ASSIGNMENT – 5 MACHINE LEARING

Q1. Ans : R-Squared and RSS both have their advantages and disadvantages .

R-Squared is easy to compare across different models but it can be misleading for small datasets or models with many variables.

On the other hand , RSS is not directly comparable across different models but can be used to compare models with different scales for the different variables . Also it is more sensitive to outliers that R-Squared .

So , I don't think that either of these two is a better measure of goodness of fit model in regression .

Q2. Ans :

TSS : TSS is a descriptive statistics that shows how far datapoints are from the centre .

It is the sum of the squared differences between the observed variables and the overall mean .

ESS : ESS is a statistical tool that is used to show how well a model explains the observed data .

It is the sum of the squared differences between the predicted values and the mean value of the variables .

RSS : RSS is a statistical method used to check amount of variance in a dataset that is not explained by the regression model .

The equation relating these three metrics with each other TSS = RSS + ESS

Q3. Ans :

Regularization plays an important rule in Machine Learning .

Sometimes Machine learning model makes their own assumptions while working on a dataset . Some times it tries to learn from the details along with the noise of the data and tries to fit each datapoint on the curve called **Overfitting .** Sometimes it can neither learn the relationship between variables in the testing data nor predict or classify a new data point , it is called **Underfitting .**

To get ride of this problem , we use regularization in Machine Learning .

Two types of Regularization techniques are there ---------

1) Lasso Regression

2) Ridge Regression

Lasso Regression modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients .

Ridge Regression modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.

Q4. Ans :

Gini Impurity : Gini Impurity is the measure of impurity of a dataset . It is calculated by subtracting the sum of the squared probabilities of each class from 1 .

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

Fig (a) : Gini Impurity

Q5. Ans : Unregularized Decision trees make very few assumptions about the training data . If left unconstrained , the tree structure will adapt itself to thee training data , fitting it very closely , indeed most likely overfitting it . Such a model is often called as nonparametric model because the number of parameters is not determined prior to training , so the model structure is free to stick closely to the data . Hence it is prone to overfitting .

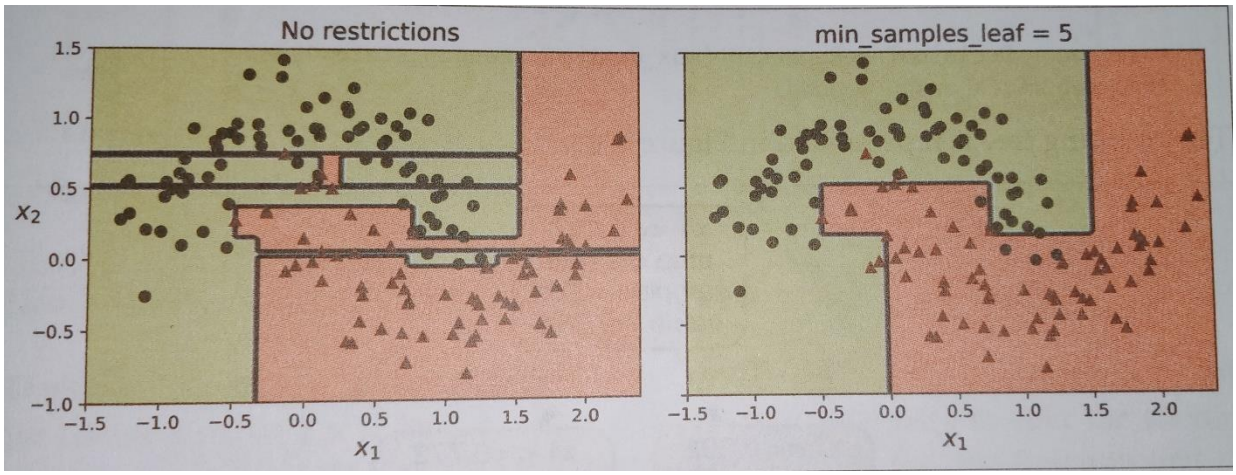


Fig (b) : Decision boundaries of an unregularized tree(left) and regularized tree(right)

ASSIGNMENT – 5 MACHINE LEARING

The unregularized tree on the left is clearly overfitting here.

