II Probability theory

(2)

$P(A) = \frac{2}{6} = 1/3$

$P(B|A) = \frac{P(A,B)}{P(A)} = \frac{1/3}{1/3} = 1$

$P(A,B) = \frac{2}{6} = 1/3$

$P(A,B,C,D) = 0$

$P(A,B,C) = 1/6$

$$P(A|B,C) = \frac{P(A,B,C)}{P(B,C)} = \frac{1/6}{1/6} = 1$$

$$P(D|A,B,C) = \frac{P(A,B,C,D)}{P(A,B,C)} = 0$$

(3)

$\{(-2,2), (-1,3), (0,1), (-2,1)\}$

(a)

$$\mu = \frac{1}{4}\left(\begin{bmatrix}-2\\2\end{bmatrix} + \begin{bmatrix}-1\\3\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}-2\\1\end{bmatrix}\right) = \begin{bmatrix}-1,25\\1,75\end{bmatrix}$$

Vamos ter $\begin{bmatrix}\Sigma_{00} & \Sigma_{01}\\ \Sigma_{10} & \Sigma_{11}\end{bmatrix}$

$$\Sigma_{00} = \frac{1}{4-1}\left[(-2+1,25)^2 + \dots + (-2+1,25)^2\right] = 0,9167$$

Como é que varia em relação à outra dim

$$\Sigma_{01} = \frac{1}{4-1}\left[(-2+1,25)(-2-1,75)\dots\right] = -0,0833$$

$$\Sigma_{10} = \Sigma_{01}$$

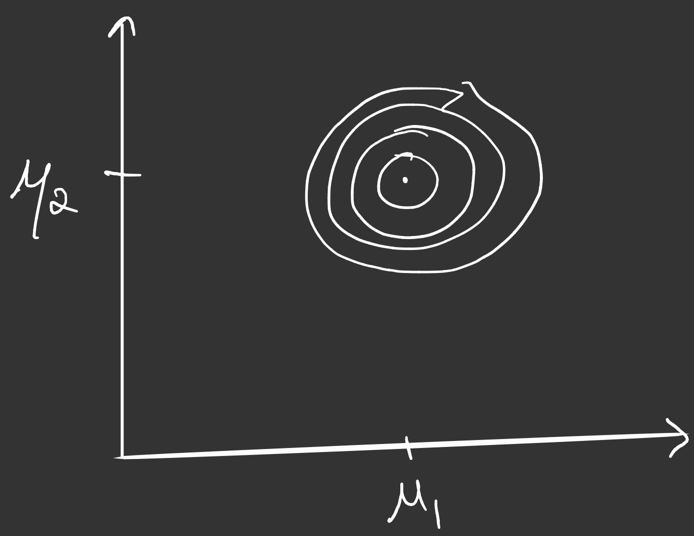
$$\Sigma_{11} = \frac{1}{4}\left[(2-1{,}75)^2 + \dots + (1-1{,}75)^2 \right] = 0{,}9167$$

$$\det(\Sigma) = \det\begin{pmatrix} 0{,}9167 & -0{,}0833 \\ -0{,}0833 & 0{,}9167 \end{pmatrix} = 0{,}9167 \times 0{,}9167 - (-0{,}0833 \times (-0{,}0833))$$

$$\Sigma^{-1} = \frac{1}{\det(\Sigma)} \begin{bmatrix} 0{,}9167 & 0{,}0833 \\ 0{,}0833 & 0{,}9167 \end{bmatrix} = \begin{bmatrix} 1{,}1 & 0{,}1 \\ 0{,}1 & 1{,}1 \end{bmatrix}$$

$$\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} \right) = \frac{1}{\det} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$N(u|\mu,\Sigma) = \frac{1}{2\pi\sqrt{0{,}0833}} \times \exp\left(-\frac{1}{2} \left(\begin{bmatrix} u_0 \\ u_1 \end{bmatrix} - \begin{bmatrix} -1{,}25 \\ 1{,}75 \end{bmatrix} \right)^T \times \begin{bmatrix} 1{,}1 & 0{,}1 \\ 0{,}1 & 1{,}1 \end{bmatrix} \left(\begin{bmatrix} u_0 \\ u_1 \end{bmatrix} - \begin{bmatrix} -1{,}25 \\ 1{,}75 \end{bmatrix} \right) \right)$$



III Bayesian learning

$$\underbrace{p(H|D)}_{\text{posterior prob}} = \frac{\overbrace{P(D|H)}^{\text{likelihood}} \cdot \overbrace{p(H)}^{\text{prior}}}{\underbrace{P(D)}_{\text{evidence}}}$$

4 a $u_{new} [1\ 1\ 1\ 1\ 1]^T$

$$P(C=0 \mid y_1=1, y_2=1, y_3=1, y_4=1, y_5=1) =$$

$$= \frac{P(C=0)\, P(y_1=1, \dots, y_5=1)}{P(y_1=1, \dots, y_5=1)}$$

$$\boxed{2^5}$$

Sem a assumption

$$P(y_1=1, \dots, y_5=1) = P(y_1=1, \dots, y_5=1, C=0) +$$

$$P(y_1=1, \dots, y_5=1, C=1) =$$

$$= \underbrace{P(y_1=1, \dots, y_5=1 \mid C=0)}_{0 \text{ (n̄ está no dataset}} \cdot P(C=0) + \overbrace{P(y_1=1, \dots, y_5=1 \mid C=1)}^{0}\, P(C=1) =$$

$$= 0 \times \frac{3}{7} + 0 \times \frac{4}{7} = 0$$

ⓒ $P(C=0 \mid y_1=1, \ldots, y_5=1) = \dfrac{P(y_1=1, \ldots, y_5=1 \mid C=0) . P(C=0)}{P(y_1=1, \ldots, y_5=1)} =$

