ME731 - Métodos em Análise Multivariada
– Análise discriminante I –

Prof. Carlos Trucíos
ctrucios@unicamp.br
ctruciosm.github.io

Instituto de Matemática, Estatística e Computação Científica,
Universidade Estadual de Campinas

Aula 21

Agenda I

Motivação

Motivação

Suponhas que estamos interessados em algum dos seguintes problemas:

1. Saber se devemos ou não dar linha de crédito a um determinado cliente.
2. Saber se, segundo um conjunto de caracteristicas (sintomas) uma pessoa está doente ou não.
3. Saber se uma nota de 100 reais (ou obra de arte, produto, etc) é original ou não.

Motivação

Suponhas que estamos interessados em algum dos seguintes problemas:

1. Saber se devemos ou não dar linha de crédito a um determinado cliente.
2. Saber se, segundo um conjunto de caracteristicas (sintomas) uma pessoa está doente ou não.
3. Saber se uma nota de 100 reais (ou obra de arte, produto, etc) é original ou não.

Como faria as análises?

Análise Discriminante

Análise Discriminante

- Faz parte do conjunto de técnicas dentro do escopo de "Aprendizagem supervisionado" (classificação supervisionada).
- Utilizando um conjunto de variáveis X, vamos a encontrar como discriminar (classificar) y.
- Utilizamos as regras de classificação para classificar novas observações das quais apenas conhecemos X mas não y.

Análise Discriminante

- Sejam P_1 e P_2, 2 populações nas quais $\mathbf{X}$, o vetor aleatório p-dimensional, está definido.
- Sejam $f_1(x)$ e $f_2(x)$ as respectivas densidades de $\mathbf{X}$ para cada uma das populações.

Análise Discriminante

- Sejam P_1 e P_2, 2 populações nas quais $\mathbf{X}$, o vetor aleatório p-dimensional, está definido.
- Sejam $f_1(x)$ e $f_2(x)$ as respectivas densidades de $\mathbf{X}$ para cada uma das populações.
- Estamos interessados em classificar uma nova observação $\mathbf{x}_0$ em alguma das populações (P_1 ou P_2).

Análise Discriminante

- Sejam P_1 e P_2, 2 populações nas quais $\mathbf{X}$, o vetor aleatório p-dimensional, está definido.
- Sejam $f_1(x)$ e $f_2(x)$ as respectivas densidades de $\mathbf{X}$ para cada uma das populações.
- Estamos interessados em classificar uma nova observação $\mathbf{x}_0$ em alguma das populações (P_1 ou P_2).
- Sejam π_1 e π_2 as probabilidade iniciais de que um elemento venha de alguma das populações (P_1 ou P_2).

Análise Discriminante

- Sejam P_1 e P_2, 2 populações nas quais $\mathbf{X}$, o vetor aleatório p-dimensional, está definido.
- Sejam $f_1(x)$ e $f_2(x)$ as respectivas densidades de $\mathbf{X}$ para cada uma das populações.
- Estamos interessados em classificar uma nova observação $\mathbf{x}_0$ em alguma das populações (P_1 ou P_2).
- Sejam π_1 e π_2 as probabilidade iniciais de que um elemento venha de alguma das populações (P_1 ou P_2).
- Uma vez observado x_0, podemos então calcular a probabilidade condicional de que pertença a P_1 ou P_2.

Análise Discriminante

$$P(1|x_0) = \frac{P(x_0|1)\pi_1}{P(x_0|1)\pi_1 + P(x_0|2)\pi_2} = \frac{f_1(x_0)\pi_1}{f_1(x_0)\pi_1 + f_2(x_0)\pi_2}$$

Análise Discriminante

$$P(1|x_0) = \frac{P(x_0|1)\pi_1}{P(x_0|1)\pi_1 + P(x_0|2)\pi_2} = \frac{f_1(x_0)\pi_1}{f_1(x_0)\pi_1 + f_2(x_0)\pi_2}$$

$$P(2|x_0) = \frac{P(x_0|2)\pi_2}{P(x_0|1)\pi_1 + P(x_0|2)\pi_2} = \frac{f_2(x_0)\pi_2}{f_1(x_0)\pi_1 + f_2(x_0)\pi_2}$$

