Session-12-Assignment-1

Aditya Mulay

July 18, 2018

# 1. Use the given link Data Set.  
# Answer the below questions:  
Yeast <- read.table("C:/Users/Aditya/Desktop/Yeast.txt", quote="\"", comment.char="")  
View(Yeast)  
yeastdata <- Yeast  
dim(yeastdata)

## [1] 1484 10

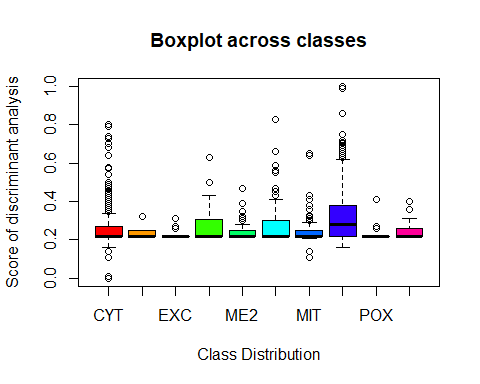
View(yeastdata)  
library(reshape)  
yeastdata <- rename (yeastdata, c(V1 = "Sequence Name",  
 V2 = "mcg",  
 V3 = "gvh",  
 V4 = "alm",  
 V5 = "mit",  
 V6 = "erl",  
 V7 = "pox",  
 V8 = "vac",  
 V9 = "nuc",  
 V10 = "Class Distribution"))  
View(yeastdata)

# # a. Perform ANOVA test on the discriminant analysis scores of nuclear localization signals of both nuclear and non-nuclear proteins by class variables (Target).  
results <- aov (yeastdata$nuc ~ yeastdata$`Class Distribution`)  
summary(results)

## Df Sum Sq Mean Sq F value Pr(>F)   
## yeastdata$`Class Distribution` 9 1.993 0.22141 22.01 <2e-16 \*\*\*  
## Residuals 1474 14.825 0.01006   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Since p value is significantly lesser than 0.05 we have strong evidence to reject null hypothesis and coclude that there is atleast one class significantly different than others.

#b. Which class is significantly different from others?  
  
plot(yeastdata$nuc ~ yeastdata$`Class Distribution`, col = rainbow(10), xlab = "Class Distribution", ylab = "Score of discriminant analysis" , main = "Boxplot across classes")



# Nuclear class is significantly different from others.