# 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?
   1. GridSearchCV() B) RandomizedCV()

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

Answer: (A)

1. In which of the below ensemble techniques trees are trained in parallel?
   1. Random forest B) Adaboost

C) Gradient Boosting D) All of the above

Answer: ( C)

1. 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?

* 1. The regularization will increase B) The regularization will decrease

C) No effect on regularization D) kernel will be changed to linear

Answer: ( C)

1. 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?

* 1. It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown.
  2. It denotes the number of children a node can have.
  3. both A & B
  4. None of the above

Answer: ( C)

1. Which of the following is true regarding Random Forests?
   1. It's an ensemble of weak learners.
   2. The component trees are trained in series
   3. In case of classification problem, the prediction is made by taking mode of the class labels predicted by the component trees.
   4. None of the above

Answer: (C )

1. What can be the disadvantage if the learning rate is very high in gradient descent?
   1. Gradient Descent algorithm can diverge from the optimal solution.
   2. Gradient Descent algorithm can keep oscillating around the optimal solution and may not settle.
   3. Both of them
   4. None of them.

Answer: (D )

1. As the model complexity increases, what will happen?
   1. Bias will increase, Variance decrease B) Bias will decrease, Variance increase C)both bias and variance increase D) Both bias and variance decrease.

Answer: (A)

1. 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?

* 1. model is underfitting B) model is overfitting

C) model is performing good D) None of the above

Answer: (C )

## Q9 to Q15 are subjective answer type questions, Answer them briefly.

1. 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: The overall Gini index before splitting is: Gorig = 1 − 0.4 2 − 0.6 2 = 0.48 The gain in the Gini index after splitting on A is: GA=T=1 – (4/7)2 – (3/7)2 = 0.4898 GA=F = 1 – (3/3)2 – (0/3) 2 = 0 Hence the corresponding gain is equal to Gorig – (7/10)GA=T –(3/10)GA=F = 0.1371. Similarly, we can compute the gain after splitting on B, which will be given by: Gorig –(4/10)GB=T -(6/10) GB=F = 0.1633. Therefore, attribute B will be chosen to split the node.

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

Answer: With that said, **random forests** are a strong modeling technique and much more robust than a single **decision tree**. They aggregate many **decision trees** to limit overfitting as well as error due to bias and therefore yield useful results.

The advantages of random forest are: It is one of the most accurate **learning** algorithms available. For many data sets, it produces a highly accurate classifier. It runs efficiently on large databases.

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

Answer: **Feature Scaling** or Standardization: It **is** a step of Data Pre Processing which **is** applied to independent variables or **features** of data. It basically helps to normalise the data within a particular range. Sometimes, it also helps in speeding up the calculations in an algorithm.

The most **common techniques** of **feature scaling** are **Normalization** and Standardization. **Normalization** is used when we want to bound our values between **two** numbers, typically, between [0,1] or [-1,1]. While Standardization transforms the data to have zero mean and a variance of 1, they make our data unitless.

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

Answer: **Gradient descent** is a first-order iterative optimization **algorithm** for finding a local minimum of a differentiable function. To find a local minimum of a function using **gradient descent**, we take steps proportional to the negative of the **gradient** (or approximate **gradient**) of the function at the current point.

**Gradient Descent** is an iterative **optimiZation** algorithm, used to find the minimum value for a function. The general idea is to initialize the parameters to random values, and then take small steps in the direction of the “slope” at each iteration.

Deep learning neural networks are trained using the stochastic **gradient descent algorithm**. ... A **learning rate** that is too large can cause the model to converge too quickly to a suboptimal solution, whereas a **learning rate** that is too small can cause the process to get stuck

Common **examples** of **algorithms** with coefficients that can be optimized using **gradient descent** are Linear Regression and Logistic Regression. ... Batch **gradient descent** is the most common form of **gradient descent** described in machine learning

1. 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: A pair of evaluations **metrics that are commonly used** when **there is a class imbalance** are precision and recall. Precision is defined as the number of true positives divided by the sum of true positives and false positives.

We can use classification **performance** metrics such as Log-Loss, **Accuracy**, AUC(Area under Curve) etc. Another example of metric for evaluation of machine learning algorithms is precision, recall, which can be used for sorting algorithms primarily used by search engines.

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

Answer: 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.

1. **F**-**Measure** = (2 \* Precision \* Recall) / (Precision + Recall)
2. **F**-**Measure** = (2 \* 0.633 \* 0.95) / (0.633 + 0.95)
3. **F**-**Measure** = (2 \* 0.601) / 1.583.
4. **F**-**Measure** = 1.202 / 1.583.
5. **F**-**Measure** = 0.759.
6. What is the difference between fit(), transform() and fit\_transform()?

Answer: "**transform**" uses a previously computed mean and std to autoscale the data (subtract mean from all values and then divide it by std). "**fit\_transform**" does both at the same time. So you can do it with 1 line of code instead of 2.