Análise Discriminante

$$P(1|x_0) = \frac{P(x_0|1)\pi_1}{P(x_0|1)\pi_1 + P(x_0|2)\pi_2} = \frac{f_1(x_0)\pi_1}{f_1(x_0)\pi_1 + f_2(x_0)\pi_2}$$

$$P(2|x_0) = \frac{P(x_0|2)\pi_2}{P(x_0|1)\pi_1 + P(x_0|2)\pi_2} = \frac{f_2(x_0)\pi_2}{f_1(x_0)\pi_1 + f_2(x_0)\pi_2}$$

Assim, atribuimos x_0 à população com a probabilidade condicional maior:

- Atribuimos x_0 a P_1 se $P(1|x_0) > P(2|x_0)$ $(\equiv \pi_1 f_1(x_0) > \pi_2 f_2(x_0))$.
- Atribuimos x_0 a P_2 se $P(2|x_0) > P(1|x_0)$ $(\equiv \pi_2 f_2(x_0) > \pi_1 f_1(x_0))$.

Análise Discriminante

$$P(1|x_0) = \frac{P(x_0|1)\pi_1}{P(x_0|1)\pi_1 + P(x_0|2)\pi_2} = \frac{f_1(x_0)\pi_1}{f_1(x_0)\pi_1 + f_2(x_0)\pi_2}$$

$$P(2|x_0) = \frac{P(x_0|2)\pi_2}{P(x_0|1)\pi_1 + P(x_0|2)\pi_2} = \frac{f_2(x_0)\pi_2}{f_1(x_0)\pi_1 + f_2(x_0)\pi_2}$$

Assim, atribuimos x_0 à população com a probabilidade condicional maior:

- Atribuimos x_0 a P_1 se $P(1|x_0) > P(2|x_0)$ $(\equiv \pi_1 f_1(x_0) > \pi_2 f_2(x_0))$.
- Atribuimos x_0 a P_2 se $P(2|x_0) > P(1|x_0)$ $(\equiv \pi_2 f_2(x_0) > \pi_1 f_1(x_0))$.

Se $\pi_1 = \pi_2$, então

- Atribuimos x_0 a P_1 se $f_1(x_0) > f_2(x_0)$.
- Atribuimos x_0 a P_2 se $f_2(x_0) > f_1(x_0)$.

Análise Discriminante

Em algumas situações, o custo por classificar em P_2 quando na verdade era P_1 (ou classificar em P_1 quando na verdade era P_2) pode não ser o mesmo. Por exemplo

Análise Discriminante

Em algumas situações, o custo por classificar em P_2 quando na verdade era P_1 (ou classificar em P_1 quando na verdade era P_2) pode não ser o mesmo. Por exemplo

- Suponha que a máquina classifica erroneamente uma nota de R\$. 10 como uma de R\$. 100 e da o troco ao cliente.
- O custo do erro (para a empresa) é de R\$. 90. Já se acontecer o contrário, o cliente recebera o dinheiro de volta via SAC e o custo será apenas a taxa paga ao funcionário terceirizado do SAC (R\$ 5).

Análise Discriminante

Seja $C(i|j)$ o custo de atribuir uma observação ao grupo i quando na verdade pertence ao grupo j.

Análise Discriminante

Seja $C(i|j)$ o custo de atribuir uma observação ao grupo i quando na verdade pertence ao grupo j. Assim, temos que

	População Correta: P_1	População Correta: P_2
Classificação P_1	0	$C(1\|2)$
Classificação P_2	$C(2\|1)$	0

Análise Discriminante

Seja $C(i|j)$ o custo de atribuir uma observação ao grupo i quando na verdade pertence ao grupo j. Assim, temos que

	População Correta: P_1	População Correta: P_2
Classificação P_1	0	$C(1\|2)$
Classificação P_2	$C(2\|1)$	0

- No momento de classificar incluiremos esta informação.
- Então, estamos interessados em classificar de forma que minimizemos o custo esperado.

