

Aprendizagem Estatística de Dados

Francisco Carvalho

Separação de Populações: Método de Fisher

➡ Duas Populações (Dois Grupos)

- Objetivo:
 - ▼ Transformar observações multi-variadas $\mathbf{x}$ em observações uni-variadas y de tal forma que as populações (classes) π_1 e π_2 sejam o mais separadas possível

Separação de Populações: Método de Fisher

➡ Duas Populações (Dois Grupos)

- Modelo:

- ▼ y é combinação linear dos $\mathbf{x}$
- ▼ Não é necessário supor que as populações π_1 e π_2 sejam normais
- ▼ Igualdade das matrizes de covariâncias de π_1 e π_2

Separação de Populações

Método de Fisher

➡ Duas Populações (Dois Grupos)

- Dados:

- ▼ n_1 observações de π_1 : $\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1}$

- ▼ $y_{11}, y_{12}, \dots, y_{1n_1}$ são as n_1 observações de π_1 expressas em termos de y

- ▼ n_2 observações de π_2 : $\mathbf{x}_{21}, \dots, \mathbf{x}_{2n_2}$

- ▼ $y_{21}, y_{22}, \dots, y_{2n_2}$ são as n_2 observações de π_2 expressas em termos de y

Separação de Populações

Método de Fisher

➡ Duas Populações (Dois Grupos)

- Critério: Separação entre π_1 e π_2
 - ▼ formulada em termos da diferença entre $\bar{y}_1$ e $\bar{y}_2$
 - ▼ expressa em unidades de desvio-padrão :

$$\text{Separação} = \frac{|\bar{y}_1 - \bar{y}_2|}{s_y}$$

Separação de Populações

Método de Fisher

➡ Duas Populações (Dois Grupos)

▼ onde

$$s_y^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2}$$

Separação de Populações

Método de Fisher

➡ Duas Populações (Dois Grupos)

- Lembre que:

$$\bar{\mathbf{x}}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} \mathbf{x}_{1j} \quad \mathbf{S}_1 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)^t$$

$$\bar{\mathbf{x}}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{x}_{2j} \quad \mathbf{S}_2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)^t$$

Separação de Populações

Método de Fisher

➡ Duas Populações (Dois Grupos)

- Note que:

$$s_y^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} = \hat{\mathbf{a}}^t \mathbf{S}_{com} \hat{\mathbf{a}}$$

$$\mathbf{S}_{com} = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{(n_1 - 1) + (n_2 - 1)}$$

Separação de Populações

Método de Fisher

➡ Duas Populações (Dois Grupos)

- Proposição

- ▼ Dentre todas as possíveis combinações lineares, é a combinação

$$\hat{y} = \hat{\mathbf{a}}^t \mathbf{x} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^t \mathbf{S}_{com}^{-1} \mathbf{x}$$

que maximiza a razão

Separação de Populações

Método de Fisher

➡ Duas Populações (Dois Grupos)

- Proposição

Distância entre as médias amostrais de y =
Variância amostral de y

$$= \frac{(\bar{y}_1 - \bar{y}_2)^2}{s_y^2} = \frac{(\hat{\mathbf{a}}^t \bar{\mathbf{x}}_1 - \hat{\mathbf{a}}^t \bar{\mathbf{x}}_2)^2}{\hat{\mathbf{a}}^t \mathbf{S}_{com} \hat{\mathbf{a}}} = \frac{(\hat{\mathbf{a}}^t \mathbf{d})^2}{\hat{\mathbf{a}}^t \mathbf{S}_{com} \hat{\mathbf{a}}}$$

Separação de Populações

Método de Fisher

➡ Duas Populações (Dois Grupos)

- Proposição

onde $\mathbf{d} = \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$

O máximo dessa razão é

$$D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^t \mathbf{S}_{com}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

- Observação: $\mathbf{S}_{com}$ tem inversa quando $n_1 + n_2 - 2 \geq p$

Separação de Populações

Método de Fisher

➡ Duas Populações (Dois Grupos)

- Regra de alocação baseada na função discriminante de Fisher

▼ Alocar $\mathbf{x}_0$ a π_1 se

$$\hat{y}_0 = \hat{\mathbf{a}}^t \mathbf{x}_0 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^t \mathbf{S}_{com}^{-1} \mathbf{x}_0 \geq \hat{m}$$

Separação de Populações

Método de Fisher

➡ Duas Populações (Dois Grupos)