Q6. Ans :

Ensemble Technique : Ensemble technique in machine learning is a very useful technique . It is a technique through which we can get a better result from the data .

Here , a combination of models performs the machine learning . training data are given to a combination of models . This group will often perform better than the best individual model .

Two important ensemble techniques in machine learning are ----------

1) Bagging
2) Boosting

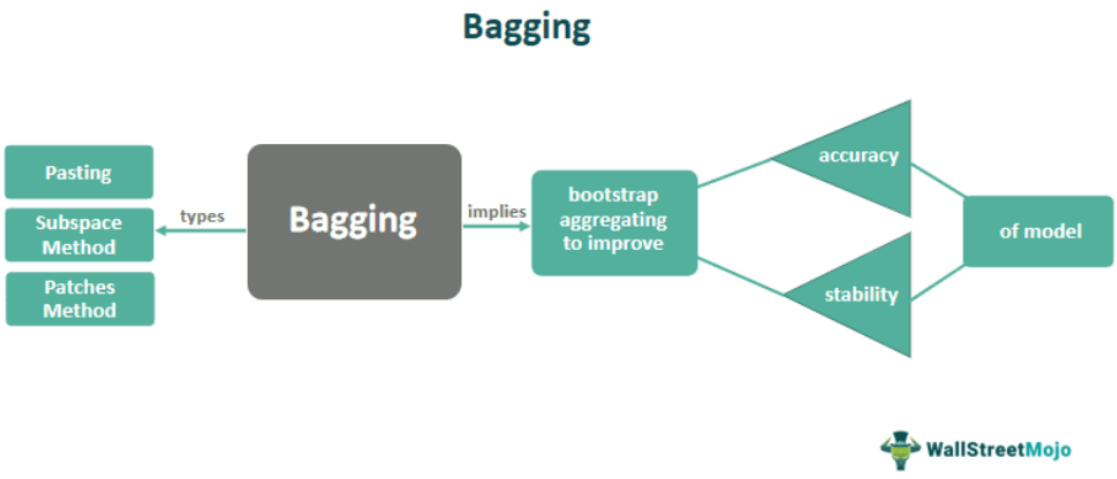
Q7. Ans :

Bagging , also known as bootstrap Aggregating is an ensemble technique where multiple models are combined to give more accurate result .

In Bagging, large datasets are divided into small subsets an each subsets are given to different models .

Here multiple base models are trained independently on the subsets and combine the results to create the final model .

One example of bagging is random forests



Fig(c) : Bagging technique

On the other hand , Boosting is an ensemble method that can combine several weak learners into a strong learner .

The general idea of most boosting methods is to train predictors sequentially , each trying to correct its predecessor .

