Gabarito - Segunda Prova
Fundamentos Estatísticos para Ciência dos Dados

12/05/2016

1. Esboce aproximadamente o vetor-direção do primeiro componente principal no gráfico a esquerda na Figura 1 ignorando os dois tipos de símbolo para os pontos. No gráfico da direita, esboce aproximadamente o vetor direção do (primeiro) discriminante linear de Fisher considerando as duas classes para os pontos.

   Figura da prova reproduzida abaixo com a solução.

   Solução: Solução na figura 1.

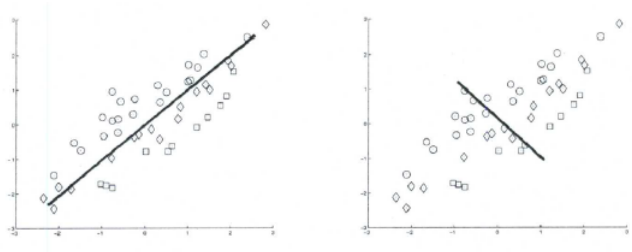

Figura 1: 1o. componente principal (esquerda) e 1o. discriminante linear (direita).

---

2. Todo objeto pertence a uma de duas classes ou populações, π_1 ou π_2. Suponha que a classe π_1 é extremamente rara (isto é, $\mathbb{P}(\in \pi_1) \approx 0$).

   - Isto significa que se classificarmos todo novo objeto na população 2 teremos um pequeno número de erros, supondo que os custos dos dois erros sejam iguais. Mostre que, com esta regra de classificação radical temos $\mathbb{P}(\text{erro}) \approx 0$.
   - Podemos tentar colocar as duas classes em pé de igualdade considerando a seguinte quantidade: $\eta = \mathbb{P}(\text{class } 2 | \in 1) + \mathbb{P}(\text{class } 1 | \in 2)$. Mostre que $\eta = 1$ com a regra de classificação radical que portanto não é uma boa regra se quisermos um η pequeno.
   - Mostre que a regra que minimiza η é aquela em que a região R de classificação é dada por $R = \{\mathbf{x} \text{ tais que } f_2(\mathbf{x}) < f_1(\mathbf{x})\}$.

   Solução: Seja $\mathbb{P}(\in \pi_1) = \varepsilon \approx 0$. Então, com a regra de classificação radical temos:

$$
\begin{aligned}
\mathbb{P}(\text{erro}) &= \mathbb{P}(\text{class } 2 \text{ e } \in 1) + \mathbb{P}(\text{class } 1 \text{ e } \in 2) \\
&= \mathbb{P}(\text{class } 2 \mid \in 1)\mathbb{P}(\in 1) + \mathbb{P}(\text{class } 1 \mid \in 2)\mathbb{P}(\in 1) \\
&= \mathbb{P}(\text{class } 2 \mid \in 1)\varepsilon + \mathbb{P}(\text{class } 1 \mid \in 2)(1-\varepsilon) \\
&= 1 \times \varepsilon + 0 \times (1-\varepsilon) \\
&= \varepsilon \approx 0
\end{aligned}
$$

   Para mostrar que $\eta = 1$ com a regra de classificação radical:

$$
\eta = \mathbb{P}(\text{class } 2 \mid \ \in 1) + \mathbb{P}(\text{class } 1 \mid \ \in 2) = 1 + 0 = 1
$$

Para minimizar η, suponha que tenhamos definido uma regra e que portanto temos o espaço das variáveis dividid em duas regiões disjuntas, R e $\bar{R}$, onde R é a região de clasificação na classe 1. Então:

$$
\begin{aligned}
\eta &= \mathbb{P}(\text{class } 2 | \in 1) + \mathbb{P}(\text{class } 1 | \in 2) \\
&= \int_{\bar{R}} f_1(\mathbf{x})\mathbf{x} + \int_R f_2(\mathbf{x})\mathbf{x} \qquad (1)
\end{aligned}
$$

Como as densidades f_1 e f_2 integram 1 no espaço $\Omega = R \cup \bar{R}$, sabemos que

$$1 = \int_\Omega f_1(\mathbf{x})\mathbf{x} = \int_R f_1(\mathbf{x})\mathbf{x} + \int_{\bar{R}} f_1(\mathbf{x})\mathbf{x}\,.$$

Substituindo em (1), temos

$$
\begin{aligned}
\eta &= \left(1 - \int_R f_1(\mathbf{x})\mathbf{x}\right) + \int_R f_2(\mathbf{x})\mathbf{x} \\
&= 1 + \int_R \left(f_2(\mathbf{x}) - f_1(\mathbf{x})\right)\mathbf{x}
\end{aligned}
$$