$$= \frac{P(y_1=1, \ldots, y_5=1 \mid C=0)}{P(y_1=1, \ldots, y_5=1)} . p(C=0) =$$

$$= \frac{p(C=0)\, p(y_1=1 \mid C=0) \times p(y_2=1 \mid C=0) \times \ldots \times p(y_5=1 \mid C=0)}{P(y_1=1, \ldots, y_5=1)}$$

$$= \frac{3/7 \times 1/3 \times 2/3 \times 2/3 \times 2/3 \times 1/3}{P(y_1=1, \ldots, y_5=5)} = \frac{0{,}0141}{P(y_1=1, \ldots, y_5=1)}$$

$$P(C=1 \mid y_1=1, \ldots, y_5=1) = \frac{4/7 \times 1/2 \times 1/4 \times 1/2 \times 1/2 \times 1/2}{P(y_1=1, \ldots, y_5=1)} =$$

$$= \frac{0{,}00892}{P(y_1=1, \ldots, y_5=1)}$$

é maior

logo $C=0$, ñ é Iphone

ⓓ igual à anterior, mas ignoramos os valores que não conhecemos.

5 $u_{new} = [100 \quad 225]^T$

$P(C=0) = 1/2, \; P(C=1) = 1/2$

$P(y_1, y_2 | C=0) \quad P(y_1, y_2 | C=1)$

$$\mu \quad \begin{array}{c} y_1 \\ y_2 \end{array} \begin{bmatrix} 93,(3) \\ 156,(6) \end{bmatrix} \qquad \begin{array}{c} y_1 \\ y_2 \end{array} \begin{bmatrix} 80 \\ 203,(3) \end{bmatrix}$$

$$\Sigma \quad \begin{bmatrix} 4433,(3) & 216,(6) \\ 216,(6) & 33,(3) \end{bmatrix} \qquad \begin{bmatrix} 100 & 50 \\ 50 & 233,(3) \end{bmatrix}$$

$$P(C=0 | y_1=100, y_2=225)$$

$$= \frac{p(C=0) \cdot p(y_1=100, y_2=225 | C=0)}{p(y_1=100, y_2=225)}$$

$$= \frac{\frac{1}{2} \times N\left(\begin{bmatrix} 100 \\ 225 \end{bmatrix} \middle| \mu = \begin{bmatrix} 93,(3) \\ 156,(6) \end{bmatrix}, \Sigma = \begin{bmatrix} 4433,(3) & 216,(6) \\ 216,(6) & 33,(3) \end{bmatrix}\right)}{P(y_1=100, y_2=225)}$$

$$= \frac{\boxed{3,4783 \times 10^{-48}}}{P(y_1=100, y_2=225)}$$

C=1, logo é NBA player

$$P(C=1 | y_1=100, y_2=225) = \frac{\boxed{0,0001}}{p(y_1=100, y_2=225)}$$

(b)

	$p(y_1 \mid C=0)$	$p(y_1 \mid C=1)$	$p(y_2 \mid C=0)$	$p(y_2 = C1)$
μ	93,(3)	80	156,(6)	203,3
σ	66,58	10	5,77	15,275

$$P(C=0 \mid y_1=100, y_2=225) \stackrel{B}{=} \frac{p(C=0)\, p(y_1=100 \mid C=0)\, p(y_2=225 \mid C=0)}{p(y_1=100, y_2=225)}$$

$$= \frac{\frac{1}{2} N(100 \mid \mu=93,(3), \sigma=66,58)\, N(225 \mid \mu=156,(6), \sigma=5,77)}{p(y_1=100, y_2=225)}$$

$$= \frac{\boxed{7,854 \times 10^{-35}}}{P(y_1=100, y_2=225)}$$

Novamente NBA player

$$P(C=1 \mid y_1=100, y_2=225) = \ldots = \frac{\boxed{2,578 \times 10^{-5}}}{P(y_1=100, y_2=225)}$$

(6) m features

classes binárias

(a) i) $p(C=0) = 1 - p(C=1)$ } 1 parâmetro (se sabemos 1, sabemos o outro)

$$\underbrace{p(y_1=v_1, \ldots, y_m=v_m \mid C=0)}_{2^m}$$

$$\underbrace{p(y_1=v_1, \ldots, y_m=v_m \mid C=1)}_{2^m}$$

$\underbrace{p(y_1 = v_1, \ldots, y_m = v_m)}_{2^m}$

Total $= 3 \times 2^m + 1$

ii) $p(C=0) = 1 - P(C=1)$ } 1 parâmetro

$P(y_i \mid C=0)$ } calcular para as m features, ou seja, m parâmetros

$p(y_i \mid C=1)$ } m parâmetros

total: $1 + 2m$

b

iii) $P(C=0) = 1 - P(C=1)$ } 1 parâmetro

$$\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_m \end{bmatrix}$$ m parâmetros

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \cdots & \Sigma_{1m} \\ \vdots & \ddots & \\ \Sigma_{m1} & \cdots & \Sigma_{mm} \end{bmatrix}$$

m parâmetros

$m \cdot \left(\frac{m-1}{2}\right) \rightarrow$ é simétrica e tira-se a diagonal

temos $m + m\left(\frac{m-1}{2}\right)$

total: $1 + 2m + m\left(\frac{m-1}{2}\right)$

iv) $P(C=0) = 1 - P(C=1)$ } 1 parâmetro

$P(y_1 \mid C=0) \Rightarrow$ m parâmetros

$P(y_2 \mid C=1) \rightarrow$ m parâmetros

$\mu\ \sigma$ } 2 parâmetros

total $= 1 + 2m \times 2 = 1 + 4m$