# SUPPORT VECTOR MACHINES

* Support vector machines are used mainly in classification problems,they use the distance metrics .
* They handle multiple continuos and categorical variables.

## WORKING

* SVM is actually a representation of multiple classes in hyperplane in multidimensional space.
* Hyperplane is generated in such a manner by the svm such that error is minimized.
* Goal of svm is to divide the datasets into classes to find maximum marginal hyperplane.
* Margin is inversely proportional to error.

## Working of SVM

## IMPORTANT TERMS

* Support vectors – Datapoints closest to the hyperplanes,helps in determining the separating line.
* Hyperplane – Decision plane which is divided by a seprating line separating objects having different classes
* Margin – Distance between the lines nearest to the separation line,large margin is preferred.
* If line of separation is Wx + b=0 then the two lines are Wx +b =1 and Wx + b =-1 ,margin is 2/|W|,error is .

## SVM KERNELS

* SVM uses kernels which transform low dimensional input space to high dimensional ones.
* Kernels converts non-seperable problems to seperable problems by adding more dimensions.
* It makes SVM more powerful,flexible and accurate.

### LINEAR KERNEL

* Dot product between any two observations
* Formula :- K(x,xi)=sum(x\*xi)
* Product between two vectors x and xi is the sum of product of each pair of input values.

### POLYNOMIAL KERNEL

* Distinguishes between curved or nonlinear input space
* Formula :- K(x,xi)= 1+ sum(x\*xi)^d
* d is degree of the polynomial

### RADIAL BASIS FUNCTION

* Used to map input space in indefinite dimensional space
* Formula :- *K*(*x*,*xi*)=*exp*(−*gamma*∗*sum*(*x*−*xi*^2))
* Gamma ranges between 0 and 1
* Gamma needs to be manually specified.

#### PROS

* Great accuracy and work well with high dimensional space
* Consumes low memory

#### CONS

* Not suitable for large datasets as training time is large
* Does’nt work well with overlapping classes.

## ERRORS

* Total error = C\*Classification error + Margin Error,where C is the factor to which we can allow the misclassification
* Small C => More classification errors,large margin
* Large C => Less classification error,small margin.