▼ Onde

$$\begin{aligned}\hat{m} &= \frac{1}{2}(\bar{y}_1 + \bar{y}_2) = \frac{1}{2}(\hat{\mathbf{a}}^t \bar{\mathbf{x}}_1 + \hat{\mathbf{a}}^t \bar{\mathbf{x}}_2) \\ &= \frac{1}{2}((\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^t \mathbf{S}_{com}^{-1} \bar{\mathbf{x}}_1 + (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^t \mathbf{S}_{com}^{-1} \bar{\mathbf{x}}_2) \\ &= \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^t \mathbf{S}_{com}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)\end{aligned}$$

Separação de Populações

Método de Fisher

➡ Duas Populações (Dois Grupos)

- Regra de alocação baseada na função discriminante de Fisher

▼ Alocar $\mathbf{x}_0$ a π_1 se

$$\hat{y}_0 - \hat{m} \geq 0$$

Alocar $\mathbf{x}_0$ a π_2 se

$$\hat{y}_0 < \hat{m}$$

ou se

$$\hat{y}_0 - \hat{m} < 0$$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Motivação
 - ▼ Obter um conjunto reduzido de variáveis,
 - ▼ combinações lineares das observações (tais como $\mathbf{a}_1^t \mathbf{x}$, $\mathbf{a}_2^t \mathbf{x}$, ...),
 - ▼ de tal modo a separar ao máximo as classes (populações)

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Interesse

- ▼ Representação das g populações (classes)
 - ⊠ em um espaço de dimensão reduzida
 - ⊠ que separa ao máximo as populações
- ▼ Regra visual ou automática de classificação
- ▼ etc

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Consideremos g populações para as quais $\mathbf{X}_{p \times 1}$ é definida
- As matrizes de covariância das g populações são consideradas iguais e de posto completo

$$\Sigma_1 = \cdots = \Sigma_g = \Sigma_{p \times p}$$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Observação

Se $\Sigma_{p \times p}$ não é de posto completo deve-se reduzir a dimensão do espaço de observação como segue

- ▼ calcula-se os autovalores e auto-vetores de $\Sigma_{p \times p}$

- ⊠ retendo-se aqueles autovalores não nulos e seus correspondentes auto-vetores

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

▼ Define-se

$$\mathbf{P}_{p \times q} = \begin{bmatrix} \mathbf{e}_1, \dots, \mathbf{e}_q \\ p \times 1 \quad p \times 1 \end{bmatrix}$$

⊠ matriz de auto-vetores correspondentes aos autovalores não nulos de $\Sigma_{p \times p}$

▼ Substitui-se $\mathbf{X}_{p \times 1}$ por $\mathbf{X}^*_{q \times 1} = \mathbf{P}^t_{q \times p} \mathbf{X}_{p \times 1}$ cuja matriz de covariâncias será

$$\otimes \text{cov}(\mathbf{X}^*) = \text{cov}(\mathbf{P}^t \mathbf{X}) = \mathbf{P}^t \text{cov}(\mathbf{X}) \mathbf{P} = \mathbf{P}^t \Sigma \mathbf{P}$$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Sejam

$$\underset{p \times 1}{\mu_1}, \dots, \underset{p \times 1}{\mu_g}$$

os vetores média das g classes (populações)

Então

$$\underset{p \times 1}{\bar{\mu}} = \frac{1}{g} \sum_{i=1}^g \underset{p \times 1}{\mu_i}$$

é a média combinada dos g grupos

Separação de Populações

Método de Fisher

- ➡ Várias Populações (Vários Grupos)
 - Seja a matriz de dispersão inter-grupos populacional

$$\mathbf{B}_0 = \sum_{i=1}^g (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^t$$

$p \times p$ $i=1$ $p \times l$ $l \times p$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Consideremos a combinação linear (escalar)

$$Y = \underset{l \times p}{\mathbf{a}^t} \underset{p \times 1}{X}$$

Temos:

$$\left. \begin{aligned} \varepsilon(Y \mid \pi_i) &= \varepsilon(\mathbf{a}^t X \mid \pi_i) = \mathbf{a}^t \varepsilon(X \mid \pi_i) = \mathbf{a}^t \underset{l \times p}{\mu_i} \underset{p \times 1}{} \\ \text{Var}(Y) &= \text{Var}(\mathbf{a}^t X) = \mathbf{a}^t \text{cov}(X) \mathbf{a} = \mathbf{a}^t \underset{l \times p}{\Sigma} \underset{p \times p}{\mathbf{a}} \underset{p \times 1}{} \end{aligned} \right\}$$
$$i = 1, \dots, g$$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Temos então:

$$\mu_{iY} = \mathbf{a}^t \mu_i$$

Para a população π_i :

$$\sigma_Y^2 = Var(Y) = \mathbf{a}^t \Sigma \mathbf{a}$$

Para todas as populações:

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Temos então:

Média global:

$$\bar{\mu}_Y = \frac{1}{g} \sum_{i=1}^g \mu_{iY} = \frac{1}{g} \sum_{i=1}^g \mathbf{a}_{l \times p}^t \mu_{p \times 1, i} = \mathbf{a}_{l \times p}^t \left[\frac{1}{g} \sum_{i=1}^g \mu_{p \times 1, i} \right] = \mathbf{a}_{l \times p}^t \bar{\mu}_{p \times 1}$$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Seja a razão

Soma das distâncias de cada média em relação
a média global de Y para cada população

Variância de Y

$$= \frac{\sum_{i=1}^g (\mu_{iY} - \bar{\mu}_Y)^2}{\sigma_Y^2} = \frac{\sum_{i=1}^g \left(\underset{l \times p}{\mathbf{a}^t} \underset{p \times 1}{\mu_i} - \underset{l \times p}{\mathbf{a}^t} \underset{p \times 1}{\bar{\mu}} \right)^2}{\mathbf{a}^t \Sigma \mathbf{a}} =$$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Seja a razão

$$\begin{aligned} &= \frac{\sum_{i=1}^g \left(\mathbf{a}^t (\mu_i - \bar{\mu}) \right)^2}{\mathbf{a}^t \Sigma \mathbf{a}} = \frac{\sum_{i=1}^g \mathbf{a}^t (\mu_i - \bar{\mu}) (\mu_i - \bar{\mu})^t \mathbf{a}}{\mathbf{a}^t \Sigma \mathbf{a}} = \\ &= \frac{\mathbf{a}^t \left[\sum_{i=1}^g (\mu_i - \bar{\mu}) (\mu_i - \bar{\mu})^t \right] \mathbf{a}}{\mathbf{a}^t \Sigma \mathbf{a}} = \frac{\mathbf{a}^t \mathbf{B}_0 \mathbf{a}}{\mathbf{a}^t \Sigma \mathbf{a}} \end{aligned}$$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- A razão

$$\frac{\sum_{i=1}^g (\mu_{iY} - \bar{\mu}_Y)^2}{\sigma_Y^2} = \frac{\mathbf{a}^t \mathbf{B}_0 \mathbf{a}}{\mathbf{a}^t \Sigma \mathbf{a}}$$

mede a variabilidade inter-grupos em relação a variabilidade intra-grupos

- Trata-se de selecionar $\mathbf{a}$ que maximiza essa razão

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Proposição(Discriminantes populacionais de Fisher)
 - ▼ Seja $\lambda_1 \geq \dots \lambda_s > 0$, onde $s \leq \min(g-1, p)$ os autovalores não nulos de $\Sigma^{-1}\mathbf{B}_0$ e
 - ▼ $\mathbf{e}_1, \dots, \mathbf{e}_s$ os correspondentes auto-vetores, normalizados pela relação
 - ▼ $\mathbf{e}_j^t \Sigma \mathbf{e}_j = 1, j=1, \dots, s.$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Proposição (Discriminantes populacionais de Fisher)

Então o vetor $\mathbf{a}$ que maximiza a razão

$$\frac{\mathbf{a}^t \mathbf{B}_0 \mathbf{a}}{\mathbf{a}^t \Sigma \mathbf{a}}$$

é $\mathbf{a}_1 = \mathbf{e}_1$ (vetor próprio associado a λ_1 , maior valor próprio de $\Sigma^{-1} \mathbf{B}_0$)

$Y_1 = \mathbf{a}_1 \mathbf{X}$ é denominado *primeiro discriminante*

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Proposição (Discriminantes populacionais de Fisher)

O vetor $\mathbf{a}$ que maximiza a razão

$$\frac{\mathbf{a}^t \mathbf{B}_0 \mathbf{a}}{\mathbf{a}^t \Sigma \mathbf{a}}$$

sujeito a $\text{cov}(Y_1, Y_2) = \text{cov}(\mathbf{a}_1 \mathbf{X}, \mathbf{a}_2 \mathbf{X}) = \mathbf{a}_1 \Sigma \mathbf{a}_2 = 0$ é $\mathbf{a}_2 = \mathbf{e}_2$ (vetor próprio associado a λ_2 , segundo maior valor próprio de $\Sigma^{-1} \mathbf{B}_0$)

