

Análise Discriminante

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.
- ▶ Em um problema de **discriminação**, o objetivo é aplicar e avaliar a contribuição de p variáveis na discriminação (separação) de k grupos **pré-definidos**;

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.
- ▶ Em um problema de **discriminação**, o objetivo é aplicar e avaliar a contribuição de p variáveis na discriminação (separação) de k grupos **pré-definidos**;
- ▶ Em um problema de **classificação**, o objetivo é construir uma regra, baseada em p variáveis e n indivíduos, que permita classificar novos indivíduos em um de k grupos **pré-definidos**.

Discriminação em dois grupos

- ▶ Considere que desejamos discriminar indivíduos de duas populações π_1 e π_2 com base em p variáveis avaliadas $x = [X_1, X_2, \dots, X_p]^t$.

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.
- ▶ Em um problema de **discriminação**, o objetivo é aplicar e avaliar a contribuição de p variáveis na discriminação (separação) de k grupos **pré-definidos**;
- ▶ Em um problema de **classificação**, o objetivo é construir uma regra, baseada em p variáveis e n indivíduos, que permita classificar novos indivíduos em um de k grupos **pré-definidos**.

Discriminação em dois grupos

- ▶ Considere que desejamos discriminar indivíduos de duas populações π_1 e π_2 com base em p variáveis avaliadas $x = [X_1, X_2, \dots, X_p]^t$.
- ▶ Considere ainda conhecidas μ_1, μ_2, Σ_1 e Σ_2 .

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.
- ▶ Em um problema de **discriminação**, o objetivo é aplicar e avaliar a contribuição de p variáveis na discriminação (separação) de k grupos **pré-definidos**;
- ▶ Em um problema de **classificação**, o objetivo é construir uma regra, baseada em p variáveis e n indivíduos, que permita classificar novos indivíduos em um de k grupos **pré-definidos**.

Discriminação em dois grupos

- ▶ Considere que desejamos discriminar indivíduos de duas populações π_1 e π_2 com base em p variáveis avaliadas $x = [X_1, X_2, \dots, X_p]^t$.
- ▶ Considere ainda conhecidas μ_1, μ_2, Σ_1 e Σ_2 .
- ▶ Sejam $f_1(x)$ e $f_2(x)$ as funções (densidade) de probabilidade de x em π_1 e π_2 .

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.
- ▶ Em um problema de **discriminação**, o objetivo é aplicar e avaliar a contribuição de p variáveis na discriminação (separação) de k grupos **pré-definidos**;
- ▶ Em um problema de **classificação**, o objetivo é construir uma regra, baseada em p variáveis e n indivíduos, que permita classificar novos indivíduos em um de k grupos **pré-definidos**.

Discriminação em dois grupos

- ▶ Considere que desejamos discriminar indivíduos de duas populações π_1 e π_2 com base em p variáveis avaliadas $x = [X_1, X_2, \dots, X_p]^t$.
- ▶ Considere ainda conhecidas μ_1, μ_2, Σ_1 e Σ_2 .
- ▶ Sejam $f_1(x)$ e $f_2(x)$ as funções (densidade) de probabilidade de x em π_1 e π_2 .
- ▶ Seja Ω o espaço amostral (conjunto de resultados possíveis para x).

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.
- ▶ Em um problema de **discriminação**, o objetivo é aplicar e avaliar a contribuição de p variáveis na discriminação (separação) de k grupos **pré-definidos**;
- ▶ Em um problema de **classificação**, o objetivo é construir uma regra, baseada em p variáveis e n indivíduos, que permita classificar novos indivíduos em um de k grupos **pré-definidos**.

Discriminação em dois grupos

- ▶ Considere que desejamos discriminar indivíduos de duas populações π_1 e π_2 com base em p variáveis avaliadas $x = [X_1, X_2, \dots, X_p]^t$.
- ▶ Considere ainda conhecidas μ_1, μ_2, Σ_1 e Σ_2 .
- ▶ Sejam $f_1(x)$ e $f_2(x)$ as funções (densidade) de probabilidade de x em π_1 e π_2 .
- ▶ Seja Ω o espaço amostral (conjunto de resultados possíveis para x).

