CC-226
Introdução à Análise de Padrões

Prof. Carlos Henrique Q. Forster

Inferência com Bayes

Distribuição multivariada- PDF conjunta

Definimos a função de densidade de probabilidade conjunta no caso de mais de uma variável:

$$p(x, y) = P(X = x \wedge Y = y)$$

Caso discreto:

$$P[(X, Y) \in A] = \sum_{A} p(x, y)$$

Caso contínuo:

$$P[(X, Y) \in A] = \int \int_{A} p(x, y) dx\, dy$$

Probabilidade de ocorrer uma instância dentro de um (hiper-)retângulo

$$P(a_1 \leq X_1 \leq b_1, \ldots, a_l \leq X_l \leq b_l) = \int_{a_1}^{b_1} \ldots \int_{a_l}^{b_l} p(x_1, \ldots, x_l) dx_1 \ldots dx_l$$

Probabilidade Marginal

Corresponde à soma de todas probabilidades conjuntas para um dado eixo

$$p_x(x) = \sum_y p(x,y)$$

$$p_y(y) = \sum_x p(x,y)$$

As funções p_x e p_y são funções de densidade de probabilidade marginal.

Covariância

Valor esperado no caso conjunto:

$$E[h(x,y)] = \int\int_{\Omega} h(x,y)p(x,y)dxdy$$

Covariância

$$Cov(x,y) = E[(x-\mu_x)\cdot(y-\mu_y)]$$

$$Cov(x,y) = \int\int_{\Omega} (x-\mu_x)(y-\mu_y)p(x,y)dxdy$$

$$Cov(x,y) = E[x\cdot y] - \mu_x \cdot \mu_y$$

Correlação

Coeficiente de correlação

$$Corr(x, y) = \rho_{x,y} = \frac{Cov(x, y)}{\sigma_x \cdot \sigma_y}$$

Propriedade

$$Corr(aX + b, cY + d) = Corr(X, Y)$$

Independência Estatística

Duas variáveis aleatórias são independentes se e só se

$$p(x, y) = p_x(x) \cdot p_y(y), \forall (x, y)$$

Se X e Y são independentes, $\rho = 0$.

Matriz de Covariância

Seja $\sigma_x y = Cov(x, y)$.

Notar que $\sigma_x y = \sigma_y x$ e que $\sigma_x x = Cov(x, x) = E[(x - \mu_x)^2] = \sigma_x^2$

A matriz de covariância Σ é definida como

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma_n^2 \end{pmatrix}$$

Para um vetor coluna de variáveis aleatórias $\mathbf{x}$ com vetor média μ

$$\Sigma = E_{\mathbf{x}}[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T]$$

Distribuição Normal Multivariada

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{m/2} \, \|\boldsymbol{\Sigma}\|^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \, \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2{}_1 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{12} & \sigma^2{}_2 & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1m} & \sigma_{2m} & \cdots & \sigma^2{}_m \end{pmatrix}$$

Exemplo em 2 dimensões

Write r.v. $\mathbf{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$ Then define $X \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to mean

$$p(\mathbf{x}) = \frac{1}{2\pi \,||\, \boldsymbol{\Sigma} \,||^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \ \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

Where the Gaussian's parameters are…

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2{}_x & \sigma_{xy} \\ \sigma_{xy} & \sigma^2{}_y \end{pmatrix}$$

