ISSN 2448-3230

ANÁLISE DISCRIMINANTE: CLASSIFICADORES

Raquel Costa da Silva

Jeniffer Johana Duarte Sanchez

Resumo

A análise discriminante é uma técnica da análise multivariada utilizada para discriminar e classificar objetos, desde os já existentes até os novos apresentados pela natureza. Neste trabalho apresentamos os principais métodos da análise discriminante e uma aplicação em dados reais. Inicialmente são apresentados alguns conceitos básicos. Em seguida são expostas regras de classificação para mais de duas populações e métodos para avaliar seu desempenho. Por fim, é feita uma análise de dados reais utilizando a plataforma R, com o intuitivo de mostrar a eficiência dos classificadores apresentados.

Palavras-chave: Análise Discriminante; Regras de Classificação; Método de Fisher.

Introdução

Considere g populações, $\pi_1, \pi_2, \ldots, \pi_g, g > 2$. Suponha que cada grupo π_j esteja associado a uma função de densidade de probabilidade (fdp) $f_f(x)$ em R^g, tal que todo individuo da população π_j possui fdp $f_j(x)$. O problema da discriminação entre dois ou mais grupos consiste em obter funções matemáticas capazes de classificar um indivíduo x em uma das populações π_j, $j = 1, \ldots, g$, com base em medidas de um número p de características, buscando minimizar a probabilidade de erro, ou seja, minimizando a probabilidade de classificar um indivíduo em uma população π_i quando na verdade ele pertence a população π_j $i \neq j$ $(i, j = 1, 2, \ldots, g)$. Para tanto, deve-se determinar uma regra que possa dizer a que população é mais provável um determinado indivíduo pertencer [Anderson, 1954].

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

Uma regra discriminante d corresponde a uma divisão do R^g em regiões disjuntas $R_1, \ldots, R_g$, onde alocamos a observação x em π_j se $x \in R_j$. Segundo Johnson (2002) uma boa regra deve resultar em poucas classificações incorretas, sendo então importante olharmos para o custo de má classificação. A Figura (1) exemplifica a divisão de R^2 em regiões R_1 e R_2.

O objetivo desse trabalho consiste em apresentar os principais métodos da análise discriminante para mais de duas populações e uma aplicação em dados reais. Inicialmente apresentando alguns conceitos básicos e em seguida expomos as regras de classificação para mais de duas populações e métodos para avaliar seu desempenho. Por fim, apresentamos uma nova aplicação, com o banco de dados ***iris***.

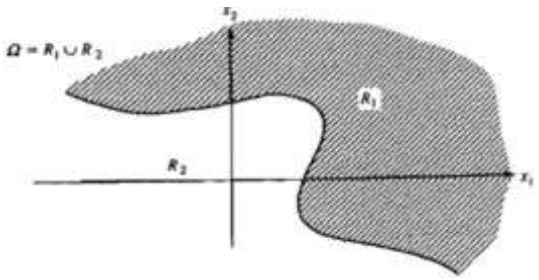

Figura 1: Divisão das regiões R_1 e R_2.

Classificação com várias populações

Seja $f_i(x)$ a densidade associada a população π_i $i = 1, \ldots, g$ e p_i a probabilidade a priori associada a π_i. Seja $c(j|i)$ o custo da aloca de um item em π_j quando na verdade pertence a π_i para $j \neq i, i = 1, \ldots, g$. Quando $j = i$, temos $c(i/i)$= 0. Finalmente, seja R_j o conjunto de $x's$ classificados como π_j, então a probabilidade de classificar um elemento em π_j quando na verdade ele pertence a π_i é dada por

