# MACHINE LEARNING 5

3).Regularization:

When we use machine learning for any dataset so at that time our model can easily be overfitted and underfitted .for avoiding such problems we use regularization techniques in machine learning so our model can be fit properly this technique reduces the chance of overfitting  and underfitting.

4).GINI impurity index measures the diversity in the dataset.

6).Ensemble Technique:

When our dataset is too large there is large number of columns and rows so our model not able to learn such a huge data .so we do the subsets of that particular data set .means N number of subsets and those subsets goes for individual learning to the individual model and then they will combine the means and gives mean accuracy as a final result .ensemble method is divided into tow categories named by bagging and boosting .bagging is used for parallel dataset and boosting is used for sequential dataset.

7).Bagging:

In bagging technique we create parallel dataset .and dataset is divided in to N number of subsets as parallel and after that this subsets are learned by the model DTC. it is used for homogeneous model and gives mean accuracy and for very large dataset there is extra trees.

Boosting:

In boosting we have large dataset. It is learned by model and that model will boost out and amplify the learning and transfer its out put to the second model and final mean accuracy comes .there is adaboost and gradient boosting.

8).Out of bag Error:

When we perform bagging or random forest at that time we create subsets and feed the model with different different subsets and rows of subsets goes to the model many of rows can goes to model more than one time and on another hand never goes to any of the model those rows will be hidden from our models .this is called out of bag error we can use those rows after for testing the model.

9).K-Fold:

This is actual cross validation called k-fold where k is stands for cv means number of cross validation .original dataset is equally partitioned into subparts for each iteration one group is selected as validation of data and remaining groups are selected for training data and this process is repeated for k times until each group is iterated as validation and remaining as training data .

11).The lowest point could be overshooted.

14).Some times model starts learning data in different way it learns on its own Way with self assumptions .some time it learns all the points in data and it underfited or overfitted the model it happens due to variance and bias inside model and to over come with this problems we use regularization and cross validation techniques.