MACHINE LEARNING WORKSHEET 4

1. A) GridSearchCV()
2. A) Random forest
3. B) The regularization will decrease.
4. A) It regularizes the tree by limiting the maximum depth upto which a tree can be grown.
5. A) Its an ensemble of weak learners.
6. C) Both of them.
7. B) Bias will decrease, variance increase.
8. B) Model is overfitting.
9. 1) Gini Index

Gini Index = [ 1 - sum(Pi)^2]\*100

Where i = 1 to n(n is number of classes), Pi = probability of the ith class

Gini Index = [1 – [ (Pa)^2 + (Pb)^2]] \* 100

Where Pa = probability of class A,

Pb = probability of class B

Gini Index = [ 1 – [ (0.4)^2 + (0.6)^2] ] \*100

= [ 1 – [ 0.16 + 0.36] ] \*100

= [ 1 – [0.52] ]\* 100

= 0.48 \* 100

= 48

2) Entropy

Entropy = - [ sum [(Pi) \* log2 (Pi ) ] ]

Where Pi = probability of ith class , log2 (Pi) = log of probability of ith class with

Base 2

Entropy = -[ [Pa \* log2 (Pa)] + [Pb \* log2(Pb) ] ]

= - [ [(0.4) \* log2 (0.4) ] + [ (0.6) \* log2(0.6) ] ]

= - [ [ 0.4 \* -(1.3) ] + [ 0.6 \* -(0.07) ] ]

= - [ (- 0.52) + (-0.04) ]

= 0.52 + 0.04

= 0.56

* Random forests are much more robust than a single decision tree.
* Random forests aggregate many decision trees to limit overfitting and error due to bias.
* Random forests give more useful results than decision trees.

1. Feature Scaling helps to normalise the data within a particular range. When all the features are in a same range it helps algorithm to understand the relative relationship better.

Different techniques used for scaling :

* Min-Max normalization
* Standardization

1. The feature scaling brings every feature in same range without upfront importance. Feature scaling helps Gradient Descent to converge much faster.
2. No, Accuracy can’t be a good metric to measure the performance of a model trained on a highly imbalanced dataset.
3. The F-Score is a measure of testset accuracy and it is defined as the weighted harmonic mean of the precision and recall of the test.

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

1. Fit() computes the mean and std to be used for later scaling. (just a computation), nothing is given to you.

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.