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

$$P\,(j|i) = \int_{R_J} f_i(x)dx$$

para $J, i = 1, \ldots, g$ com

$$P(i|i) = 1 - \sum_{\substack{j=1 \\ j \neq 1}}^{g} P(j|i).$$

O custo esperado condicional de má classificação de uma observação x de π_1 em π_2 ou π_g é

$$\text{CEM (1)} = P\,(2|1)\,c(2|1) + P(3|1)\,c(3|1) + \ldots + P\,(g|1)\,c(g|1)$$

$$= \sum_{j=2}^{g} P\,(j|1)c\,(j|1).$$

Este custo condicional esperado ocorre com probabilidade a priori de p_1. Analogamente, podemos obter os custos esperados condicionais de má classificação, CEM (2), ..., CEM (g). Dessa forma, multiplicando cada CEM (i) por sua probabilidade a priori e somando os termos obtidos, encontramos o CEM global

$$\text{CEM} = p_1 CEM\,(1) + \ldots + p_g CEM\,(g)$$

$$= p_1\left(\sum_{j=2}^{g} P(j|1)c\,(j|1)\right) + \ldots + p_g\left(\sum_{j=1}^{g-1} P(j|g)c\,(j|g)\right)$$

$$= \sum_{i=1}^{g} p_i \left(\sum_{\substack{i=1 \\ i \neq j}}^{g} P(j|i)c\,(j|i)\right).$$

Determinar uma regra de classificação ótima, implica na escolha de regiões de classificação que minimizem tal CEM. Sendo assim, temos que as regras de classificação são definidas alocando x para a população π_j para a qual

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

$$\sum_{\substack{i=1 \\ i\neq j}}^{g} p_i f_i(x) c\,(j|i)$$

é o menor possível [Ver Anderson (1954)]. No caso de empate, x pode ser alocado em qualquer uma das populações empatadas.

Agora, suponha, sem perda de generalidade, que todos os custos de má classificação são iguais a 1. Então, a observação x será alocada a população π_j para o qual

$$\sum_{\substack{i=j \\ i\neq j}}^{g} p_i f_i(x),$$

é o menor possível. Note que tal mínimo será atingido quando $p_j f_j(x)$ for máximo. Dessa forma, podemos reescrever as regiões de classificação, no caso de custos de má classificação idênticos, como:

Aloca-se x para π_j se:

$$p_j f_j(x) > p_i f_i(x), \forall\, i \neq j.$$

Equivalentemente:

Aloca-se x a π_j se:

$$\ln p_j\, f_j(x) > \ln p_i\, f_i(x), \forall\, i \neq j$$

É interessante notar que a regra de classificação obtida nesse caso é equivalente à regra obtida pela maximização da probabilidade a posteriori de classificação vista para o caso de duas populações. No contexto de várias populações temos:

$$p(\pi_j | x) = \frac{p_j\, f_j(x)}{\sum_{I=1}^{g} p_i f_i(x)}.$$

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

Classificação com populações normais

Suponha agora que as g populações sejam normais multivariadas com vetores de média μ_i e matrizes de covariância, ou seja,

$$f_i(x) = \frac{1}{(2\pi)^{p/2} \mid \Sigma_i \mid^{1/2}} \exp\left[\frac{-1}{2} \ (x - \mu_i)^T \ \Sigma_i^{-1}(x - \mu\_i)\right].$$

Suponha ainda que os custos de má classificação são idênticos. Nessa situação, temos a regra de classificação definida como:

Aloca-se x a π_j se:

$$\ln p_j \ f_j(x) = \ln p_j - \left(\frac{p}{2}\right)\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| - \frac{1}{2}(x - \mu_j)^T \Sigma_j^{-1}(x - \mu_j)$$

$$= \max \ln p_i f_i\,(x).$$

A constante (p/2) ln (2 π) pode ser ignorada, uma vez que é a mesma para todas as populações. Definimos então o escore de discriminação quadrático para a i-ésima população, considerando custos de má classificação idênticos, como

$$d^Q(x) = -\frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(x - \mu_i)^T \ \Sigma_i^{-1}(x - \mu_i) + \ln p_i$$

Usando os escores de discriminação, a regra de classificação acima pode ser reescrita como:

