Linear Discriminant Analysis

Jason

Wednesday, May 06, 2015

1. Linear Discriminant Analysis(LDA)

It mainly uses the Bayes' theroem. So, basically, we have two terms:

- Prior probability: $p(Y=y)=\pi_k$
- Density function: $f_k(x)=Pr(X=x|Y=y)$

And using above terms, we can compute posterior probability: $p_k=Pr(Y=y|X=x)=\frac{\pi_k f_k(x)}{\sum_{l=1}^{K}\pi_l f_l(x)}$

When estimaing the prior probability π_k, we simply calculate the sample proportion in our data, which is $\frac{n_k}{n}$. However, to obtain $f_k(x)$, we have to make some assumption that the density function is from an underlying distribution and then estimate the parameters by data. Normally, as we know, we will assume that it is from normal distribution.

$$f_k(x)=\frac{1}{\sqrt{2\pi}\sigma_k}exp(-\frac{(x-\mu_k)^2}{2\sigma_k^2})$$

If all σ are equal, we can get

$$p_k=\frac{\pi_k\frac{1}{\sqrt{2\pi}\sigma}exp(-\frac{(x-\mu_k)^2}{2\sigma^2})}{\sum_{l=1}^{K}\pi_l\frac{1}{\sqrt{2\pi}\sigma}exp(-\frac{(x-\mu_l)^2}{2\sigma^2})}$$

If the probability of this obervation belonging to l given that $X=x$ is the largest among $1\ k$, then we will classify it into l category. Because all the denominator are the same, we only have to compare the numerator, i.e.

$$\pi_k\frac{1}{\sqrt{2\pi}\sigma}exp(-\frac{(x-\mu_k)^2}{2\sigma^2})$$

After removing constant and take log, we get

$$\delta_k(x)=x\cdot\frac{\mu_k}{\sigma^2}-\frac{\mu_k^2}{2\sigma^2}+log(\pi_k)$$

For each parameter, we can use the following estimator to estimate their value.

- $\hat{\mu}_k=\frac{1}{n_k}\sum_{i:y_i=k}x_i$
- $\hat{\sigma}^2=\frac{1}{n-K}\sum_{k=1}^{K}\sum_{i:y_i=k}(x_i-\mu_i)^2$
- $\hat{\pi}_k=\frac{n_k}{n}$

When the number of independent variable, p, is bigger than one, our distributon will become multivariate normal distribution. Its pdf is:

$$f(x) = \frac{1}{(2\pi)^{p/2}(|\sum|)^{1/2}} exp(-\frac{1}{2}(x-\mu)^T \sum^{-1}(x-\mu))$$

We assume that the observations in the kth class are drawn from a multivariate normal distribution $N(\mu_k, \sum)$. As the approach in p=1, we can use Bayes' theroem to compute the probability and predict the observation is from kth group if its probability is the largest.

Assumptions

- Multivariate normal distribution for independent variables.
- Equal variance and covariance, i.e. same covariance matrix.

2. Textbook Example

```
library(ISLR)
library(car)

Default$default <- as.numeric(as.character(
recode(Default$default, "'Yes'='1'; 'No'='0'")))
```

```
library(MASS)
#Model fit
LDA <- lda(default ~ balance + student, data=Default)
#Fitted value
Prediction <- predict(LDA, Default)$class
t <- table(Predict=Prediction, True=Default$default)
#addmargins: comput all margin of the table
#ftable: make the table format nicer
ftable(addmargins(t))
```

```
##          True     0     1   Sum
## Predict
## 0              9644   252  9896
## 1                23    81   104
## Sum            9667   333 10000
```

```
threshold <- 0.2
Prediction_new <- (predict(LDA, Default)$posterior[, 2] > threshold)*1

t_new <- table(Predict=Prediction_new, True=Default$default)
ftable(addmargins(t_new))
```

```
##          True     0     1   Sum
## Predict
## 0              9432   138  9570
## 1               235   195   430
## Sum            9667   333 10000
```

Confusion matrix	True 0	True 1
Predicted 0	True negative(TN)	False negative(FN)
Predicted 1	False positive(FP)	True positive(TP)