Análise Discriminante

- O custo esperado de classificar x_0 em P_2 é dado por

$$\mathbb{E}(\text{Custo de classificar em } P_2) = C(2|1)P(1|x_0) + 0P(2|x_0)$$

Análise Discriminante

- O custo esperado de classificar x_0 em P_2 é dado por

$$\mathbb{E}(\text{Custo de classificar em } P_2) = C(2|1)P(1|x_0) + 0P(2|x_0)$$

- O custo esperado de classificar x_0 em P_1 é dado por

$$\mathbb{E}(\text{Custo de classificar em } P_1) = C(1|2)P(2|x_0) + 0P(1|x_0)$$

Análise Discriminante

- O custo esperado de classificar x_0 em P_2 é dado por

$$\mathbb{E}(\text{Custo de classificar em } P_2) = C(2|1)P(1|x_0) + 0P(2|x_0)$$

- O custo esperado de classificar x_0 em P_1 é dado por

$$\mathbb{E}(\text{Custo de classificar em } P_1) = C(1|2)P(2|x_0) + 0P(1|x_0)$$

Assim, atribuimos x_0 a, por exemplo, P_2 se seu custo esperado é menor, *i.e*

$$C(2|1)P(1|x_0) < C(1|2)P(2|x_0),$$

$$\equiv C(2|1)\pi_1 f_1(x_0) < C(1|2)\pi_2 f_2(x_0) \equiv \frac{\pi_1 f_1(x_0)}{C(1|2)} < \frac{\pi_2 f_2(x_0)}{C(2|1)}$$

Análise Discriminante

Regra de decisão

$$x_0 \in P_2 \text{ se } \frac{\pi_1 f_1(x_0)}{C(1|2)} < \frac{\pi_2 f_2(x_0)}{C(2|1)}. \text{ Caso contrário, } x_0 \in P_1.$$

Análise Discriminante

Regra de decisão

$$x_0 \in P_2 \text{ se } \frac{\pi_1 f_1(x_0)}{C(1|2)} < \frac{\pi_2 f_2(x_0)}{C(2|1)}. \text{ Caso contrário, } x_0 \in P_1.$$

Assuma que f_1 e f_2 são as densidades da distribuição normal multivariada p-dimensional e que elas tem a mesma matriz de covariância.

Análise Discriminante

Regra de decisão

$x_0 \in P_2$ se $\dfrac{\pi_1 f_1(x_0)}{C(1|2)} < \dfrac{\pi_2 f_2(x_0)}{C(2|1)}$. Caso contrário, $x_0 \in P_1$.

Assuma que f_1 e f_2 são as densidades da distribuição normal multivariada p-dimensional e que elas tem a mesma matriz de covariância.

$$f_i(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu_i)'\Sigma^{-1}(x-\mu_i)} \tag{1}$$

Análise Discriminante

Regra de decisão

$$x_0 \in P_2 \text{ se } \frac{\pi_1 f_1(x_0)}{C(1|2)} < \frac{\pi_2 f_2(x_0)}{C(2|1)}. \text{ Caso contrário, } x_0 \in P_1.$$

Assuma que f_1 e f_2 são as densidades da distribuição normal multivariada p-dimensional e que elas tem a mesma matriz de covariância.

$$f_i(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu_i)'\Sigma^{-1}(x-\mu_i)} \tag{1}$$

Se substituirmos (1) na regra de decisão e aplicarmos logaritmo, temos que

Análise Discriminante

$$\frac{\pi_1 f_1(x_0)}{C(1|2)} < \frac{\pi_2 f_2(x_0)}{C(2|1)}$$

$$\log(f_1(x_0)) + \log(\pi_1/C(1|2)) < \log(f_2(x_0)) + \log(\pi_2/C(2|1))$$

$$-\frac{1}{2}(x_0-\mu_1)'\Sigma^{-1}(x_0-\mu_1)+\log(\frac{\pi_1}{C(1|2)}) < -\frac{1}{2}(x_0-\mu_2)'\Sigma^{-1}(x_0-\mu_2)+\log(\frac{\pi_2}{C(2|1)})$$

Análise Discriminante

$$\frac{\pi_1 f_1(x_0)}{C(1|2)} < \frac{\pi_2 f_2(x_0)}{C(2|1)}$$

$$\log(f_1(x_0)) + \log(\pi_1/C(1|2)) < \log(f_2(x_0)) + \log(\pi_2/C(2|1))$$

$$-\frac{1}{2}(x_0-\mu_1)'\Sigma^{-1}(x_0-\mu_1)+\log(\frac{\pi_1}{C(1|2)}) < -\frac{1}{2}(x_0-\mu_2)'\Sigma^{-1}(x_0-\mu_2)+\log(\frac{\pi_2}{C(2|1)})$$

