# 13) Linear Discriminant Analysis

Vitor Kamada

February 2018

# Reference

Tables, Graphics, and Figures from

**An Introduction to Statistical Learning**

James et al. (2017): Chapters: 4.4, 4.6.3, 4.6.4

# Bayes Rule

$$P(Y=k|X=x) = \frac{P(Y=k)P(X=x|Y=k)}{P(X=x)}$$

$$= \frac{P(Y=k)P(X=x|Y=k)}{\sum_{l=1}^{K} P(Y=l)P(X=x|Y=l)}$$

$$Posterior = \frac{Prior \times Likelihood}{Marginal}$$

# Bayes' Theorem for Classification

$$p_k(X) = Pr(Y = k|X = x)$$

$$= \frac{\pi_k f_k(x)}{\sum_{l=1}^{K} \pi_l f_l(x)}$$

$$f_k(x) = Pr(X = x|Y = k)$$

# Linear Discriminant Analysis (LDA)

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

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

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

## K=2, and $\pi_1 = \pi_2$

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

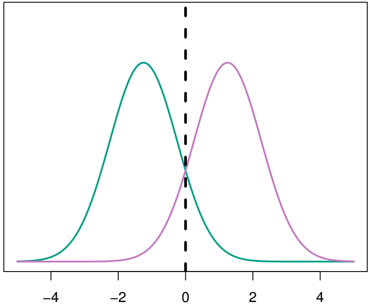
$$\delta_1 = \delta_2$$

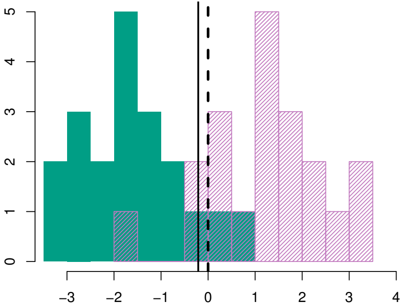
$$2x\mu_1 - \mu_1^2 = 2x\mu_2 - \mu_2^2$$

$$2x(\mu_1 - \mu_2) = \mu_1^2 - \mu_2^2$$

$$x = \frac{\mu_1^2 - \mu_2^2}{2(\mu_1 - \mu_2)} = \frac{\mu_1 + \mu_2}{2}$$

## $\mu_1 = -1.25,\ \mu_2 = 1.25,\ \sigma_1^2 = \sigma_2^2 = 1,$ **and** $\pi_1 = \pi_2 = 0.5$



$$\hat{\delta}_k(x) = x\frac{\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2} + log(\hat{\pi}_k)$$

$$\hat{\mu}_k = \frac{1}{n_k}\sum_{i:y_i=k} x_i$$

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

$$\hat{\pi}_k = \frac{n_k}{n}$$

# Default Data

| | | True default status | | |
|---|---|---|---|---|
| | | No | Yes | Total |
| *Predicted* | No | 9,644 | 252 | 9,896 |
| *default status* | Yes | 23 | 81 | 104 |
| | Total | 9,667 | 333 | 10,000 |

$$\text{Training Error: } \frac{23+252}{10000} = 2.75\%$$

$$\text{Specificity: } \frac{9644}{9667} \cong 99.8\%$$

$$\text{Sensitivity: } \frac{81}{333} \cong 24.3\%$$

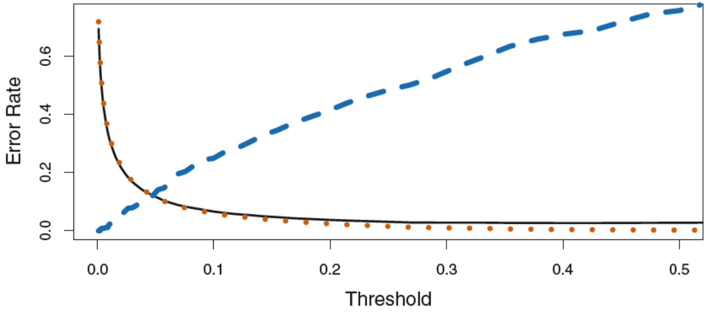
# $Pr(default = Yes|X = x) > 0.2$

| | | True default status | | |
|---|---|---|---|---|
| | | No | Yes | Total |
| *Predicted* | No | 9,432 | 138 | 9,570 |
| *default status* | Yes | 235 | 195 | 430 |
| | Total | 9,667 | 333 | 10,000 |