Para minimizar η devemos tornar a integral o mais negativa possível. Isto é obtido tomando

$$R = \{\mathbf{x} \text{ tais que } f_2(\mathbf{x}) < f_1(\mathbf{x})\}$$

Se você ainda não está completamente convencido de que esta integral pode ser negativa, veja que

$$0 = 1 - 1 = \int_\Omega f_1(\mathbf{x})\mathbf{x} - \int_\Omega f_2(\mathbf{x})\mathbf{x} = \int_\Omega \left(f_1(\mathbf{x}) - f_2(\mathbf{x})\right)\mathbf{x}$$

Como $f_1 \neq f_2$ e são ambas $geq0$ integrando 1, não podemos ter $f_1(\mathbf{x}) \geq f_2(\mathbf{x})$ em todo ponto $\mathbf{x}$ de Ω. Seja

$$R = \{\mathbf{x} \text{ tais que } f_2(\mathbf{x}) < f_1(\mathbf{x})\}$$

Então podemos decompor

$$0 = \int_\Omega \left(f_1(\mathbf{x}) - f_2(\mathbf{x})\right)\mathbf{x} = \int_{\bar{R}} \left(f_2(\mathbf{x}) - f_1(\mathbf{x})\right)\mathbf{x} + \int_R \left(f_2(\mathbf{x}) - f_1(\mathbf{x})\right)\mathbf{x}$$

O primeiro termo é positivo e o segundo termo é negativo. Além disso, R é a maior região em que $f_2(\mathbf{x}) - f_1(\mathbf{x}) < 0$. Qualquer outro ponto fora de R contribui um valor positivo para a integral de interesse.

---

3. Considere um vetor aleatório bi-dimensional $\mathbf{X} = (X_1, X_2)' \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Marque V ou F justificando sua resposta:

   - Se X_1 e X_2 são independentes, a nuvem dos n pontos $(x_{i1}, x_{i2}), i = 1, \ldots, n$ com as instâncias observadas e a curva de nível $f(\mathbf{x}) = c$ com a densidade tem necessariamente a forma de um círculo.
   - Se X_1 e X_2 são independentes, a nuvem de pontos $(x_{i1}, x_{i2}), i = 1,\ ldots, n$ com as instâncias observadas e a curva de nível $f(\mathbf{x}) = c$ com a densidade tem de ter a forma elíptica com os eixos principais da elipse alinhados com os eixos das abscissas e ordenadas (SE NECESSÁRIO, considere um círculo como uma forma especial de uma elipse)
   - A matriz de covariância $\boldsymbol{\Sigma}$ é simétrica apenas se X_1 e X_2 são independentes.
   - A matriz de covariância $\boldsymbol{\Sigma}$ é diagonal se X_1 e X_2 são independentes.
   - A matriz de covariância $\boldsymbol{\Sigma}$ é diagonal se X_1 e X_2 possuem correlação positiva.

   Solução:

- F: As curvas de nível tem a forma de uma elipse com os seus dois eixos proporcionais aos autovalores da matriz $\mathbf{\Sigma}$. Se os dois autovalores form iguais, a elipse vai se tornar um círculo. As variáveis não precisam ser independentes.
- V: Se X_1 e X_2 são independentes, $\mathbf{\Sigma}$ é uma matriz diagonal e portanto os dois autovetores são $\mathbf{e}_1 = (1, 0)$ e $\mathbf{e}_2 = (0, 1)$. Portanto, os eixo da elipse são paralelos aos eixos das coordenadas.
- F: $\mathbf{\Sigma}$ é sempre simétrica pois

$$\mathbf{\Sigma}_{12} = \text{Cov}(X_1, X_2) = \sigma_1\sigma_2\rho_{12} = \text{Cov}(X_2, X_1) = \mathbf{\Sigma}_{21}$$

  onde $\rho_{12} = \text{Corr}(X_1, X_2)$.
- V: Se X_1 e X_2 são independentes então $0 = \text{Cov}(X_1, X_2) = \mathbf{\Sigma}_{12}$.
- F: $\mathbf{\Sigma}$ é diagonal se, e somente se, $\text{Corr}(X_1, X_2) = 0$.