Where we insist that Σ is symmetric non-negative definite

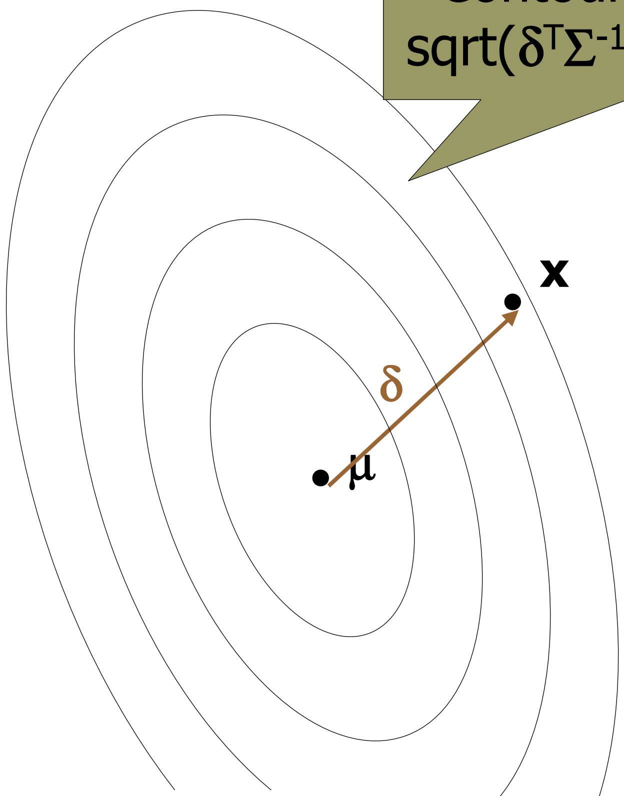
It turns out that E[X] = μ and Cov[X] = Σ. (Note that this is a resulting property of Gaussians, not a definition)

Distância de Mahalanobis

Contours defined by sqrt($\delta^T\Sigma^{-1}\delta$) = constant

1. Begin with vector **x**
2. Define $\delta = \mathbf{x} - \mu$
3. Count the number of contours crossed of the ellipsoids formed Σ^{-1}

   D = this count = sqrt($\delta^T\Sigma^{-1}\delta$) = Mahalonobis Distance between **x** and μ



$$p(\mathbf{x}) = \frac{1}{2\pi \,||\, \mathbf{\Sigma} \,||^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \; \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

Gaussiana com eixos alinhados

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2{}_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sigma^2{}_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \sigma^2{}_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2{}_{m-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sigma^2{}_m \end{pmatrix}$$

Gaussiana esférica

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \sigma^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sigma^2 \end{pmatrix}$$

Propriedades das Gaussianas (A. Moore)

$\begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix} \rightarrow$ Margin-alize $\rightarrow \mathbf{U}$

$$\text{IF } \begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix}\boldsymbol{\mu}_u\\ \boldsymbol{\mu}_v\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{uu} & \boldsymbol{\Sigma}_{uv}\\ \boldsymbol{\Sigma}_{uv}^T & \boldsymbol{\Sigma}_{vv}\end{pmatrix}\right) \quad \text{THEN } \mathbf{U} \sim \mathrm{N}(\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_{uu})$$

Matrix ***A***

$\mathbf{X} \rightarrow$ Multiply $\rightarrow \mathbf{AX}$

$$\text{IF } \mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{AND} \quad \mathbf{Y} = \mathbf{AX} \quad \text{THEN } \mathbf{Y} \sim \mathrm{N}\left(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\,\mathbf{A}^T\right)$$

$\mathbf{X}, \mathbf{Y} \rightarrow + \rightarrow \mathbf{X} + \mathbf{Y}$

$$\text{if } \mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \text{ and } \mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y) \text{ and } \mathbf{X} \perp \mathbf{Y}$$
$$\text{then } \mathbf{X} + \mathbf{Y} \sim \mathrm{N}\left(\boldsymbol{\mu}_x + \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y\right)$$

$\begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix} \rightarrow$ Condition-alize $\rightarrow \mathbf{U} \mid \mathbf{V}$

$$\text{IF } \begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix}\boldsymbol{\mu}_u\\ \boldsymbol{\mu}_v\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{uu} & \boldsymbol{\Sigma}_{uv}\\ \boldsymbol{\Sigma}_{uv}^T & \boldsymbol{\Sigma}_{vv}\end{pmatrix}\right) \quad \text{THEN } \mathbf{U} \mid \mathbf{V} \sim \mathrm{N}\left(\boldsymbol{\mu}_{u|v}, \boldsymbol{\Sigma}_{u|v}\right)$$

$$\text{where} \quad \boldsymbol{\mu}_{u|v} = \boldsymbol{\mu}_u + \boldsymbol{\Sigma}_{uv}^T \boldsymbol{\Sigma}_{vv}^{-1}(\mathbf{V} - \boldsymbol{\mu}_v) \qquad \boldsymbol{\Sigma}_{u|v} = \boldsymbol{\Sigma}_{uu} - \boldsymbol{\Sigma}_{uv}^T \boldsymbol{\Sigma}_{vv}^{-1} \boldsymbol{\Sigma}_{uv}$$