Por definição, a distância de Mahalanobis ao quadrado entre x_0 e média da população i é $D_i^2 = (x_0 - \mu_i)'\Sigma^{-1}(x_0 - \mu_i)$

Análise Discriminante

$$\frac{\pi_1 f_1(x_0)}{C(1|2)} < \frac{\pi_2 f_2(x_0)}{C(2|1)}$$

$$\log(f_1(x_0)) + \log(\pi_1/C(1|2)) < \log(f_2(x_0)) + \log(\pi_2/C(2|1))$$

$$-\frac{1}{2}(x_0-\mu_1)'\Sigma^{-1}(x_0-\mu_1)+\log(\frac{\pi_1}{C(1|2)}) < -\frac{1}{2}(x_0-\mu_2)'\Sigma^{-1}(x_0-\mu_2)+\log(\frac{\pi_2}{C(2|1)})$$

Por definição, a distância de Mahalanobis ao quadrado entre x_0 e média da população i é $D_i^2 = (x_0 - \mu_i)'\Sigma^{-1}(x_0 - \mu_i)$

$$-\frac{1}{2}\underbrace{(x_0 - \mu_1)'\Sigma^{-1}(x_0 - \mu_1)}_{D_1^2} + \log(\frac{\pi_1}{C(1|2)}) < -\frac{1}{2}\underbrace{(x_0 - \mu_2)'\Sigma^{-1}(x_0 - \mu_2)}_{D_2^2} + \log(\frac{\pi_2}{C(2|1)})$$

Análise Discriminante

Classificamos x_0 em P_2 se

$$\frac{1}{2}D_2^2 - \log(\frac{\pi_2}{C(2|1)}) < \frac{1}{2}D_1^2 - \log(\frac{\pi_1}{C(1|2)}))$$

Análise Discriminante

Classificamos x_0 em P_2 se

$$\frac{1}{2}D_2^2 - \log(\frac{\pi_2}{C(2|1)}) < \frac{1}{2}D_1^2 - \log(\frac{\pi_1}{C(1|2)}))$$

Note que se $\pi_1 = \pi_2$ e $C(2|1) = C(1|2)$, então a regra se reduz a classicar x_0 em P_2 se

$$D_2^2 < D_1^2.$$

Análise Discriminante

$$
\begin{aligned}
D_2^2 &< D_1^2 \\
(x_0-\mu_2)'\Sigma^{-1}(x_0-\mu_2) &< (x_0-\mu_1)'\Sigma^{-1}(x_0-\mu_1) \\
-2\mu_2'\Sigma^{-1}x_0 + 2\mu_1\Sigma^{-1}x_0 + \mu_2'\Sigma^{-1}\mu_2 - \mu_1'\Sigma^{-1}\mu_1 &< 0 \\
2(\mu_1-\mu_2)'\Sigma^{-1}x_0 - (\mu_1-\mu_2)'\Sigma^{-1}(\mu_1+\mu_2) &< 0 \\
(\mu_1-\mu_2)'\Sigma^{-1}x_0 - (\mu_1-\mu_2)'\Sigma^{-1}\frac{(\mu_1+\mu_2)}{2} &< 0 \\
\underbrace{(\mu_1-\mu_2)'\Sigma^{-1}}_{\alpha'}[x_0 - \underbrace{\frac{(\mu_1+\mu_2)}{2}}_{\mu}] &< 0
\end{aligned}
$$

Análise Discriminante

$$D_2^2 < D_1^2$$
$$(x_0 - \mu_2)'\Sigma^{-1}(x_0 - \mu_2) < (x_0 - \mu_1)'\Sigma^{-1}(x_0 - \mu_1)$$
$$-2\mu_2'\Sigma^{-1}x_0 + 2\mu_1\Sigma^{-1}x_0 + \mu_2'\Sigma^{-1}\mu_2 - \mu_1'\Sigma^{-1}\mu_1 < 0$$
$$2(\mu_1 - \mu_2)'\Sigma^{-1}x_0 - (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 + \mu_2) < 0$$
$$(\mu_1 - \mu_2)'\Sigma^{-1}x_0 - (\mu_1 - \mu_2)'\Sigma^{-1}\frac{(\mu_1 + \mu_2)}{2} < 0$$
$$\underbrace{(\mu_1 - \mu_2)'\Sigma^{-1}}_{\alpha'}[x_0 - \underbrace{\frac{(\mu_1 + \mu_2)}{2}}_{\mu}] < 0$$