One example of boosting method is **AdaBoost .**

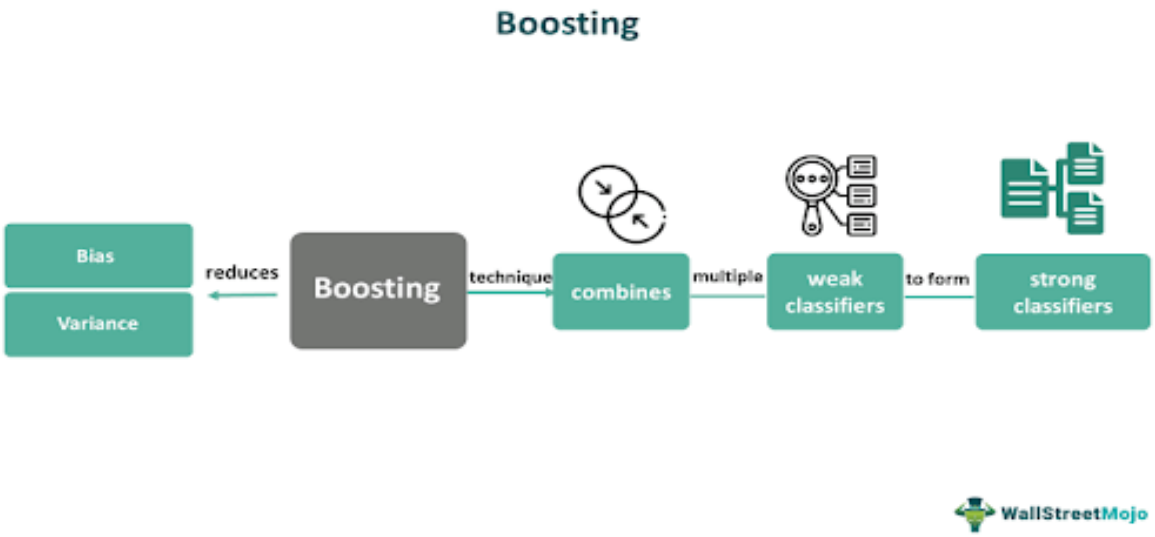


Fig (d) : Boosting

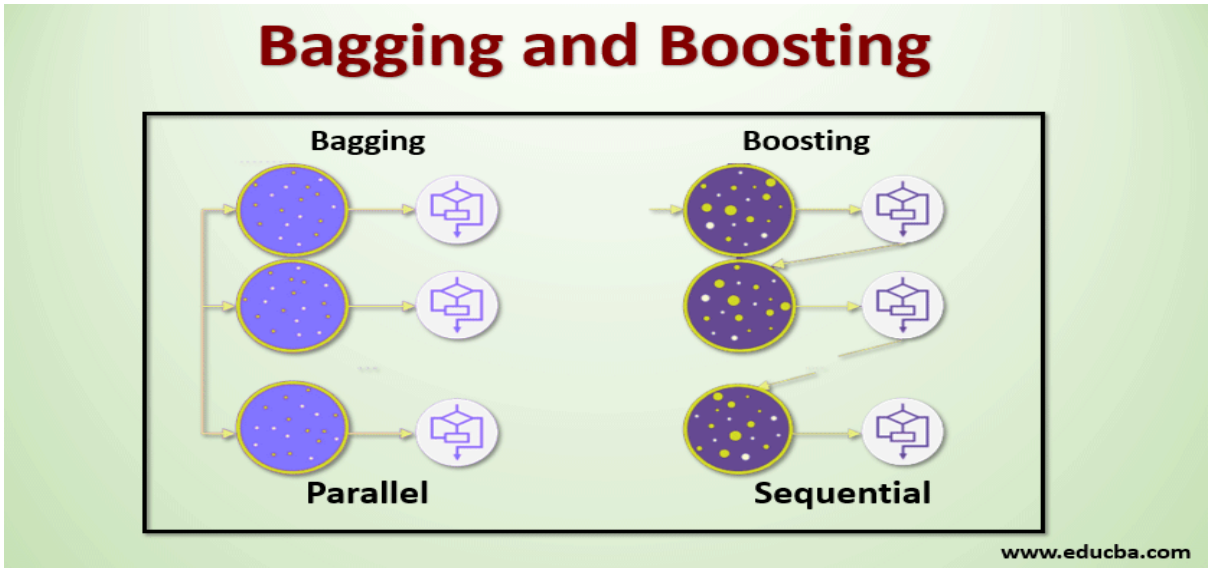


Fig (e) : Bagging Vs Boosting

Q8. Ans :

Out-of-bag(OOB) error is a method for measuring the prediction error of machine learning models that use Bagging ensemble method .

In bagging , some training instances may be sampled several times for any given predictor , while others may not be sampled at all . By the method of BaggingClassifier , it can be shown that only 63% of the training instances are sampled on average for each predictor . The remaining 37% instances that are not sampled are called out-of-bag instances .

In random forest , OOB error is the average error for each prediction , calculated using only the trees that did not have the data point in their bagging sample .

Q9. Ans :

K-fold cross velidation is a technique used in Machine learning to assess the performamce of the predictive models

1) K-cross validation first of all split the dataset into k subsets or folds
2) Then train the models on all the subsets except the (k-1) subset , that is for evaluating the trained model
3) Iterate k times with the different subset reserved for testing each time .
4) Average the model against each of the folds
5) Finalize the model
6) Test the model against test set

10.Ans :

Hyperparameter tuning is a process of finding a set of hyperparameter values for the machine learning algorithm while applying these algorithms to any dataset . That combination of hyperparameters maximize the model performance and helpful to produce better result with fewer errors .