$\mathbf{U} \mid \mathbf{V} \rightarrow$, $\mathbf{V} \rightarrow$ Chain Rule $\rightarrow \begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix}$

$$\text{IF } \quad \mathbf{U} \mid \mathbf{V} \sim \mathrm{N}\left(\mathbf{AV}, \boldsymbol{\Sigma}_{u|v}\right) \text{ and } \mathbf{V} \sim \mathrm{N}(\boldsymbol{\mu}_v, \boldsymbol{\Sigma}_{vv})$$

$$\text{THEN } \begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ with } \boldsymbol{\Sigma} = \begin{pmatrix}\mathbf{A}\boldsymbol{\Sigma}_{vv}\mathbf{A}^T + \boldsymbol{\Sigma}_{u|v} & \mathbf{A}\boldsymbol{\Sigma}_{vv}\\ (\mathbf{A}\boldsymbol{\Sigma}_{vv})^T & \boldsymbol{\Sigma}_{vv}\end{pmatrix}$$

Noção da dimensionalidade

- Considere uma variável X e cada componente x_i independente

  Considere uma amostra de tamanho 10000 e fazemos a poda de 5% dos valores extremos para cada variável.

  Quanto sobra da amostra para diferentes dimensões (no. de atributos) dos dados?

Solução

- Para uma variável, multiplica por 95%:
  - 10000*0.95 = 9500
- Para duas dimensões, multiplica novamente por 95%:
  - 9025
- Para 10 dimensões, multiplica por $(0.95)^{10}$:
  - 5987 (quase metade dos dados foram podados)
- Para 50 dimensões:
  - 769
- Para 100 dimensões:
  - 59 (a grande maioria está nos 5% outliers de alguma variável)

Probabilidade Condicional

$$P\left(X \mid Y\right) = \frac{P\left(X \cap Y\right)}{P\left(Y\right)}$$

Exercício

- Num programa de TV você deve escolher uma dentre três portas das quais uma delas tem um prêmio. O apresentador escolhe uma das portas que você não escolheu e abre para mostrar que o prêmio não está lá. Você é dado a oportunidade de trocar a porta que escolheu pela outra que ainda não está aberta. Qual é a melhor opção?

Exercício http://www.regentsprep.org/regents/math/algebra/APR3/PracCond.htm

A survey of middle school students asked: What is your favorite winter sport? The results are summarized below:

Favorite Winter Sport

Grade	Snowboarding	Skiing	Ice Skating	TOTAL
6th	68	41	46	155
7th	84	56	70	210
8th	59	74	47	180
TOTAL	211	171	163	545

Using these 545 students as the sample space, a student from this study is randomly selected.

Exercício (cont)

a.) What is the probability of selecting a student whose favorite sport is skiing?

b.) What is the probability of selecting a 6th grade student?

c.) If the student selected is a 7th grade student, what is the probability that the student prefers ice-skating?

d.) If the student selected prefers snowboarding, what is the probability that the student is a 6th grade student?

e.) If the student selected is an 8th grade student, what is the probability that the student prefers skiing or ice-skating?

Part a: 171/545

Part b: 155/545 = 31/109

Part c: 70/210 = 1/3

Part d: 68/211

Part e: (74 + 47)/180 = 121/180

Exercício

- Na tabela dada, identificar ou calcular:
  - Distribuição de probabilidade conjunta P(A,B)
  - Probabilidades marginais de cada variável P(A) e P(B)
  - Probabilidades condicionais P(A|B) e P(B|A)
- Qual a relação entre P(A|B) e P(B|A)?
  (que tal montar uma tabela dividindo uma pela outra?)