$Y_2 = \mathbf{a}_2 \mathbf{X}$ é denominado *segundo discriminante*

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Proposição (Discriminantes populacionais de Fisher)

De um modo geral, o vetor $\mathbf{a}$ que maximiza a razão

$$\frac{\mathbf{a}^t \mathbf{B}_0 \mathbf{a}}{\mathbf{a}^t \Sigma \mathbf{a}}$$

sujeito a $\text{cov}(Y_k, Y_i) = \text{cov}(\mathbf{a}_k \mathbf{X}, \mathbf{a}_i \mathbf{X}) = \mathbf{a}_k \Sigma \mathbf{a}_i = 0$, $i < k$, é $\mathbf{a}_k = \mathbf{e}_k$ (vetor próprio associado a λ_k).

$Y_k = \mathbf{a}_k \mathbf{X}$ é denominado o *k-esimo discriminante*

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Em geral, Σ e μ_i , $i = 1, \dots, g$ são desconhecidos.

Suponha que o conjunto de treinamento consiste de amostras aleatórias

n_i observações da população (classe) π_i , $i = 1, \dots, g$

Seja X_i a matriz de dados da população π_i e $\mathbf{x}_{ij}$ a j -ésima observação para $i = 1, \dots, g$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Então

Vetores de média amostral

$$\hat{\mu}_i = \bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij}, i = 1, \dots, g$$

$p \times 1$ $p \times 1$ n_i $p \times 1$

Matrizes de covariâncias

$$\mathbf{S}_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^t = \hat{\Sigma}_i, i = 1, \dots, g$$

$p \times p$ $n_i - 1$ $p \times 1$ $1 \times p$ $p \times p$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Vetor de médias global

$$\hat{\bar{\boldsymbol{\mu}}}_{p \times 1} = \bar{\mathbf{X}}_{p \times 1} = \frac{\sum_{i=1}^g n_i \bar{\mathbf{X}}_i}{\sum_{i=1}^g n_i} = \frac{\sum_{i=1}^g \sum_{j=1}^{n_i} \mathbf{X}_{ij}}{\sum_{i=1}^g n_i}$$

Matriz de dispersão inter-grupos amostral

$$\hat{\mathbf{B}}_0 = \sum_{i=1}^g n_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^t$$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Matriz de dispersão intra-grupos amostral

$$\mathbf{W}_{p \times p} = \sum_{i=1}^g (n_i - 1) \mathbf{S}_i = \sum_{i=1}^g \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i) (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^t$$

$p \times 1$ $1 \times p$

Estimativa de Σ

$$\hat{\Sigma}_{p \times p} = \frac{\mathbf{W}}{\sum_{i=1}^g n_i - 1} = \mathbf{S}_{com}$$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Observações

- ▼ $\mathbf{W} = [(n_1 + \dots + n_g) - g] \mathbf{S}_{com}$

- ▼ O mesmo $\hat{\mathbf{a}}$ que maximiza

$$\frac{\hat{\mathbf{a}}^t \hat{\mathbf{B}}_0 \hat{\mathbf{a}}}{\hat{\mathbf{a}}^t \mathbf{S}_{com} \hat{\mathbf{a}}}$$

também maximiza

$$\frac{\hat{\mathbf{a}}^t \hat{\mathbf{B}}_0 \hat{\mathbf{a}}}{\hat{\mathbf{a}}^t \mathbf{W} \hat{\mathbf{a}}}$$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Problema

Encontrar $\hat{\mathbf{a}}$ que maximiza

$$\frac{\hat{\mathbf{a}}^t \hat{\mathbf{B}}_0 \hat{\mathbf{a}}}{\hat{\mathbf{a}}^t \mathbf{W} \hat{\mathbf{a}}}$$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Proposição (Discriminantes amostrais de Fisher)

Seja $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_s > 0$ onde $s \leq \min(g-1, p)$ os autovalores não nulos de $\mathbf{W}^{-1}\hat{\mathbf{B}}_0$ e $\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_s$ os correspondentes auto-vetores, normalizados pela relação $\hat{\mathbf{e}}^t \mathbf{S}_{com} \hat{\mathbf{e}} = 1$