Aloca-se x para π_j se:

$$d_j^Q(x) = \max\left[\ d_1^Q(x), d_2^Q(x), \ldots, d_g^Q(x)\right].$$

Na prática, μ_i e Σ_i são desconhecidos, mas pode-se utilizar amostras de treinamento para construir estimativas para os mesmos. Sendo assim, podemos construir uma regra de classificação utilizando a estimativa do escore de discriminação quadrático

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

$$\hat{d}_i^Q(x) = -\frac{1}{2}\ln|S_i| - \frac{1}{2}(x - \bar{x_i})^T S_i^{-1}(x - \bar{x_i}) + \ln p_i.$$

Tal regra de classificação é dada por:

Aloca-se x a π_j se:

$$\hat{d}_j^Q = \max\left[\hat{d}_1^Q(x), \hat{d}_2^Q(x), \ldots, \hat{d}_g^Q(x)\right].$$

Uma simplificação é possível se as matrizes de covariância populacionais, Σ_i, são idênticas. Quando $\Sigma_i = \Sigma$ para $i = 1, 2, \ldots, g$ o escore discriminante torna-se

$$d_j^Q(x) = -\frac{1}{2}\ln|\Sigma| - \frac{1}{2}\, x^T\Sigma^{-1}\, x + \mu_i^T\, \Sigma^{-1}x - \frac{1}{2}\,\mu_i\Sigma^{-1}\mu_i + \ln p_i.$$

Note que os dois primeiros termos são idênticos para todos $d_i^Q(x_i)$, e, consequentemente, eles podem ser ignoradas para efeitos de alocação. Os demais termos consistem de uma constante $c_i = \ln p_i - \frac{1}{2}\,\mu_i^T\Sigma^{-1}\,\mu_i$ e uma combinação linear das componentes de x. Dessa forma, podemos definir o escore discriminante linear por

$$d_i(x) = \mu_i^T\,\Sigma^{-1}\,x - \tfrac{1}{2}\,\mu_i^T\Sigma^{-1}\mu_i + \ln p_i.$$

Uma estimativa $\hat{d}_i\,(x)$, dos escores discriminantes lineares $d_i(x)$, baseia-se na estimativa combinada de Σ.

$$S_c = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2 + \ldots + (n_g - 1)S_g}{n_1 + n_2 + \cdots + n_g - g}$$

e é dada por

$$\hat{d}_i(x) = \bar{x}_i^T S_c^{-1}x - \frac{1}{2}\,\bar{x}_i^T\,S_c^{-1}\bar{x} + \ln p_i$$

Consequentemente, temos a seguinte regra de classificação:

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

Aloca-se x para π_j se:

$$\hat{d}_j(x) = \max \left[\hat{d}_1(x),\ \hat{d}_2(x), \ldots,\ \hat{d}_g(x)\right].$$

Método de Fisher para várias populações

Fisher também propôs uma extensão de seu método discriminante para o caso de várias populações. O principal objeto do método é separar populações, mas também pode ser utilizado para classificação. Assim como no caso de duas populações, não é necessário supor a normalidade das mesmas, no entanto há a suposição de que as g matrizes de covariâncias são idênticas [ver Fisher, 1936].

Seja $\bar{\mu}$ o vetor de médias das g populações combinadas e B_μ a matriz de somas dos produtos cruzados

$$B_\mu = \sum_{i=1}^{g} (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^T,$$

onde $\bar{\mu} = \frac{1}{g} \sum_{i=1}^{g} \mu_i$. Consideramos a combinação linear

$$Y = l^T X,$$

que possui valor esperado, para a população π_i , igual a

$$\mathrm{E(Y)} = l^T E(X \mid \pi_i) = l^T \mu_i$$

e variância

$$\mathrm{Var\ (Y)} = l^T Cov\,(X) l = l^T \Sigma\, l\,,$$

Para todo as populações. Consequentemente. O valor esperado $\mu_{iY} = l^T \mu_i$ muda de acordo com a população da qual X provém. Defina

