|  |
| --- |
| **Assignment 12.1**  **Problem Statement**  1. Use the given link Data Set.  Answer the below questions: |
| yeast <-read.table("C:/Users/Mymaster/Desktop/Data Analytics/Assignments\_Questions/yeast.txt", quote="\"", comment.char="") |
| View(yeast) |
| yeastdata<- yeast |
| dim(yeastdata) |
| install.packages("reshape") |
| 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) |
| #Answer the below questions: |
| # 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) |
|  |
| 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=heat.colors(10), xlab="Class Distribution", ylab="Score of discriminant analysis" , main="Boxplot across classes") |
|  |
| Nuclear class is significantly different from others. |