Hyperparameter tuning in machine learning is done because it is an important process in model development . The correct selection of Hyperparameter leads to better model performance and better results .

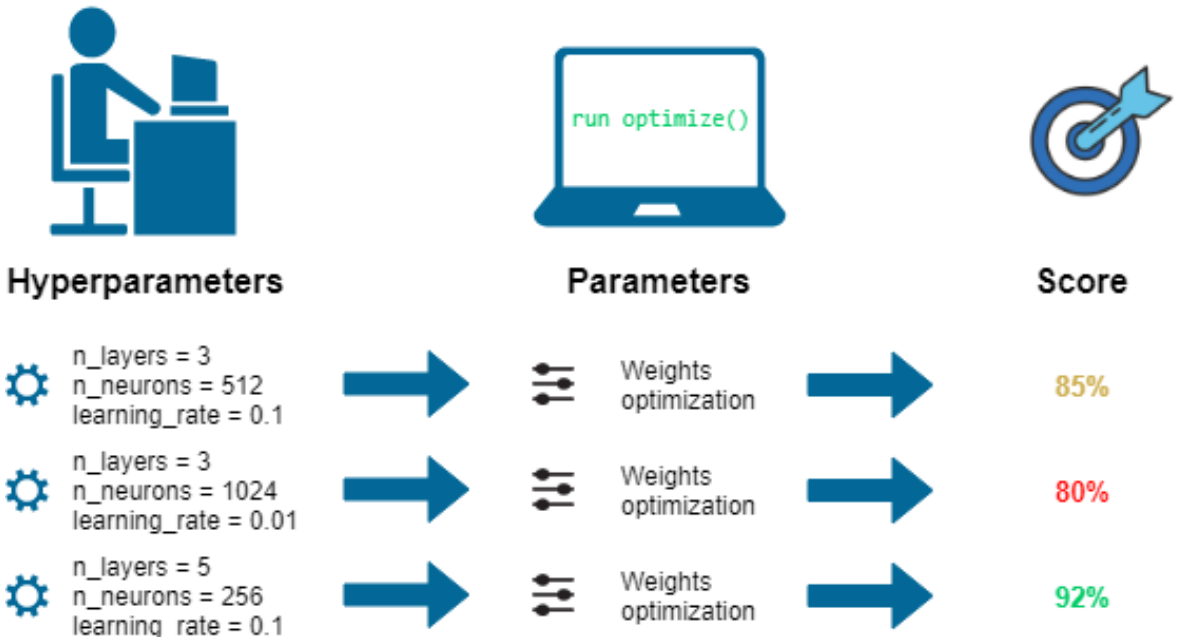


Fig (f) : Hyperparameter tuning

Q11. Ans :

Gradient descent is a generic optimization algorithm that is capable of finding optimal solution to a wide range of problems . The general Idea of gradient descent is to tweak parameters iteratively in order to minimize the cost function .

The Gradient descent measures the local gradient of the error function with regard to the parameter vector (theta) .

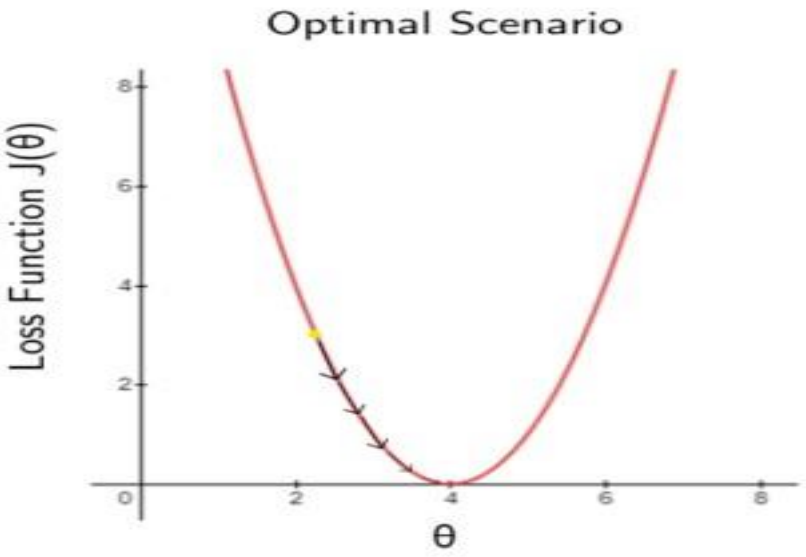


Fig (g) : Optimal scenario of gradient descent

The model parameters are initialized randomly and get tweaked repeatedly to minimize the cost function , the learning step size is proportional to the slope of the cost function . So the step gradually gets smaller as the cost approaches to minimum .

