1.) Classify the Size Categorie using SVM.

Solution:

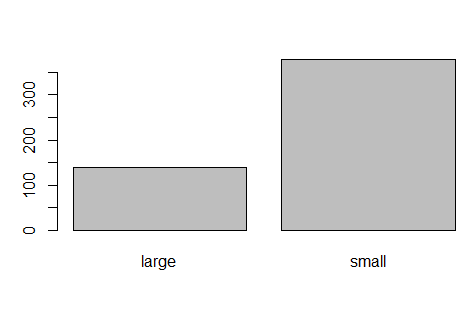
**Business Problem:** To build a model to classify size of forest fire

**Datasets:**

Independent Variable: month, day, FFMC, DMC, DC, ISI, temp, RH, wind, rain, area

Dependent Variable: size\_category

**EDA:**



There are 139 large and 378 small size category of forest fire

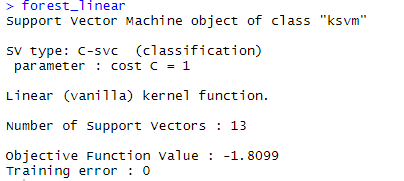
Data Partitioning

Data is partitioned into train and test data sets by 7: 3 ratio of entire forest data sets.

**Model Building :**

A model is built on train datasets using “ksvm ” algorithm form kernel lab package

Model Summary



Linear model is build using vanilladot as kernel method.

Number of support vectors are 13. Support vectors are those points which are affected by the increase of margin size. Error is less, it seems accuracy is high and is good model.

**Model Evaluation:**

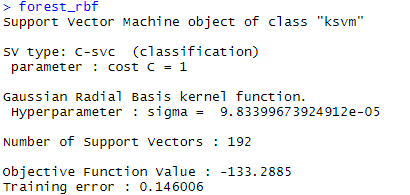
Model performance is checked on test data sets

Accuracy = 98.05%

Improvement of model performance

1.) Using rbfdot kernel

Model Summary

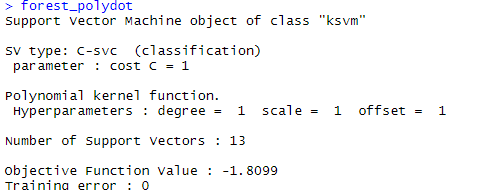


Here Error rate is 0.22865, Support vectors is 245

Accuracy = 82.467%

2.) Using polydot Kernel

Model Summary



Training error is 0.099174, Number of support vectors = 124

Accuracy = 98.05 %

**Better the accuracy better the model**

2.) Prepare a classification model using SVM for salary data

Solution:

**Business Problem:** To provide a model to classify salary data

**Datasets:**

**Independent Variable:** Salary

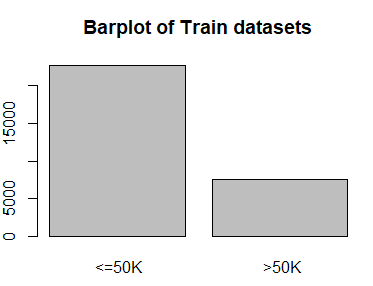
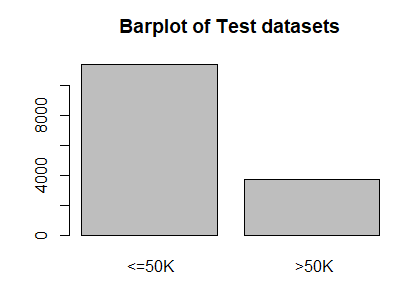
**Dependent Variable:** age, workclass, education educationno, maritalstatus,

occupation, relationship, race, sex, capitalgain, capitalloss, hoursperweek,

native

**EDA:**

There are two sets of data sets given as train and test data sets

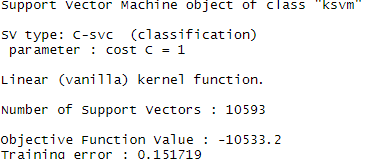
<=50k = 22653 <=50k = 11360

>50k = 7508 >50k = 3700

**Model Building :**

A model is built on train datasets using “ksvm ” algorithm form kernel lab package

Model Summary



Linear model is build using vanilladot as kernel method.

Number of support vectors are 10593. Support vectors are those points which are affected by the increase of margin size. Training error = 0.151719, it seems accuracy is high and is good model.

**Model Evaluation:**

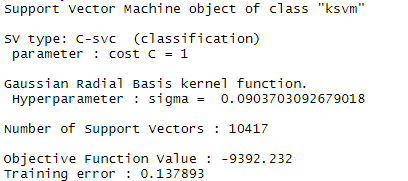
Model performance is checked on test data sets

Accuracy = 84.62%

Improvement of model performance

1.) Using rbfdot kernel

Model Summary

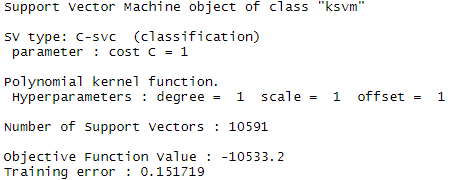


Here Error rate is 0.137893 , Support vectors are 10417, It seems better model then the previous model.

Accuracy = 85.43%

2.) Using polydot Kernel

Model Summary



Training error is 0.151719, Number of support vectors = 10591

Accuracy = 90 %

**Better the accuracy better the model**