$x_0 \in P_2$ se $\alpha'(x_0 - \mu) < 0$ e $x_0 \in P_1$ se $\alpha'(x_0 - \mu) \geq 0$.

Análise Discriminante

Resumo

- $x_0 \in P_2$ se
$$\frac{f_1(x)\pi_1}{C(2|1)} < \frac{f_2(x)\pi_2}{C(1|2)}.$$
- Além do mais, sob normalidade e assumindo $\Sigma_1 = \Sigma_2$, $\pi_1 = \pi_2$ e $C(2|1) = C(1|2)$, $x_0 \in P_2$ se
$$D_2^2 < D_1^2 \quad \text{ou equivalentemente} \quad \underbrace{(\mu_1 - \mu_2)'\Sigma^{-1}}_{\alpha'}[x_0 - \underbrace{\frac{(\mu_1 + \mu_2)}{2}}_{\mu}] < 0$$

Implementação I

Implementação I

```
ad_me731 <- function(x, y, new_obs) {
  mu_1 <- apply(x, 2, mean)
  mu_2 <- apply(y, 2, mean)
  n_1 <- nrow(x)
  n_2 <- nrow(y)
  S <- ((n_1 - 1)*cov(x) + (n_2 - 1)*cov(y))/(n_1 + n_2 - 2)
  iS <- solve(S)
  mu <- 0.5*(mu_1 + mu_2)
  alpha <- iS %*% (mu_1 - mu_2)
  ifelse(t(alpha) %*% (new_obs - mu) < 0, "P2", "P1")
}
```

Implementação I

```
Sigma = matrix(c(1, 0.2, 0.2, 1), ncol = 2)
mu1 = c(1, 1)
mu2 = c(2.5, 2.5)
x = MASS::mvrnorm(100, mu1, Sigma)
y = MASS::mvrnorm(100, mu2, Sigma)
new_obs = MASS::mvrnorm(1000, mu1, Sigma)
classifica = c()
for (i in 1:1000) classifica[i] = ad_me731(x, y, new_obs[i, ])
table(classifica)
```

```
## classifica
##  P1  P2
## 838 162
```

Implementação I

```
Sigma = matrix(c(0.5, 0.2, 0.2, 0.5), ncol = 2)
mu1 = c(1, 1)
mu2 = c(2.5, 2.5)
x = MASS::mvrnorm(100, mu1, Sigma)
y = MASS::mvrnorm(100, mu2, Sigma)
new_obs = MASS::mvrnorm(1000, mu1, Sigma)
classifica = c()
for (i in 1:1000) classifica[i] = ad_me731(x, y, new_obs[i, ])
table(classifica)
```

```
## classifica
##  P1  P2
## 881 119
```

Implementação I

```
data <- data.frame(rbind(x, y))
data$Grupo <- c(rep("P1", nrow(x)), rep("P2", nrow(y)))
ad <- MASS::lda(Grupo ~ ., data)
classifica_lda <- predict(ad, data.frame(new_obs))$class
table(classifica_lda)
```

```
## classifica_lda
##  P1  P2
## 881 119
```

Implementação I

```
table(classifica, classifica_lda)
```

```
##             classifica_lda
## classifica  P1  P2
##         P1 881   0
##         P2   0 119
```

Implementação I

```
ad$prior    # Probabilidades iniciais (nossos Pis)
```

```
##  P1  P2
## 0.5 0.5
```

```
ad$means    # Vetores de médias
```

```
##           X1       X2
## P1 0.9709064 1.008752
## P2 2.5148020 2.415932
```

```
ad$scaling # O que acham que é?
```

```
##          LD1
## X1 0.8970200
## X2 0.7408648
```

Erro de classificação

Erro de classificação

- No exemplo vimos que, mesmo todos os $x_0 \in P_1$, algumas vezes foi classificado como P_2. Isto é um erro de classificação.