---

4. A Figura 2 mostra dados bi-dimensionais extraídos de uma distribuição normal multivariada.

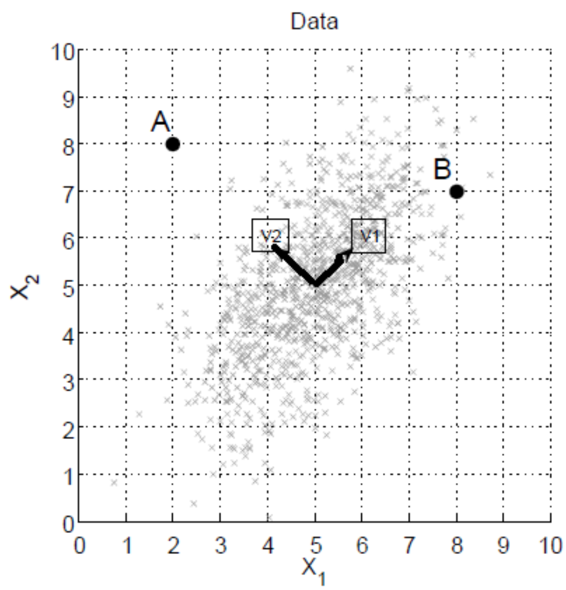

Figura 2: Dados bi-dimensionais extraídos de uma distribuição normal multivariada.

- Qual é o valor esperado de cada variável? Estime a resposta visualmente e arredonde para o número inteiro mais próximo.
- O valor do elemento $(1, 2)$ da matriz de covariância $\mathbf{\Sigma}$ é positivo, negativo ou zero?
- Defina $\mathbf{e}_1$ e $\mathbf{e}_2$ como as direções do primeiro e segundo componentes principais, ambos com comprimento igual a 1. Estas direções definem uma nova base do $\mathbb{R}^2$ onde cada ponto $\mathbf{x}$ é tranformado para $\mathbf{z} = (z_1, z_2)$ com

$$z_1 = (\mathbf{x} - \mu)'\mathbf{e}_1$$

  e

$$z_2 = (\mathbf{x} - \mu)'\mathbf{e}_2$$

  Esboce E ROTULE $\mathbf{e}_1$ e $\mathbf{e}_2$ na figura 2. Os vetores devempartir do centro da distribuição-nuvem de pontos.
- A covariância $Cov(Z_1, Z_2)$ é negativa, positiva ou aproximadamente zero?
- Como A e B estão mais ou menos a uma mesma distância do centro eles possuem igual chance de serem observados. V ou F? Justifique.

- A probabilidade de observar um ponto tão afastado de μ quanto B é aproximadamente 5%, 50% ou 95%?

Solução:

- $\mu_1 \approx 5 \approx \mu_2$ e $\sigma_1 \approx 2 \approx \sigma_2$. (Eu pedi apenas o valor esperado na prova).
- Positivo pois a correlação entre as variáveis é positiva: quando X_1 está acima de sua média μ_1, a variável X_2 também tende a estar acima de sua média μ_2.
- $\mathbf{e}_1$ e $\mathbf{e}_2$ esboçados na Figura 2.
- $Cov(Z_1, Z_2) = 0$. Este item é mais difícil. A justificativa correta é que

$$\begin{aligned}
\mathrm{Cov}(Z_1, Z_2) &= \mathrm{Cov}((\mathbf{x}-\mu)'\mathbf{e}_1, (\mathbf{x}-\mu)'\mathbf{e}_2) \\
&= \mathrm{Cov}(\mathbf{e}_1'(\mathbf{x}-\mu), \mathbf{e}_2'(\mathbf{x}-\mu)) \\
&= \mathbf{e}_1'\mathrm{Cov}((\mathbf{x}-\mu), (\mathbf{x}-\mu))\mathbf{e}_2 \\
&= \mathbf{e}_1'\mathrm{Cov}((\mathbf{x}-\mu), (\mathbf{x}-\mu))\mathbf{e}_2 \\
&= \mathbf{e}_1'\mathbf{\Sigma}\mathbf{e}_2 \\
&= \mathbf{e}_1'(\lambda_2\mathbf{e}_2) \quad \text{pois } \mathbf{e}_2 \text{ é autovetor} \\
&= \lambda_2\, \mathbf{e}_1'\mathbf{e}_2 = 0 \quad \text{pois os autovetores são ortogonais.}
\end{aligned}$$

- A distância correta é a de Mahalanobis. A está numa região do plano com probabilidade muito menor que B. É por isto que não vemos outras instâncias com valores similares aos de A.
- Estar tão ou mais afastado que B não acontece com frequência. A probabilidade é aproximadamente 5%.

---

5. Seja $\mathbf{X} = (X_1, X_2, X_3)' \sim N_3(\mathbf{0}, \mathbf{\Sigma})$ com

$$\mathbf{\Sigma} = \begin{bmatrix} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 9 \end{bmatrix}$$

- Ache a distribuição de $Y = 3X_1 + 2X_2 + X_3$.
- Para que valores de a e b as variáveis $X_1 + aX_3$ e $X_1 + bX_3$ possuem covariância 0 e portanto são independentes.

