**PRINCIPLE COMPONENT ANALYSIS (PCA)**

Linear dimensionality reduction using Singular Value Decomposition of

the data to project it to a lower dimensional space.

We get data with specified number of components, now we can apply models on it.

Number of components in data : 38

Number of components specified : 24

Model applied : Logistic regression

Accuracies :

KFold

0.8693798449612402

Training

0.8890173410404625

Test

0.8440860215053764

**SUPPORT VECTOR MACHINES (SVM)**

• It is intended for the binary classification setting in which there are two classes

• It can be extended to more than two classes also

• It performs well in a variety of settings

a) Maximal Marginal Classifier – It requires that the classes are separable by linear boundary. Cannot be applied to most data sets as they are not linearly separable

b) Support Vector Classifier – Extension of Maxima Marginal Classifier and can be applied to broader Range of classes where linearly separable criterion is relaxed

c) Support Vector Machine – Extension of Support Vector Classifier to accommodate non-linear class boundary

• The three together are also sometimes referred to as “Support Vector Machines”

Accuracy of SVM (linear) Classifier on training set: 0.902

Accuracy of SVM (linear) Classifier on test set: 0.860

Accuracy of SVM(rbf) Classifier on training set: 1.000

Accuracy of SVM(rbf) Classifier on test set: 0.817

Accuracy of SVM (sigmoid) Classifier on training set: 0.883

Accuracy of SVM (sigmoid) Classifier on test set: 0.847

Accuracy of SVM (Poly) Classifier on training set: 1.000

Accuracy of SVM (Poly) Classifier on test set: 0.820

Hyperparameter tunning

Linear model

Training accuracy : 0.900578034682081

Testing accuracy : 0.8602150537634409

RBF model

Training accuracy : 0.9017341040462428

Testing accuracy : 0.8494623655913979

Sigmoid model

Training accuracy : 0.8971098265895954

Testing accuracy : 0.8971098265895954

Poly model

Training accuracy : 0.991907514450867

Testing accuracy : 0.8279569892473119

**CLUSTERING TECHNIQUES**

1. **K- means clustering**

Aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster

1. **Hierarchical clustering**

Hierarchical clustering seeks to build a hierarchy of clusters

**Agglomerative** : It is a ‘bottom up' approach where each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy

**Divisive** : This is a 'top down' approach