Erro de classificação

- No exemplo vimos que, mesmo todos os $x_0 \in P_1$, algumas vezes foi classificado como P_2. Isto é um erro de classificação.

$$\mathbb{P}(\text{Classificar em } P_2 | X \in P_1) = \mathbb{P}(\alpha'(X - \mu) < 0 | X \in P_1)$$

Erro de classificação

- No exemplo vimos que, mesmo todos os $x_0 \in P_1$, algumas vezes foi classificado como P_2. Isto é um erro de classificação.

$$\mathbb{P}(\text{Classificar em } P_2 | X \in P_1) = \mathbb{P}(\alpha'(X - \mu) < 0 | X \in P_1)$$

$$X \sim N(\mu_1, \Sigma)$$

$$X - \mu \sim N(\frac{\mu_1 - \mu_2}{2}, \Sigma)$$

$$\alpha'(X - \mu) \sim N(\underbrace{\alpha'(\frac{\mu_1 - \mu_2}{2})}_{\frac{(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)}{2}}, \underbrace{\alpha'\Sigma\alpha}_{(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)})$$

Erro de classificação

$$\alpha'(X-\mu) \sim N(\frac{(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)}{2}, (\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2))$$

Erro de classificação

$$\alpha'(X-\mu) \sim N(\frac{(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)}{2}, (\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2))$$

Se denotarmos $\delta^2 = (\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)$, então

$$\alpha'(X-\mu) \sim N(\frac{\delta^2}{2}, \delta^2)$$

Erro de classificação

$$\alpha'(X-\mu) \sim N\left(\frac{(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)}{2}, (\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)\right)$$

Se denotarmos $\delta^2 = (\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)$, então

$$\alpha'(X-\mu) \sim N\left(\frac{\delta^2}{2}, \delta^2\right)$$

Se $X \in P_1$:

$$\mathbb{P}(\alpha'(X-\mu) < 0) = \mathbb{P}\Bigg(\underbrace{\frac{\alpha'(X-\mu)-\frac{\delta^2}{2}}{\delta}}_{Z} < -\frac{\frac{\delta^2}{2}}{\delta}\Bigg) = \mathbb{P}\left(Z < -\frac{\delta}{2}\right)$$

Erro de classificação

Então, se temos n novas observações de P_1, esperamos que $n \times P(Z < -\frac{\delta}{2})$ sejam incorretamente classificadas.

Erro de classificação

Então, se temos n novas observações de P_1, esperamos que $n \times P(Z < -\frac{\delta}{2})$ sejam incorretamente classificadas.

No nosso caso:

```
n = 1000
delta = sqrt(t(mu1 - mu2) %*% solve(Sigma) %*% (mu1 - mu2))
c(pnorm(-delta/2), n * pnorm(-delta/2))
```

```
## [1]   0.1024469 102.4469469
```

Erro de classificação

- Como não conhecemos Σ, μ_1 nem μ_2, na prática estimamos essa probabilidade de má classificação

```
mu_1 <- apply(x, 2, mean)
mu_2 <- apply(y, 2, mean)
n_1 <- nrow(x)
n_2 <- nrow(y)
S <- ((n_1 - 1)*cov(x) + (n_2 - 1)*cov(y))/(n_1 + n_2 - 2)
iS <- solve(S)
delta = sqrt(t(mu_1 - mu_2) %*% iS %*% (mu_1 - mu_2))
c(pnorm(-delta/2), n * pnorm(-delta/2))
```

```
## [1]   0.1124277 112.4277476
```

Erro de classificação

- Temos calculado $\mathbb{P}(\text{Classificar em } P_2|X \in P_1)$ no caso em que $P_1 \sim N(\mu_1, \Sigma)$, $P_2 \sim N(\mu_2, \Sigma)$, $\pi_1 = \pi_2$, $C(2|1) = C(1|2)$.