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Então o vetor $\hat{\mathbf{a}}$ que maximiza a razão

$$\frac{\hat{\mathbf{a}}^t \hat{\mathbf{B}}_0 \hat{\mathbf{a}}}{\hat{\mathbf{a}}^t \mathbf{W} \hat{\mathbf{a}}}$$

$$\text{é } \hat{\mathbf{a}}_1 = \hat{\mathbf{e}}_1$$

$\hat{Y}_1 = \hat{\mathbf{a}}_1^t \mathbf{x}$ é chamado de *primeira discriminante amostral*

Separação de Populações

Método de Fisher

➡ Várias Populações (Vários Grupos)

- Proposição(Discriminantes amostrais de Fisher)

A escolha de $\hat{\mathbf{a}}_2 = \hat{\mathbf{e}}_2$ fornece a segunda discriminante amostral $\hat{Y}_2 = \hat{\mathbf{a}}_2^t \mathbf{x}$

Da mesma forma, a escolha de $\hat{\mathbf{a}}_k = \hat{\mathbf{e}}_k$ fornece a k-ésima discriminante amostral

$$\hat{Y}_k = \hat{\mathbf{a}}_k^t \mathbf{x}$$

$$k < s$$

Separação de Populações

Método de Fisher

- ➡ Várias Populações (Vários Grupos)
 - Proposição(Discriminantes amostrais de Fisher)
Será satisfeita a condição

$$\hat{\mathbf{a}}_i^t \mathbf{S}_{com} \hat{\mathbf{a}}_k = \begin{cases} 1, \text{se } i = k \leq s \\ 0, \text{senão} \end{cases}$$

Separação de Populações

Método de Fisher

➡ Classificação baseada nos discriminantes de Fisher

- Seja $Y_k = \mathbf{a}_k^t X$ o k-ésimo discriminante populacional $k \leq s$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_s \end{bmatrix} \text{ tem vetor de médias } \underset{s \times 1}{\mu_{i\mathbf{Y}}} = \begin{bmatrix} \mu_{iY_1} \\ \vdots \\ \mu_{iY_s} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^t \mu_i \\ \vdots \\ \mathbf{a}_s^t \mu_i \end{bmatrix}$$

$$p / \pi_i, i = 1, \dots, g$$

Separação de Populações

Método de Fisher

- ➡ Classificação baseada nos discriminantes de Fisher
 - e matriz de covariâncias

$$\mathbf{I}_{S \times S}$$

Separação de Populações

Método de Fisher

➡ Classificação baseada nos discriminantes de Fisher

- Distância quadrática entre $\mathbf{Y} = \mathbf{y}$ e μ_{iY}

$$(\mathbf{y} - \mu_{iY})^t \mathbf{I} (\mathbf{y} - \mu_{iY}) = \sum_{j=1}^s (y_j - \mu_{iY_j})^2$$

Separação de Populações

Método de Fisher

➡ Classificação baseada nos discriminantes de Fisher

- Regra de Classificação

Atribua $\mathbf{y}$ a população π_k se a distancia

quadrática entre $\mathbf{y}$ e μ_{kY} é menor do que as

distâncias quadráticas entre $\mathbf{y}$ e μ_{iY}

$i \neq k$

Separação de Populações

Método de Fisher

➡ Classificação baseada nos discriminantes de Fisher

- Usando-se apenas $r \leq s$ discriminantes
Alocar $\mathbf{x}$ a π_k se

$$\sum_{j=1}^r (y_j - \mu_{kY_j})^2 =$$

$$\sum_{j=1}^r [\mathbf{a}_j(\mathbf{x} - \mu_k)]^2 \leq \sum_{j=1}^r [\mathbf{a}_j(\mathbf{x} - \mu_i)]^2, i \neq k$$

Separação de Populações

Método de Fisher

➡ Classificação baseada nos discriminantes de Fisher

- Regra de classificação de Fisher baseada nos discriminantes amostrais

Alocar $\mathbf{x}$ a π_k se

$$\sum_{j=1}^r (\hat{y}_j - \hat{y}_{kj})^2 =$$

$$\sum_{j=1}^r [\hat{\mathbf{a}}_j(\mathbf{x} - \bar{\mathbf{x}}_k)]^2 \leq \sum_{j=1}^r [\hat{\mathbf{a}}_j(\mathbf{x} - \bar{\mathbf{x}}_i)]^2, i \neq k$$