# Machine Learning Assignment – 7

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

**A) GridSearchCV()**

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

**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**

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**

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

**B) The component trees are trained in series**

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.**

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

**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?

**B) model is overfitting**

10. What are the advantages of Random Forests over Decision Tree?
Builds a robust model.

- Does not suffer from overfitting problem.
- Can use for both classification and regression problems.
- Gives highly accurate predictions.
- Powerful than other non-linear models.

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

Feature Scaling is a method to transform the numeric features in a dataset to a standard range so that the performance of the machine learning algorithm improves. It can be achieved by normalizing or standardizing the data values. This scaling is generally preformed in the data pre-processing step when working with machine learning algorithm.

Example, if we have weight of a person in a dataset with values in the range 15kg to 100kg, then feature scaling transforms all the values to the range 0 to 1 where 0 represents lowest weight and 1 represents highest weight instead of representing the weights in kgs.

Standardization and Normalization are the 2 techniques we can use for scaling.

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

- It makes the training faster
- It prevents the optimization from getting stuck in local optima
- It gives a better error surface shape
- Weight decay and Bayes optimization can be done more conveniently

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?

Imbalanced data can cause issues in understanding the performance of a model. When evaluating performance on imbalanced data, models that only predict well for the majority class will seem to be highly performant when looking at simple metrics such as accuracy, whilst in actuality the model is performing poorly.

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

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.

F-Score = (2 * Precision * Recall) / (Precision + Recall)

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

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 (fit transform in python) 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.

fit_transform() or fit transform sklearn: 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.