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

$$\bar{\mu}_Y = \frac{1}{g}\sum_{i=1}^{g}\mu_{iY} = \frac{1}{g}\sum_{i=1}^{g} l^T\mu_i = l^T\left(\frac{1}{g}\sum_{i=1}^{g}\mu_i\right) = l^T\bar{\mu},$$

E forme a razão

$$\frac{\sum_{i=1}^{g}(\mu_{iY}-\bar{\mu}_Y)^2}{\sigma_Y^2} = \frac{\sum_{i=1}^{g}\left(l^T\mu_i - l^T\bar{\mu}\right)^2}{l^T\Sigma\, l}$$

$$= \frac{l^T\left(\sum^T(\mu_i-\bar{\mu})(\mu_i-\bar{\mu})^T\right)l}{l^T\Sigma\, l},$$

ou

$$\frac{\sum_{i=1}^{g}(\mu_{iY}-\bar{\mu}_Y)^2}{\sigma_y^2} = \frac{l^T B_\mu\, l}{l^T\Sigma\, l}.$$

A razão anterior mede a variabilidade entre os grupos relativamente à variabilidade comum dentro dos grupos. Podemos escolher o vetor l de modo a maximizar essa razão. Comumente, a matriz de covariância e os vetores de média da população são desconhecidos, mas dispõe-se de uma amostra de treinamento a partir da qual podemos estimar os parâmetros desconhecidos.

Seja X_i uma matriz $p \times n_i$ de dados da população π_i e x_{ij}^T sua j-ésima coluna. Após obter os vetores de médias amostrais

$$\bar{x_i} = \frac{1}{n_i}\sum_{i=1}^{n_i} x_{ij}$$

e as matrizes de covariância S_i, $i = 1,2,\ldots,g,$ definimos o vetor de médias global

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

$$\bar{x} = \frac{\sum_{i=1}^{g} n_i \bar{x_i}}{\sum_{i=1}^{g} n_i} = \frac{\sum_{i=1}^{g} \sum_{j=1}^{n_i} x_{ij}}{\sum_{i=1}^{g} n_i}.$$

Em seguida, analogamente a B_μ , definimos a matriz

$$\hat{B}_\mu = \sum_{i=1}^{g} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T,$$

e a matriz

$$\mathrm{W} = \sum_{i=1}^{g} (n_i - 1) S_i = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x_i})(x_{ij} - \bar{x_i})^T.$$

Consequentemente, W / $(n_1 + n_2 + \cdots + n_g - g) = S_c$ é o estimador de Σ. Note que W é a constante $(n_1 + n_2 + \cdots + n_g - g)$ vezes S_c , então o mesmo $\hat{l}$ que maximiza $\hat{l}^T \hat{B}_\mu \hat{l} / \hat{l}^T S_c \hat{l}$ também maximiza $\hat{l}^T \hat{B}_\mu \hat{l} / \hat{l}^T W \hat{l}$. Além disso, podemos representar $\hat{l}$ a partir dos autovetores $\hat{e_i}$ de $W^{-1} B_\mu$, já que se $W^{-1} \hat{B}_\mu \hat{e} = \hat{\lambda}\, \hat{e}$ então $S_c^{-1} \hat{B}_\mu \hat{e} = \hat{\lambda} (n_1 + n_2 + \cdots + n_g - g)\, \hat{e}$.