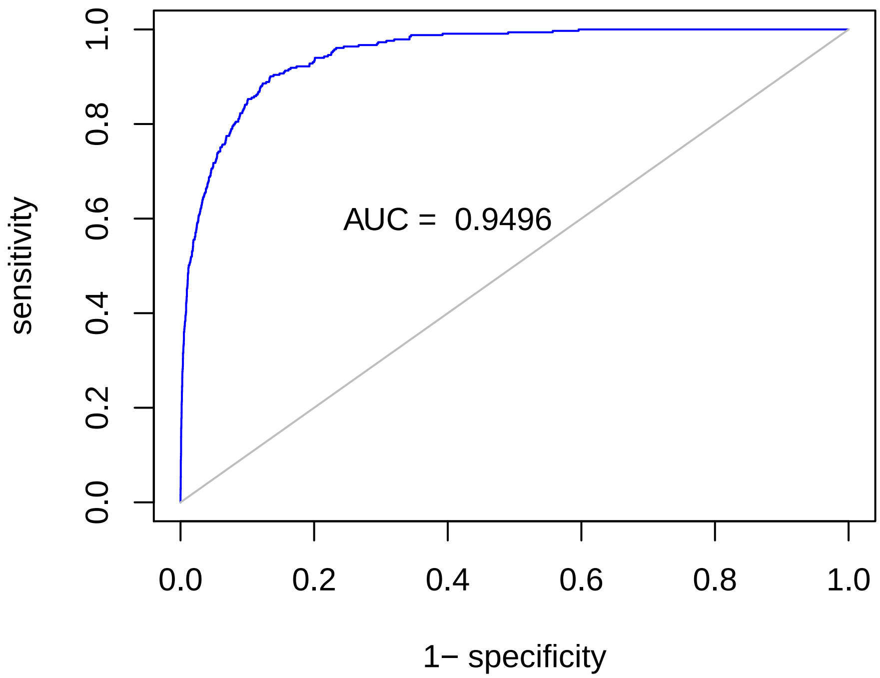
There are several measure that can help us to determine the performance of our model or classifier.

1. Accuracy: $\frac{TP+TN}{TP+FP+TN+FN}$
2. Specificity: $\frac{TN}{TN+FP}$
3. True Positive Rate (Sensitivity, Recall): $\frac{TP}{TP+FN}$
4. False Positive Rate (Type I error, 1 - Specificity): $\frac{FP}{TN+FP}$
5. Positive Predicted Value (Precision, 1 - False Discovery Rate): $\frac{TP}{TP+FP}$
6. Negative Predicted Value: $\frac{TN}{TN+FN}$

```
library(caret)
#It will give you all measure
confusionMatrix(t, positive="1")
```

```
## Confusion Matrix and Statistics
##
##        True
## Predict    0    1
##       0 9644  252
##       1   23   81
##
##                Accuracy : 0.9725
##                  95% CI : (0.9691, 0.9756)
##     No Information Rate : 0.9667
##     P-Value [Acc > NIR] : 0.0004973
##
##                   Kappa : 0.3606
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.2432
##             Specificity : 0.9976
##          Pos Pred Value : 0.7788
##          Neg Pred Value : 0.9745
##              Prevalence : 0.0333
##          Detection Rate : 0.0081
##    Detection Prevalence : 0.0104
##       Balanced Accuracy : 0.6204
##
##        'Positive' Class : 1
##
```

```r
library(AUC)
plot(roc(predict(LDA, Default)$posterior[, 2], as.factor(Default$default)),
     col="blue")
auc_value <- auc(roc(predict(LDA, Default)$posterior[, 2],
                     as.factor(Default$default)))
text(0.4, 0.6, paste("AUC = ", round(auc_value, 4)))
```



3. Textbook Graph

```r
x <- seq(-5, 5, by=0.01)
```

```r
par(mfrow=c(1, 2))

plot(x, dnorm(x, 1.25, 1), ylab="", xlim=c(-5, 5), type="l",
     lwd=2, col="mediumorchid3")
lines(x, dnorm(x, -1.25, 1), lwd=2, col="forestgreen")
abline(v=0, lty=2, lwd=2)

#Example
x1 <- rnorm(20, 1.25, 1)
x2 <- rnorm(20, -1.25, 1)
g <- rep(c(1, 2), each=20)
```

```r
hist(x2, col="forestgreen", xlim=c(-5, 5), border="forestgreen",
     main="", xlab="", ylab="")
hist(x1, add=T, col="mediumorchid3", density=20)
abline(v=0, lty=2, lwd=2)
abline(v=(mean(x1) + mean(x2))/2, lwd=2)
```