

## R practical – Discriminant Analysis

1. Market The Smarket data set, part of the ISLR library, includes the percentage returns for the S&P 500 stock index over 1250 days, from the beginning of 2001 until the end of 2005. For each date, we have a record of the percentage returns for each of the five previous trading days (Lag1 to Lag5), on the number of shares traded on the previous day (Volume), the percentage return on the date in question (Today) and on whether the market was up (Up) or down (Down) on this date.

You can view information on the data using

```
install.packages("ISLR")
library(ISLR)
names(Smarket)
summary(Smarket)
cor(Smarket[, -9])
```

Comment on the correlation matrix produced using the code above. [Answer](#)

When interest lies in discrimination or classification, we often split the data into two parts and then use one part to fit the model and the other part to assess the predictive ability of the model.

We will first create a vector corresponding to the observations from 2001 through 2004 to which we'll perform our analysis and then create a held out data set of observations from 2005 to assess the performance of our methods in predicting the market (Up or Down) in 2005.

```
train <- (Smarket$Year < 2005)
Smarket_2005 <- Smarket [! train, ]
dim(Smarket_2005)
```

Perform linear discriminant analysis on the data before 2005 using Lag1 and Lag2

```
library(MASS)
lda_fit <- lda(Direction~Lag1+Lag2 ,data=Smarket, subset=train)
lda_fit
```

Comment on the LDA output. [Answer](#)

Plot the linear discriminant function and comment on the result

```
install.packages("klaR")
library(klaR)
partimat(x=Smarket[, 2:3], grouping=Smarket$Direction, method="lda",
```

```
col.mean=1, image.colors = c("grey", "white"), prec = 400)
```

### Answer

Using the LDA performed on the 2001-2004 data, predict the market in 2005 for each day using Lag1 and Lag2 and calculate the confusion matrix. Comment on the result.

```
lda_pred <- predict(lda_fit , Smarket_2005)
table(Smarket_2005$Direction, lda_pred$class)
```

### Answer

Now repeat the above steps for quadratic discriminant analysis and comment on the result.

```
qda_fit <- qda(Direction~Lag1+Lag2 ,data=Smarket ,subset =train)
partimat(x=Smarket[,2:3], grouping=Smarket$Direction, method="qda",
         col.mean=1, image.colors = c("grey", "white"), prec = 400)
qda_pred <- predict(qda_fit , Smarket_2005)
table(Smarket_2005$Direction, qda_pred$class)
```

### Answer

## 2. Diabetes

The Pima.tr data set in MASS includes information on women of Pima Indian heritage tested for diabetes. Use `?Pima.tr` to find out more.

Plot all pairwise scatter plots and comment on what you see. Perform LDA using all of the available variables and comment on the resulting confusion matrix.

### Answer