Discriminante linear amostral de Fisher

Sejam $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_s$ os s ≤ min $(g - 1, p)$ autovalores não nulos de $W^{-1} \widehat{B_\mu}$ e $\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_s$ seus correspondentes autovetores normalizados tal que $\hat{e}^T S_c \hat{e} = 1$. Então o vetor de coeficientes $\hat{l}$ que maximiza a razão

$$\frac{\hat{l}^T \widehat{B_\mu} \hat{l}}{l^T W \hat{l}} = \frac{\hat{l} \left[\sum_{i-1}^{g} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T\right] \hat{l}}{\hat{l} \left[\sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x_i})(x_{ij} - \bar{x_i})^T\right] \hat{l}},$$

é dado por $\hat{l}_1 = \hat{e_1}$. A contradição linear $\hat{l}_1^T x$ é chamada de primeiro discriminante amostral. A escolha $\hat{l}_2 = \hat{e}_2$ produz o segundo discriminante amostral, $\hat{l}_2^T x$ e assim em diante, com $\hat{l}_k^T x = \hat{e}_k^T x,\ k \leq s,$ sendo o k-ésimo discriminante amostral.

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

Exemplo 1: Consideremos os observações em duas variáveis $(\mathrm{p}=2)$ de três populações

$(\mathrm{g}=3)$. Vamos encontrar os discriminantes de Fisher.

$$\pi_1: X_1 = \begin{bmatrix} -2 & 5 \\ 0 & 3 \\ -1 & 1 \end{bmatrix}, \quad \text{assim} \quad n_1 = 3, \quad \bar{x}_1 = \begin{bmatrix} -1 \\ 3 \end{bmatrix} \quad \text{e} \quad S_1 = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

$$\pi_2: X_2 = \begin{bmatrix} 0 & 6 \\ 2 & 4 \\ 1 & 2 \end{bmatrix}, \quad \text{assim} \quad n_2 = 3, \quad \bar{x}_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \quad \text{e} \quad S_2 = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

$$\pi_3: X_3 = \begin{bmatrix} 1 & -2 \\ 0 & 0 \\ -1 & 4 \end{bmatrix}, \quad \text{assim} \quad n_3 = 3, \quad \bar{x}_3 = \begin{bmatrix} 0 \\ -2 \end{bmatrix} \quad \text{e} \quad S_3 = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$$

Sendo assim, a matriz de covariância combinada é dada por

$$S_c = \frac{3-1}{9-3}\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} + \frac{3-1}{9-3}\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} + \frac{3-1}{9-3}\begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{3} \\ -\frac{1}{3} & 4 \end{bmatrix}.$$

Além disso,

$$\bar{x} = \begin{bmatrix} 0 \\ \frac{5}{3} \end{bmatrix} \quad \text{e} \quad \hat{B}_\mu = \sum_{i=1}^{3} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T = \begin{bmatrix} 2 & 1 \\ 1 & \frac{62}{3} \end{bmatrix}.$$

Por sua vez,

$$W = \sum_{i=1}^{3}\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T = (n_1 + n_2 + n_3 - 3)S_c = \begin{bmatrix} 6 & -2 \\ -2 & 24 \end{bmatrix},$$

e, consequentemente,

$$W^{-1} = \frac{1}{140}\begin{bmatrix} 24 & 2 \\ 2 & 6 \end{bmatrix} \quad \text{e} \quad W^{-1}\hat{B}_\mu = \begin{bmatrix} 0.3571 & 0.4667 \\ 0.0714 & 0.9000 \end{bmatrix}.$$

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

Para determinar os $s \leq \min(g-1, p) = \min(2,2) = 2$ autovalores de $W^{-1}\hat{B}_\mu$ precisamos resolver

$$|W^{-1}\hat{B}_\mu - \lambda I| = \left|\begin{bmatrix} 0.3571-\lambda & 0.4667 \\ 0.0714 & 0.9000-\lambda \end{bmatrix}\right| = 0$$ ou

$(0.3571 - \lambda)(0.9000 - \lambda) - (0.4667)(0.0714) = \lambda^2 - 1.2571\,\lambda + 0.2881 = 0$

