Model-Based Predictive Models

LDA & QDA

BAYES FORMULA

- The Bayes theorem gives us the following formula to compute the probability that the record belongs to class Ci:

$$P(C_i|X_1,\ldots,X_p) = \frac{P(X_1,\ldots,X_p|C_i)P(C_i)}{P(X_1,\ldots,X_p|C_1)P(C_1) + \cdots + P(X_1,\ldots,X_p|C_m)P(C_m)}.$$

Where

Ci : classes of interest

$X_1, X_2, \ldots X_p$: Variables which co-exist with Classes of interest

Bayes Theorem

$$P(C_i|X_1,\ldots,X_p) = \frac{P(X_1,\ldots,X_p|C_i)P(C_i)}{P(X_1,\ldots,X_p|C_1)P(C_1)+\cdots+P(X_1,\ldots,X_p|C_m)P(C_m)}.$$

- $P(C_i)$ are called prior probabilities. We can find them by dividing the incidences of occurrence of C_i by total number of observations.
- In place of $P\left(X_1, X_2, \ldots X_p \middle| C_i\right)$, we can also write a continuous function like probability density function of normal distribution as $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$

$$P(C_i|X_1) = \frac{P(C_i)\frac{1}{\sigma_i\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu_i}{\sigma_i})^2}}{\sum_{i=1}^{p} P(C_i)\frac{1}{\sigma_i\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu_i}{\sigma_i})^2}} \quad \ldots\ldots \text{(I)}$$

LDA – Univariate

- We can estimate the parameters (μ_k, σ_k^2) from the data and use the expression (I) as classifying the observation to that class i for which $P(C_i|X_1, X_2, \ldots X_p)$ will be maximum. But we have a better approach than this by solving this expression to $\delta_i(x)$ given below.
- We assume here $\sigma_i = \sigma$, a constant for all the classes.
- We can solve expression (I) by taking log of terms of both sides which finally results into the following expression

$$\boxed{\delta_i(x) = x\,\frac{\mu_i}{\sigma^2} - \frac{\mu_i^2}{2\sigma^2} + \log(\mathrm{P}(C_i))}$$

- Observe here that the function $\delta_i(x)$ is linear function in x. Hence the term Linear Discriminant Analysis.
- Each test observation is assigned to that class i, for which $\delta_i(x)$ is maximum. This function is a one-dimensional form of linear discriminating function.

Multivariate LDA

- The operations done on one variable can be extended to multiple variables and the expression $\delta_i(x)$ can be written as

$$\boxed{\delta_i(\bar{x}) \;=\; x^T \textstyle\sum^{-1} \mu_i - \frac{1}{2}\mu_i^T \textstyle\sum^{-1} \mu_i + \log(P(C_i))}$$

Where

$\sum$: Covariance Matrix

x: vector of variables x_i

μ_i: Mean of variable x_i

Note: We assume that the covariance matrix is same for all the classes

Multivariate QDA

- In Quadratic Discriminant Analysis, we assume that the covariance matrix $\sum$ is different for each class i.
- $\sum_i$: Covariance matrix for class i.
- Hence the discriminating function changes to

$$\boxed{\delta_k(\bar{x}) = -\frac{1}{2}(x-\mu_i)^T \sum\nolimits_i^{-1} (x-\mu_i) + \log(P(C_i))}$$

Assumptions of LDA & QDA

- Predictors are all numeric
- Predictors have a multivariate normal distribution
- LDA : All the variances and covariances for all the classes are same
- QDA : All the variances and covariances for each class is different

LDA & QDA in R

- LDA & QDA in R can be performed with the functions from the package **MASS**.

Syntax :

lda(formula, data, …)

qda(formula, data, …)

Example

- Consider the dataset Glass in the package **mlbench**
- A data frame with 214 observation containing examples of the chemical analysis of 6 different types of glass.
- The problem is to forecast the type of class on basis of the chemical analysis.
- The study of classification of types of glass was motivated by criminological investigation.
- At the scene of the crime, the glass left can be used as evidence (if it is correctly identified!)

R Program & Output

```
library(MASS)
fit.lda <- lda(Type ~ . , data = training)

pred.lda <- predict(fit.lda , newdata = validation)

confusionMatrix(pred.lda$class,validation$Type)
```

```
Confusion Matrix and Statistics

          Reference
Prediction  1  2  3  5  6  7
         1 17  5  3  0  0  0
         2  4 16  2  2  2  2
         3  0  0  0  0  0  0
         5  0  0  0  1  0  0
         6  0  1  0  0  0  0
         7  0  0  0  0  0  6

Overall Statistics

               Accuracy : 0.6557
                 95% CI : (0.5231, 0.7727)
    No Information Rate : 0.3607
    P-Value [Acc > NIR] : 2.711e-06

                  Kappa : 0.4931
```

Example

- Consider the dataset $\mathrm{Vehicle}$ from package mlbench.
- The purpose is to classify a given silhouette as one of four types of vehicle, using a set of features extracted from the silhouette.
- The vehicle may be viewed from one of many different angles.
- The features were extracted from the silhouettes by the HIPS (Hierarchical Image Processing System) extension BINATTS, which extracts a combination of scale independent features utilising both classical moments based measures such as scaled variance, skewness and kurtosis about the major/minor axes and heuristic measures such as hollows, circularity, rectangularity and compactness.

R Program & Output - LDA

```
##################### LDA ###########################
fit.lda <- lda(Class ~ . , data = training)

pred.lda <- predict(fit.lda , newdata = validation)

confusionMatrix(pred.lda$class,validation$Class)
```

```
Confusion Matrix and Statistics

          Reference
Prediction bus opel saab van
      bus   63    3    0   0
      opel   0   31   17   1
      saab   1   25   44   5
      van    1    4    4  53

Overall Statistics

               Accuracy : 0.7579
                 95% CI : (0.7002, 0.8095)
    No Information Rate : 0.2579
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.677
```

R Program & Output - QDA

```
###################### QDA ############################
fit.qda <- qda(Class ~ . , data = training)

pred.qda <- predict(fit.qda , newdata = validation)

confusionMatrix(pred.qda$class,validation$Class)
```

```
Confusion Matrix and Statistics

          Reference
Prediction bus opel saab van
      bus   64    0    0   0
      opel   0   44   18   3
      saab   0   18   45   3
      van    1    1    2  53

Overall Statistics

               Accuracy : 0.8175
                 95% CI : (0.7641, 0.8631)
    No Information Rate : 0.2579
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7565
```