But when the learning rate is too high in gradient descent , this might make the algorithm diverge , with large and large values , failing to find a good solution .

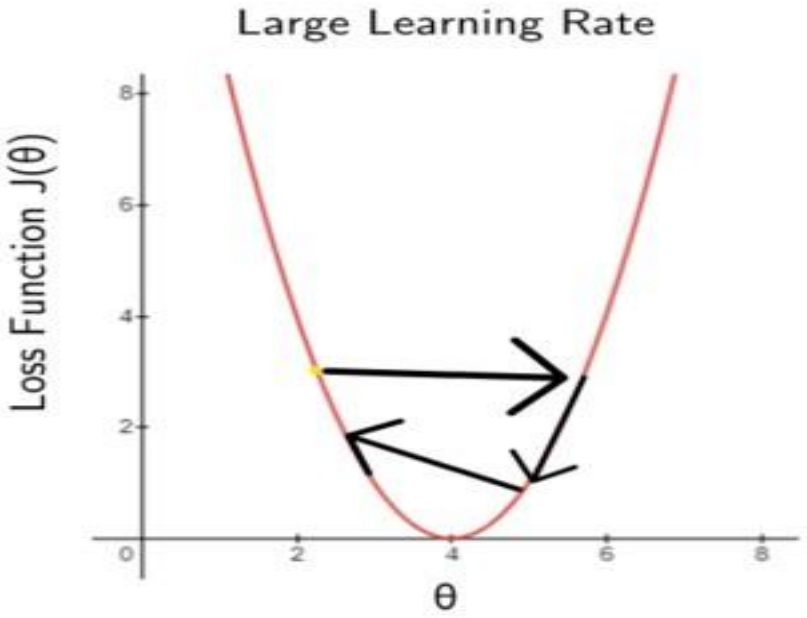


Fig (h) : Learning rate is too high

Q12. Ans :

Logistic Regression is a powerful classification tool but it is not suited for directly handle Non-linear data because –

It can only create a straight line to separate classes because logistics regression assumes a linear relationship between the features and the log odds of the target variables .

But Non-linear data often has complex decision boundaries that actually can not be captured by a straight line .

