1. Use the given link Data Set.

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).

Ans:

>data<-read.table("http://archive.ics.uci.edu/ml/machine-learning-databases/yeast/yeast.data",

col.names=c("Sequence Name","mcg","gvh","alm","mit","erl","pox","vac","nuc","Class"))

>df<-data.frame(data$nuc,data$Class)

>class.ano<-aov(data$nuc~data$Class,data = df)

>summary(class.ano)

result:

Df Sum Sq Mean Sq F value Pr(>F)

data$Class 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

b. Which class is significantly different from others?

Ans:

as the p value is very small no class is significantly different from others

OR

>data<-read.table("http://archive.ics.uci.edu/ml/machine-learning-databases/yeast/yeast.data",

col.names=c("Sequence Name","mcg","gvh","alm","mit","erl","pox","vac","nuc","Class"))

>head(data)

>table(data$nuc)

>table(data$Class)

>df<-data.frame(data$nuc,data$Class)

>plot(data$nuc~data$Class,data = df)

by seeing the box plots we can say that all classes are same

no class is significantly different from others