula: 2 x [(Precision x Recall) / (Precision + Recall)].

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.

The fit() Method
The fit function computes the formulation to transform the column based on Standard scaling but doesn’t apply the actual transformation. The computation is stored as a fit object. The fit method doesn’t return anything.

The transform() Method
The transform method takes advantage of the fit object in the fit() method and applies the actual transformation onto the column. So, fit() and transform() is a two-step process that completes the transformation in the second step. Here, Unlike the fit() method the transform method returns the actually transformed array.

The fit_transform() Method
As we discussed in the above section, fit() and transform() is a two-step process, which can be brought down to a one-shot process using the fit_transform method. When the fit_transform method is used, we can compute and apply the transformation in a single step.