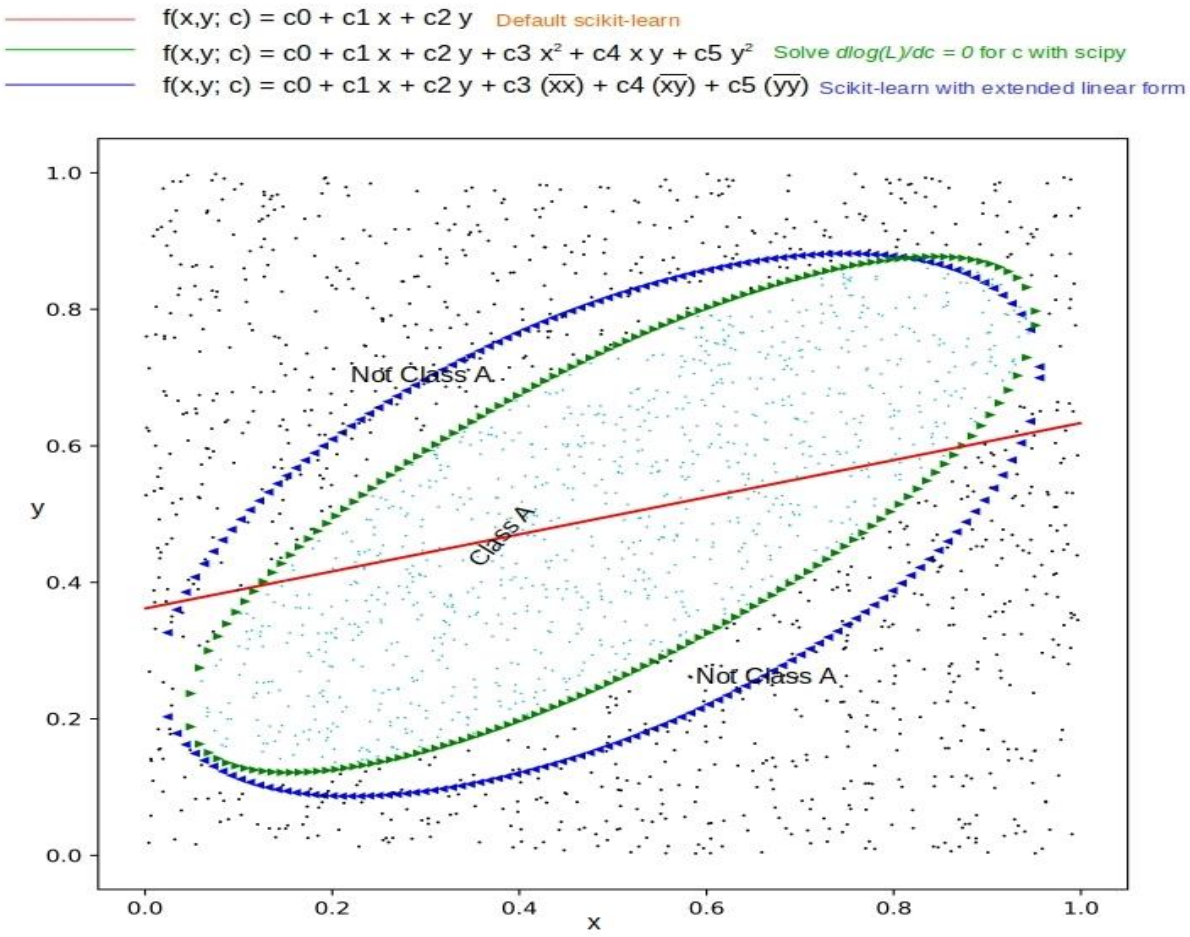


Fig (I) : Logistic regression for Non-Linear data

Q13. Ans :

AdaBoost is a boosting technique . Here , when training an AdaBoost classifier , the algorithm first trains a basic classifier and uses it to make predictions on the training set . The algorithm then increases the relative weight of misclassified training instances .Then it trains the second classifier using the updated weights and again makes the prediction on the training set , updates the instance weights and so on .

While , just like AdaBoost , gradient boosting works by sequentially adding predictors to an ensemble , each one correcting its predecessor , but instead of tweaking the instance weights at every iteration like AdaBoost , this method tries to fit the new predictor to the residual errors made by the previous predictor .

ASSIGNMENT – 5 MACHINE LEARING

Q14. Ans :

In Machine learning , the bias – variance trade off describes the relationship between model's complexity , the accuracy of it's predictions and how well it can make the predictions on the new fresh data that were not used to train the model .

Q15. Ans :

Linear Kernel : In SVM , linear kernel is used when the data is linearly separable . It is the kernel that is used most commonly in the dataset . Generally Linear kernel is used on those datasets where large number of features are there .

Training SVM with linear kernel is faster than with any other kernel .

Only the optimization of C regularization parameter is required while training the SVM with Linear kernel .