Usando a fórmula quadrática, encontramos $\hat{\lambda}_1 = 0.9556$ *e* $\hat{\lambda}_2 = 0.3015$. Os auto vetores normalizados $\hat{l}_1$ e $\hat{l}_2$ são obtidos resolvendo

$$(W^{-1}\hat{B}_\mu - \lambda_i I)\hat{l}_i = 0,\ i = 1,2$$

e normalizando os resultados de tal forma que $\hat{l}_i^T S_c \hat{l}_i = 1$. Por exemplo, a solução de

$$(W^{-1}\hat{B}_0 - \hat{\lambda}_i I)\hat{l}_1 = \begin{bmatrix} 0.3571-0.9556 & 0.4667 \\ 0.714 & 0.9000-0.9556 \end{bmatrix} = \begin{bmatrix} \hat{l}_{11} \\ \hat{l}_{12} \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

e, após a normalização, $\hat{l}_i^T S_c \hat{l}_i = 1$,

$$\hat{l}_1^T = [0.386 \quad 0.495].$$

De forma similar, se obtém

$$\hat{l}_2^T = [0.938 \quad -0.112].$$

Logo, os dois discriminantes são

$$\hat{y}_1 = \hat{l}_1^T x = [0.386\ \ 0.495]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0.386\,x_1 + 0.495\,x_2$$

$$\hat{y}_2 = \hat{l}_2^T x = [0.938\ \ -0.112]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0.938\,x_1 - 0.112\,x_2.$$

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

Seja $Y_k = \hat{l}_k^T X$ o k-ésimo discriminante, k $\leq$ s. Concluímos que

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_3 \end{bmatrix}$$

Tem vetor média

$$\mu_{iY} = \begin{bmatrix} \mu_{iY_1} \\ \vdots \\ \mu_{iY_s} \end{bmatrix} = \begin{bmatrix} l_1^T \mu_i \\ \vdots \\ l_s^T \mu_i \end{bmatrix}$$

sob a população π_i e matriz de covariância I para todas a populações.

Como as componentes de Y têm variâncias unitárias e covariâncias nulas, a medida adequada de distância quadrada de $Y = y$ para μ_{iY} é:

$$(Y - \mu_{iY})^T (Y - \mu_{iY}) \Sigma_{j=1}^{s} \left(y_j - \mu_{iY_j}\right)^2.$$

Uma regra de classificação razoável é aquela que atribui y para população π_j se a distância quadrada de y para μ_{jy} for menor do que a distância quadrada de y para μ_{iy} para $i \neq j$.

Se apenas r dos discriminantes são usados para alocação, a regra é:

Aloca-se x em π_j se

$$\sum_{k=1}^{r} (y_k - \mu_{jY_k})^2 = \sum_{k=1}^{r} [l_k^T (x - \mu_j)]^2 \leq \sum_{k=1}^{r} [l_k^T (x - \mu_i)^2] \ , \forall\, i \neq j.$$

Podemos ainda utilizar a regra de classificação baseada nos discriminantes amostrais. Tal regra é dada por:

Aloca-se x em π_j se

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

$$\sum_{k=1}^{r}(\hat{y}_k - \bar{y}_{jk})^2 = \sum_{k=1}^{r}[\hat{l}_k^T(x - \bar{x}_j)]^2 \leq \sum_{k=1}^{r}\left[\hat{l}_k^T(x - \bar{x}_j)^2\right], \forall\, i \neq j,$$

em que $\bar{y}_{jk} = \hat{l}_k\bar{x}_j$ e r ≤ s. Quando as probabilidades a priori são $p_1 = p_2 = \cdots = p_g = \frac{1}{g}$ e

$r = s$, a regra acima é equivalente a baseada no maior escore discriminante linear.

Metodologia

Realizou-se um estudo na plataforma R com o objetivo de analisar o desempenho das regras de classificação para mais de duas populações. Forarm usadas 150 observações de flores do banco de dados iris. Iris é um género de plantas com flor, muito apreciado pelas suas diversas espécies, que ostentam flores de cores muito vivas e são flores muito frequentes em jardins.

Os dados contêm quatro caractériscas de três espécies de flores Iris, da familia das Iridáceas: comprimento da sépala, largura da sépala, comprimento da pétala e largura da pétala. Os classificadores utilizam esses 4 atributos para construir funções discriminantes e depois constuimos a matriz confusão que deve classificar corretamente os 3 tipos de flores.