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.
- ▶ Em um problema de **discriminação**, o objetivo é aplicar e avaliar a contribuição de p variáveis na discriminação (separação) de k grupos **pré-definidos**;
- ▶ Em um problema de **classificação**, o objetivo é construir uma regra, baseada em p variáveis e n indivíduos, que permita classificar novos indivíduos em um de k grupos **pré-definidos**.

Discriminação em dois grupos

- ▶ Considere que desejamos discriminar indivíduos de duas populações π_1 e π_2 com base em p variáveis avaliadas $x = [X_1, X_2, \dots, X_p]^t$.
- ▶ Considere ainda conhecidas μ_1, μ_2, Σ_1 e Σ_2 .
- ▶ Sejam $f_1(x)$ e $f_2(x)$ as funções (densidade) de probabilidade de x em π_1 e π_2 .
- ▶ Seja Ω o espaço amostral (conjunto de resultados possíveis para x).

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.
- ▶ Em um problema de **discriminação**, o objetivo é aplicar e avaliar a contribuição de p variáveis na discriminação (separação) de k grupos **pré-definidos**;
- ▶ Em um problema de **classificação**, o objetivo é construir uma regra, baseada em p variáveis e n indivíduos, que permita classificar novos indivíduos em um de k grupos **pré-definidos**.

Discriminação em dois grupos

- ▶ Considere que desejamos discriminar indivíduos de duas populações π_1 e π_2 com base em p variáveis avaliadas $x = [X_1, X_2, \dots, X_p]^t$.
- ▶ Considere ainda conhecidas μ_1, μ_2, Σ_1 e Σ_2 .
- ▶ Sejam $f_1(x)$ e $f_2(x)$ as funções (densidade) de probabilidade de x em π_1 e π_2 .
- ▶ Seja Ω o espaço amostral (conjunto de resultados possíveis para x).

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.
- ▶ Em um problema de **discriminação**, o objetivo é aplicar e avaliar a contribuição de p variáveis na discriminação (separação) de k grupos **pré-definidos**;
- ▶ Em um problema de **classificação**, o objetivo é construir uma regra, baseada em p variáveis e n indivíduos, que permita classificar novos indivíduos em um de k grupos **pré-definidos**.

Discriminação em dois grupos

- ▶ Considere que desejamos discriminar indivíduos de duas populações π_1 e π_2 com base em p variáveis avaliadas $x = [X_1, X_2, \dots, X_p]^t$.
- ▶ Considere ainda conhecidas μ_1, μ_2, Σ_1 e Σ_2 .
- ▶ Sejam $f_1(x)$ e $f_2(x)$ as funções (densidade) de probabilidade de x em π_1 e π_2 .
- ▶ Seja Ω o espaço amostral (conjunto de resultados possíveis para x).

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.
- ▶ Em um problema de **discriminação**, o objetivo é aplicar e avaliar a contribuição de p variáveis na discriminação (separação) de k grupos **pré-definidos**;
- ▶ Em um problema de **classificação**, o objetivo é construir uma regra, baseada em p variáveis e n indivíduos, que permita classificar novos indivíduos em um de k grupos **pré-definidos**.

Discriminação em dois grupos

- ▶ Considere que desejamos discriminar indivíduos de duas populações π_1 e π_2 com base em p variáveis avaliadas $x = [X_1, X_2, \dots, X_p]^t$.
- ▶ Considere ainda conhecidas μ_1, μ_2, Σ_1 e Σ_2 .
- ▶ Sejam $f_1(x)$ e $f_2(x)$ as funções (densidade) de probabilidade de x em π_1 e π_2 .
- ▶ Seja Ω o espaço amostral (conjunto de resultados possíveis para x).

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.
- ▶ Em um problema de **discriminação**, o objetivo é aplicar e avaliar a contribuição de p variáveis na discriminação (separação) de k grupos **pré-definidos**;
- ▶ Em um problema de **classificação**, o objetivo é construir uma regra, baseada em p variáveis e n indivíduos, que permita classificar novos indivíduos em um de k grupos **pré-definidos**.

Discriminação em dois grupos

- ▶ Considere que desejamos discriminar indivíduos de duas populações π_1 e π_2 com base em p variáveis avaliadas $x = [X_1, X_2, \dots, X_p]^t$.
- ▶ Considere ainda conhecidas μ_1, μ_2, Σ_1 e Σ_2 .
- ▶ Sejam $f_1(x)$ e $f_2(x)$ as funções (densidade) de probabilidade de x em π_1 e π_2 .
- ▶ Seja Ω o espaço amostral (conjunto de resultados possíveis para x).