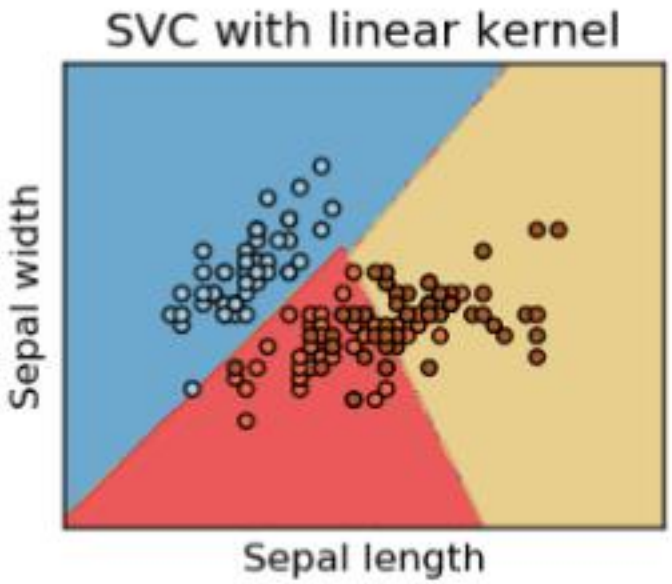
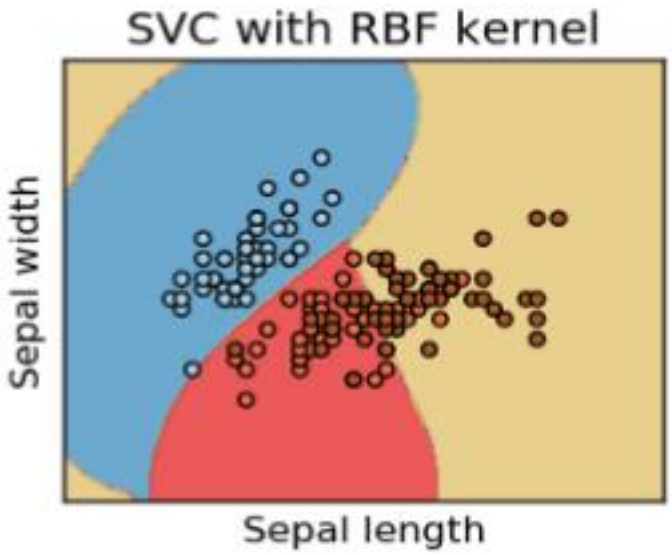


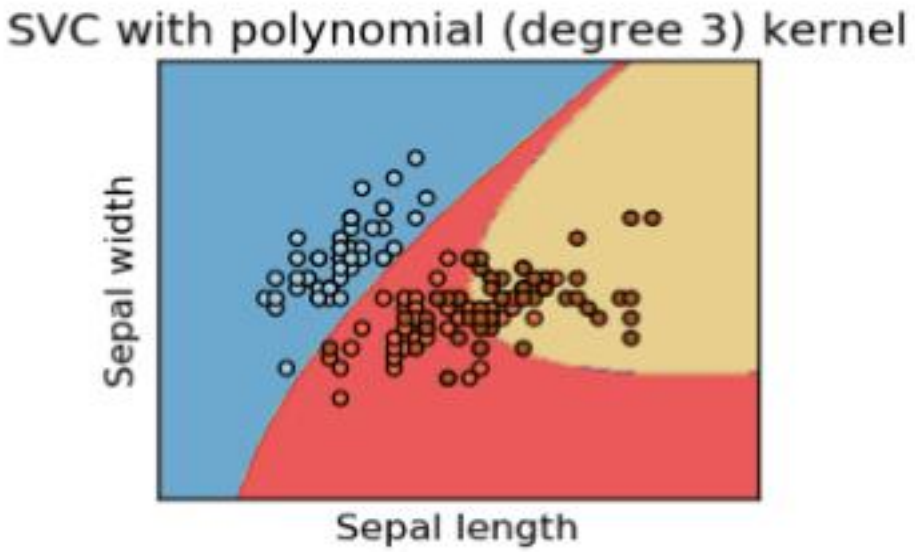
Fig (j) : SVC with linear kernel

RBF kernel : RBF is a kernel function used in SVM which measures the 'similarity' between two data points based on their Euclidean distance . The basic idea is that closer data points have a higher similarity score , while farther points has lower score . This similarity score than influences the decision boundary in the SVM .



Fig(k) : SVC with RBF kernel

Polynomial kernel : Polynomial kernel is used to capture non-linear relationship in data . It creates a higher dimensional feature space by mapping input data onto polynomial combinations of features . It allows SVM to learn non-linear decision boundaries , enhancing their ability to handle complex patterns .



Fig(l) : SVC with polynomial kernel (degree 3)

ASSIGNMENT – 5 MACHINE LEARING