

## Assignments for session on “Non- Linear Models”

**a. Perform ANOVA test on the discriminant analysis scores of nuclear localization signals of both nuclear and non-nuclear proteins by class variables (Target).**

```
yeast <- read.table('C:/Users/sarang_dani/Desktop/yeast.txt')
View(yeast)
yeastdata <- yeast
dim(yeastdata)
install.packages("reshape")
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)

#Perform the ANOVA between "nuc" and Target variable
results <- aov (yeastdata$nuc ~ yeastdata$`Class Distribution`)
summary(results)
```

```
> 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, the p-value is significantly less than 0.05, the null hypothesis is not accepted and so there is at least one class significantly different than others.

**b. Which class is significantly different from others?**

As per the below box plot we can conclude that Nuclear Class is significantly different than other classes.

**Boxplot across classes**