Training Error: $\frac{235+138}{10000} = 3.73\%$

Sensitivity: $\frac{195}{333} \cong 58.5\%$

# Error Rates as Function of Posterior Probability




**Black:** overall error rate

**Blue:** fraction of defaulting that are incorrectly classified

**Orange:** fraction of errors among the non-defaulting

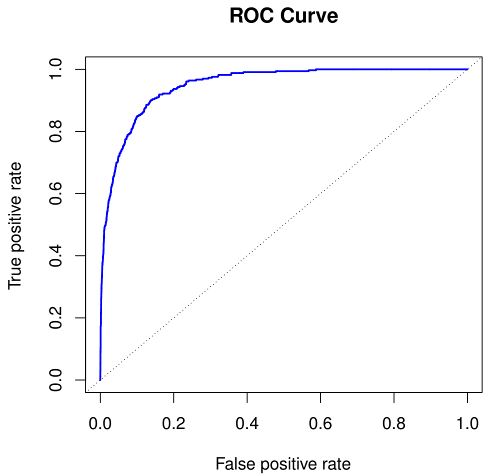
# Terminology

| | | *Predicted class* | | |
|---|---|---|---|---|
| | | − or Null | + or Non-null | Total |
| *True* | − or Null | True Neg. (TN) | False Pos. (FP) | N |
| *class* | + or Non-null | False Neg. (FN) | True Pos. (TP) | P |
| | Total | $N^*$ | $P^*$ | |

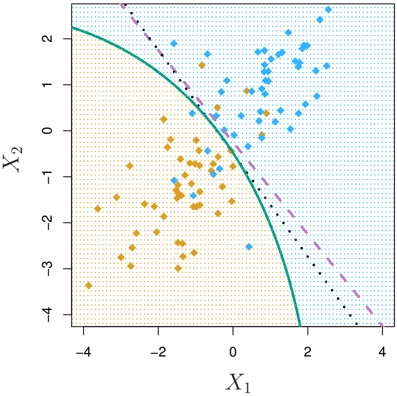
| Name | Definition | Synonyms |
|---|---|---|
| False Pos. rate | FP/N | Type I error, 1−Specificity |
| True Pos. rate | TP/P | 1−Type II error, power, sensitivity, recall |
| Pos. Pred. value | $TP/P^*$ | Precision, 1−false discovery proportion |
| Neg. Pred. value | $TN/N^*$ | |

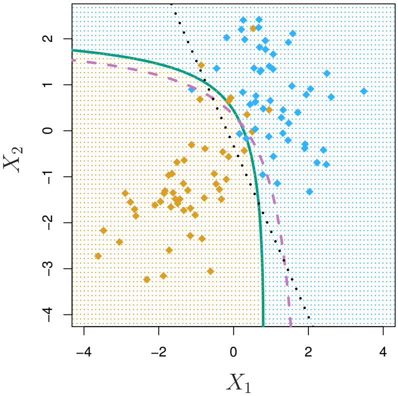
# Receiver Operating Characteristics (ROC)

$$AUC = 0.95$$



# LDA vs Quadratic Discriminant Analysis (QDA)



# Linear Discriminant Analysis in R

```
library(MASS); library(ISLR); attach(Smarket)
train=(Year<2005)
Smarket.2005=Smarket[!train,]
Direction.2005=Direction[!train]
lda.fit=lda(Direction~Lag1+Lag2,data=Smarket,
subset=train)
```

# lda.fit

```
Prior probabilities of groups:
    Down       Up
0.491984 0.508016

Group means:
            Lag1        Lag2
Down  0.04279022  0.03389409
Up   -0.03954635 -0.03132544

Coefficients of linear discriminants:
            LD1
Lag1 -0.6420190
Lag2 -0.5135293
```

# lda.pred=predict(lda.fit, Smarket.2005)

lda.class=lda.pred$class

table(lda.class,Direction.2005)

| | Direction.2005 | |
|---|---|---|
| lda.class | Down | Up |
| Down | 35 | 35 |
| Up | 76 | 106 |

mean(lda.class==Direction.2005)

55.95%

## lda.pred=predict(lda.fit, Smarket.2005)

lda.class=lda.pred$class

table(lda.class,Direction.2005)

```
          Direction.2005
lda.class Down  Up
     Down   35  35
     Up     76 106
```

mean(lda.class==Direction.2005)

55.95%

# qda.fit=qda(Direction~Lag1+Lag2,data=Smarket, subset=train)

qda.fit

```
Prior probabilities of groups:
    Down       Up
0.491984 0.508016

Group means:
            Lag1        Lag2
Down  0.04279022  0.03389409
Up   -0.03954635 -0.03132544
```

# qda.class=predict(qda.fit,Smarket.2005)$class

table(qda.class,Direction.2005)

```
           Direction.2005
qda.class Down  Up
     Down   30  20
     Up     81 121
```

mean(qda.class==Direction.2005)

59.92%