Erro de classificação

- Temos calculado $\mathbb{P}(\text{Classificar em } P_2|X \in P_1)$ no caso em que $P_1 \sim N(\mu_1, \Sigma)$, $P_2 \sim N(\mu_2, \Sigma)$, $\pi_1 = \pi_2$, $C(2|1) = C(1|2)$.
- Podemos também calcular $\mathbb{P}(\text{Classificar em } P_1|X \in P_2)$

Erro de classificação

- Temos calculado $\mathbb{P}(\text{Classificar em } P_2|X \in P_1)$ no caso em que $P_1 \sim N(\mu_1, \Sigma)$, $P_2 \sim N(\mu_2, \Sigma)$, $\pi_1 = \pi_2$, $C(2|1) = C(1|2)$.
- Podemos também calcular $\mathbb{P}(\text{Classificar em } P_1|X \in P_2)$
- Calcular as probabilidades para diversos outros modelos não sempre é facil.

Erro de classificação

- Temos calculado $\mathbb{P}(\text{Classificar em } P_2 | X \in P_1)$ no caso em que $P_1 \sim N(\mu_1, \Sigma)$, $P_2 \sim N(\mu_2, \Sigma)$, $\pi_1 = \pi_2$, $C(2|1) = C(1|2)$.
- Podemos também calcular $\mathbb{P}(\text{Classificar em } P_1 | X \in P_2)$
- Calcular as probabilidades para diversos outros modelos não sempre é facil.
- Outra opção é estimar essas probabilidades é utilizando dados de *treinamento* e *teste* ou através de *validação cruzada* (alguns utiliam re-substituição mas este método é muito otimista).

Erro de classificação

- Temos calculado $\mathbb{P}(\text{Classificar em } P_2|X \in P_1)$ no caso em que $P_1 \sim N(\mu_1, \Sigma)$, $P_2 \sim N(\mu_2, \Sigma)$, $\pi_1 = \pi_2$, $C(2|1) = C(1|2)$.
- Podemos também calcular $\mathbb{P}(\text{Classificar em } P_1|X \in P_2)$
- Calcular as probabilidades para diversos outros modelos não sempre é facil.
- Outra opção é estimar essas probabilidades é utilizando dados de *treinamento* e *teste* ou através de *validação cruzada* (alguns utiliam re-substituição mas este método é muito otimista).
- Ambos os métodos permitem "pular" o cálculo dessas probilidades todas de forma analítica ao custo de um maior esforço computacional.

Análise Discriminante Quadrática

Análise Discriminante Quadrática

- Análise Discriminante Linear assume que a matriz de covariância Σ é a mesma para ambos os grupos.

Análise Discriminante Quadrática

- Análise Discriminante Linear assume que a matriz de covariância Σ é a mesma para ambos os grupos.
- O que acontece se $\Sigma_1 \neq \Sigma_2$?

Análise Discriminante Quadrática

- Análise Discriminante Linear assume que a matriz de covariância Σ é a mesma para ambos os grupos.
- O que acontece se $\Sigma_1 \neq \Sigma_2$? ADL perde performance.

Análise Discriminante Quadrática

- Análise Discriminante Linear assume que a matriz de covariância Σ é a mesma para ambos os grupos.
- O que acontece se $\Sigma_1 \neq \Sigma_2$? ADL perde performance.
- Uma alternativa é desenvolver todas as contas considerando $\Sigma_1 \neq \Sigma_2$. Neste caso, $x_0 \in P_2$ se

$$
\begin{aligned}
&-\frac{1}{2}x_0'(\Sigma_1^{-1} - \Sigma_2^{-1})x_0 + (\mu_1'\Sigma_1^{-1} - \mu_2'\Sigma_2^{-1})x_0 - \frac{1}{2}\log(\frac{|\Sigma_1|}{|\Sigma_2|}) \\
&+ \frac{1}{2}(\mu_1'\Sigma_1^{-1}\mu_1 - \mu_2'\Sigma_2^{-1}\mu_2) - \log\left[\frac{C(1|2)\pi_2}{C(2|1)\pi_1}\right]
\end{aligned}
$$

Referências

Referências

- Härdle, W. K., & Simar, L. (2019). Applied Multivariate Statistical Analysis. Fifth Editon. Springer Nature. Capítulo 14.
- Mardia, K. V., Kent, J. T., & Bibby, J, M. (1979). Multivariate Analysis. Academic Press. Capítulo 11.
- Peña, D. (2002). Análisis de Datos Multivariantes. Mc Graw Hill. Capítulo 13.