O Problema da classificação e discriminação

- ▶ Dois tipos de problemas recorrentes em análise de dados, e que podem ser abordados por técnicas multivariadas, são **discriminação** e **classificação**.
- ▶ Em um problema de **discriminação**, o objetivo é aplicar e avaliar a contribuição de p variáveis na discriminação (separação) de k grupos **pré-definidos**;
- ▶ Em um problema de **classificação**, o objetivo é construir uma regra, baseada em p variáveis e n indivíduos, que permita classificar novos indivíduos em um de k grupos **pré-definidos**.

Discriminação em dois grupos

- ▶ Considere que desejamos discriminar indivíduos de duas populações π_1 e π_2 com base em p variáveis avaliadas $x = [X_1, X_2, \dots, X_p]^t$.
- ▶ Considere ainda conhecidas μ_1, μ_2, Σ_1 e Σ_2 .
- ▶ Sejam $f_1(x)$ e $f_2(x)$ as funções (densidade) de probabilidade de x em π_1 e π_2 .
- ▶ Seja Ω o espaço amostral (conjunto de resultados possíveis para x).

Classificação em mais de duas população normais

- ▶ Utilizando o escore quadrático $d_i^Q(x)$ de discriminação, podemos simplificar a regra de classificação. Classificamos x em π_i se

$$d_i^Q(x) = \max_j [d_j^Q(x)]$$

para $j = 1, 2, \dots, k$.

Classificação em mais de duas população normais

Para dados amostrais

- ▶ Podemos obter uma regra estimada substituindo os parâmetros μ_i e Σ_i , pelos respectivos estimadores \bar{x}_i e S_i , $i = 1, 2, \dots, k$.
- ▶ O estimador da função quadrática $d_i^Q(x)$ é representado por $Q_i(x)$ e é dado por

$$Q_i(x) = -\frac{1}{2} \ln(|S_i|) - \frac{1}{2} (x - \bar{x}_i)^t S_i^{-1} (x - \bar{x}_i) + \ln(p_i)$$

para $i = 1, 2, \dots, k$ e, pela regra estimada de mínima probabilidade total de

Classificação em mais de duas população normais

- ▶ Utilizando o escore quadrático $d_i^Q(x)$ de discriminação, podemos simplificar a regra de classificação. Classificamos x em π_i se

$$d_i^Q(x) = \max_j [d_j^Q(x)]$$

para $j = 1, 2, \dots, k$.

Classificação em mais de duas população normais

Para dados amostrais

- ▶ Podemos obter uma regra estimada substituindo os parâmetros μ_i e Σ_i , pelos respectivos estimadores \bar{x}_i e S_i , $i = 1, 2, \dots, k$.
- ▶ O estimador da função quadrática $d_i^Q(x)$ é representado por $Q_i(x)$ e é dado por

$$Q_i(x) = -\frac{1}{2} \ln(|S_i|) - \frac{1}{2} (x - \bar{x}_i)^t S_i^{-1} (x - \bar{x}_i) + \ln(p_i)$$

para $i = 1, 2, \dots, k$ e, pela regra estimada de mínima probabilidade total de

Classificação em mais de duas população normais

- ▶ Utilizando o escore quadrático $d_i^Q(x)$ de discriminação, podemos simplificar a regra de classificação. Classificamos x em π_i se

$$d_i^Q(x) = \max_j [d_j^Q(x)]$$

para $j = 1, 2, \dots, k$.

Classificação em mais de duas população normais

Para dados amostrais

- ▶ Podemos obter uma regra estimada substituindo os parâmetros μ_i e Σ_i , pelos respectivos estimadores \bar{x}_i e S_i , $i = 1, 2, \dots, k$.
- ▶ O estimador da função quadrática $d_i^Q(x)$ é representado por $Q_i(x)$ e é dado por

$$Q_i(x) = -\frac{1}{2} \ln(|S_i|) - \frac{1}{2} (x - \bar{x}_i)^t S_i^{-1} (x - \bar{x}_i) + \ln(p_i)$$

para $i = 1, 2, \dots, k$ e, pela regra estimada de mínima probabilidade total de