**CSE – 6005 – Machine Learning**

**Lab Experiment – 11 -** Evaluate Machine Learning Algorithm for Balanced and Unbalanced Dataset

Here we took the default dataset available in the caret library “PimaIndiansDiabetes” dataset. Originally it’s a unbalanced dataset. I made it as a balanced dataset by deleting some transactions and applied the following code on the both datasets. The difference between many algorithms can be observed in the resulting graphs.

# load libraries

library(mlbench)

library(caret)

# load the dataset

data(PimaIndiansDiabetes)

# prepare training scheme

control <- trainControl(method="repeatedcv", number=10, repeats=3)

# CART

set.seed(7)

fit.cart <- train(diabetes~., data=PimaIndiansDiabetes, method="rpart", trControl=control)

# LDA

set.seed(7)

fit.lda <- train(diabetes~., data=PimaIndiansDiabetes, method="lda", trControl=control)

# SVM

set.seed(7)

fit.svm <- train(diabetes~., data=PimaIndiansDiabetes, method="svmRadial", trControl=control)

# kNN

set.seed(7)

fit.knn <- train(diabetes~., data=PimaIndiansDiabetes, method="knn", trControl=control)

# Random Forest

set.seed(7)

fit.rf <- train(diabetes~., data=PimaIndiansDiabetes, method="rf", trControl=control)

# collect resamples

results <- resamples(list(CART=fit.cart, LDA=fit.lda, SVM=fit.svm, KNN=fit.knn, RF=fit.rf))

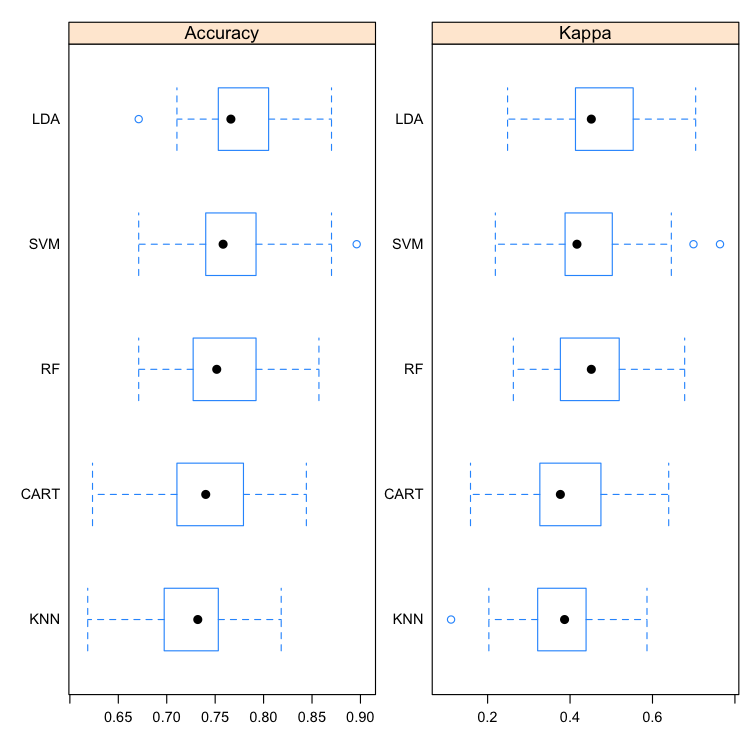
summary(results)

# box and whisker plots to compare models

scales <- list(x=list(relation="free"), y=list(relation="free"))

bwplot(results, scales=scales)

**Box Plot for UnBalanced Dataset**

****

**Box Plot for Balanced Dataset**