Resultados e discussões

Para a análise feita nessa seção foi utilizada a versão 3.0.2 da plataforma R e banco de dados ***iris***. Os dados da flor Iris é um conjunto de dados multivariados introduzido por Sir Ronald Aylmer Fisher (1936) como um exemplo de análise discriminante. O conjunto de dados consiste de 50 amostras de cada uma das três espécies de flores Iris (setosa, virginica e versicolor). Quatro características foram medidas em cada amostra: o comprimento da sépala (X_1), largura da sépala (X_2), comprimento da pétala (X_3), e largura da pétala (X_4), A Figura (2) mostra as três espécies de Iris presentes nos dados.

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

Figura 2: Espécies de flores Iris.

Para a análise foram selecionadas aleatoriamente 50 observações para a amostra de treinamento, as quais forneceram os seguintes discriminantes:

$$Y_1 = 0.8727474X_1 + 2.1175976X_2 - 2.5315621X_3 - 2.1376389X_4.$$

$$Y_2 = -1.622659X_1 - 1.011072X_2 + 1.938316X_3 - 3.466623X_4.$$

Em seguida, as 100 observações restantes foram utilizadas para construir a matriz de confusão: As porcentagens de acertos de classificação foram de 100% para a setosa, 96.875% para a versicolor e 97.14286% para a virginica. A porcentagem de acerto global foi de 98%.

		População prevista		
		Setosa	Versicolor	Virginica
População verdadeira	Setosa	33	0	0
	Versicolor	0	31	1
	Virginica	0	1	34

A Figura (3) mostra o gráfico de dispersão dos discriminantes para cada observação.

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co

ISSN 2448-3230

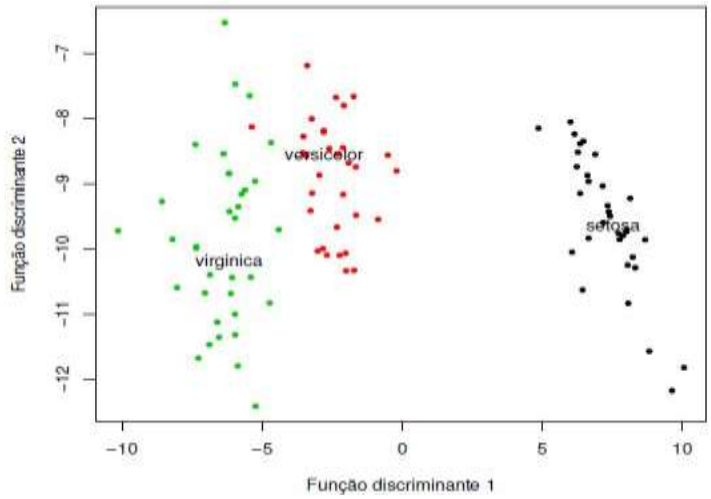

Figura 3: Gráfico de dispersão.

Considerações Finais

Os classificadores propostos se mostram eficiêntes ao classificar os dados da base de dados iris, os erros verificados foram relativamente pequenos. Através da análise dos resultados também é possivel verificar que as classes de versicolor e virginica tem maior semelhança entre si do que a outra classe de setosa que apresentou 100% de acertos.

Referências

[1] Anderson, T.W. *An Introduction to multivariate statistical analysis*. 3rd. ed 1954.

[2] Fisher, R.A. *The use of multiple measurements in taxonomic problems*. Annals of Eugenics 7 (1936): 179–188.

[3] Johnson, R.A., and Wichern, D.W. *Applied multivariate statistical analysis*. Vol. 5. No. 8. Upper Saddle River, NJ: Prentice hall 2002.

IF Sertão Pernambucano , raquelmendiola@yahoo.com.br
USP, jjduartes@unal.edu.co