Solução: $Y = 3X_1 + 2X_2 + X_3 = (3,2,1)(X_1, X_2, X_3)'$ é uma normal com valor esperado

$$\mathbb{E}(Y) = (3,2,1)\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = (3,2,1)\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0$$

e variância

$$\mathbb{V}(Y) = (3,2,1)\ \mathbf{\Sigma}\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = (3,2,1)\begin{bmatrix} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 9 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 65$$

Para obtermos covariância zero e portanto independência, calculamos

$$\begin{aligned}
\mathrm{Cov}(X_1 + aX_3, X_1 + bX_3) &= \mathrm{Cov}((1,0,a)\mathbf{X}, (1,0,b)\mathbf{X}) \\
&= (1,0,a)\mathbf{\Sigma}\begin{bmatrix} 1 \\ 0 \\ b \end{bmatrix} \\
&= 4 + 2b + 2a + 9ab = 0
\end{aligned}$$

Esta é uma equação linear com duas variáveis e portanto possui infinitas soluçõoes. Por exemplo, tomando $a = 0$ temos $b = -2$. Tomando $a = 1$, temos $b = -6/11$.

---

6. Mostre que a matriz de covariância

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1.0 & 0.63 & 0.45 \\ 0.63 & 1.0 & 0.35 \\ 0.45 & 0.35 & 1.0 \end{bmatrix}$$

do vetor $\mathbf{X} = (X_1, X_2, X_3)' \sim N_3(\mathbf{0}, \boldsymbol{\Sigma})$ pode ser gerada pelo modelo de análise fatorial com $m = 1$ fator e dado por

$$\begin{aligned} X_1 &= 0.9F_1 + \epsilon_1 \\ X_2 &= 0.7F_1 + \epsilon_2 \\ X_3 &= 0.5F_1 + \epsilon_3 \end{aligned}$$

com $var(F_1) = 1$, $\text{Cov}(F_1, \boldsymbol{\epsilon}) = \mathbf{0}$ e

$$\boldsymbol{\Psi} = \text{Cov}(\epsilon) = \begin{bmatrix} 0.19 & 0 & 0 \\ 0 & 0.51 & 0 \\ 0 & 0 & 0.75 \end{bmatrix}$$

Solução: Basta obter a matriz de covariância do vetor

$$\begin{bmatrix} 0.9F_1 + \epsilon_1 \\ 0.7F_1 + \epsilon_2 \\ 0.7F_1 + \epsilon_2 \end{bmatrix} = \begin{bmatrix} 0.9 \\ 0.7 \\ 0.7 \end{bmatrix} F_1 + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_2 \end{bmatrix}$$

Pela independência entre F_1 e o vetor ϵ, a matriz de covariância é a soma das covariâncias:

$$\text{Cov}\left(\begin{bmatrix} 0.9 \\ 0.7 \\ 0.7 \end{bmatrix} F_1 + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_2 \end{bmatrix} \right) = \begin{bmatrix} 0.9 \\ 0.7 \\ 0.7 \end{bmatrix} \mathbb{V}(F_1) + \boldsymbol{\Psi} = \boldsymbol{\Sigma}$$

Se você não se sente a vontade com as manipulações matriciais acima, você pode verificar a igualdade termo a termo na matriz. Por exemplo, vamos verificar que os elementos diagonais de $\boldsymbol{\Sigma}$ são os memsos daqueles obtidos com o modelo fatorial:

$$\begin{aligned} 1.0 &= \boldsymbol{\Sigma}_{11} = \mathbb{V}(X_1) \\ &= \mathbb{V}(0.9F_1 + \epsilon_1) \text{ de acordo com o modelo fatorial} \\ &= \mathbb{V}(0.9F_1) + \mathbb{V}(\epsilon_1) \text{ pela independência entre } F_1 \text{ e } \epsilon_1 \\ &= (0.9)^2 \mathbb{V}(F_1) + 0.19 \\ &= 0.81 \times 1.0 + 0.19 \end{aligned}$$

De forma análoga verificamos que $\mathbb{V}(X_2)$ e $\mathbb{V}(X_3)$ coincidem com os valores obtidos através do modelo fatorial.

Para os elementos fora da digonal,

$$\begin{aligned} 0.63 &= \boldsymbol{\Sigma}_{12} = \text{Cov}(X_1, X_2) \\ &= \text{Cov}(0.9F_1 + \epsilon_1, 0.7F_1 + \epsilon_2) \text{ de acordo com o modelo fatorial} \\ &= \text{Cov}(0.9F_1, 0.7F_1) + \text{Cov}(\epsilon_1, \epsilon_2) \text{ pela independência entre } F_1 \text{ e os } \epsilon\text{'s} \\ &= (0.9)(0.7)\mathbb{V}(F_1) + 0 \\ &= 0.63 \times 1.0 \end{aligned}$$

Os outros são obtidos de maneira análoga.