HW2

Saturday, 28 September 2024 11:43

1. a)

likelihood ↓ prior ↓

$$P(C=A \mid x_1=1, x_2=2) = \frac{P(x_1=1, x_2=2 \mid C=A) \times p(C=A)}{p(D)}$$

posterior prob — evidence

$P(x_1=1, x_2=2 \mid C=A)$ Naive Bayes

↓

$P(x_1=1 \mid C=A)\ P(x_2=2 \mid C=A)$

We are going to test for A and B and then compare (we only need the numerator because $P(x=x)$ is equal in both)

$$\frac{\overset{4/8}{P(C=A)}\ \overset{N_1(\mu_1,\sigma_1)}{P(x_1=1 \mid C=A)} \times \overset{N_2(\mu_2,\sigma_2)}{P(x_2=2 \mid C=A)}}{\cancel{P(x=x)}} \quad \boxed{\begin{matrix}<\\>\\=\end{matrix}} \quad \frac{\overset{4/8}{P(C=B)}\ P(x_1=1 \mid C=B) \times P(x_2=2 \mid C=B)}{P(x=x)}$$

$$\boxed{N = \frac{1}{\sqrt{2\pi} \times \sigma} \exp\left[-\frac{1}{2\sigma^2} \cdot (x-\bar{x})^2\right],}$$

for x_1: $\bar{x}_{1A} = 1.25$

and $C=A$ $\sigma_{1A} = \sqrt{Var(x_{1A})} = \sqrt{\frac{(0.6-1.25)^2+(1-1.25)^2+(1.6-1.25)^2+(1.8-1.25)^2}{4-1}} = 0.55$

$N(1 \mid 1.25, 0.55) \to N_{1A} = 0.654$

and $C=B$: $\bar{x}_{1B} = 2.75$

$\sigma_{1B} = \sqrt{Var(x_{1B})} = \sqrt{\frac{(2-2.75)^2+(2-2.75)^2+(3-2.75)^2+(4-2.75)^2}{4-1}} = 0.957$

$N(1 \mid 2.75, 0.957) \to N_{1B} = 0.078$

for x_2: $\bar{x}_2 = 1.2$

and $C=A$ $\sigma_{2A} = \sqrt{\frac{(0.4-1.2)^2+(1.1-1.2)^2+(1.5-1.2)^2+(1.8-1.2)^2}{4-1}} = 0.605$

$N(2 \mid 1.2, 0.605) \to N_{2A} = 0.275$

and $C=B$: $\bar{x}_{2B} = 0.55$

$\sigma_{2B} = \sqrt{\frac{(0-0.55)^2+(1-0.55)^2+(0-0.55)^2+(1.2-0.55)^2}{4-1}} = 0.64$

$N(2 \mid 0.55, 0.64) \to N_{2B} = 0.048$

$$\frac{\frac{4}{8} \times N_{1A} \times N_{2A}}{P(x=x)} \quad \boxed{\begin{matrix}<\\>\\=\end{matrix}} \quad \frac{\frac{4}{8} \times N_{1B} \times N_{2B}}{P(x=x)}$$

$N_{1A} \times N_{2A} > N_{1B} \times N_{2B}$

$0.1798 > 0.00374$

So, the most probable class for query vector $x = (1,2)$ is A

---

b) → 2 dimensional

em vez de σ_1 e σ_2

usar $\Sigma \begin{array}{c|cc} & x_1 & x_2 \\ x_1 & Var(x_1) & \\ x_2 & cov(x_1,x_2) & Var(x_2) \end{array}$

$$\frac{\overset{4/8}{P(C=A)}\ \overset{N_A(x \mid \bar{x}, \Sigma)}{P(x_1=1, x_2=2 \mid C=A)}}{\cancel{P(x=x)}} \quad \boxed{\begin{matrix}<\\>\\=\end{matrix}} \quad \frac{\overset{4/8}{P(C=B)}\ \overset{N_B(x \mid \bar{x}, \Sigma)}{P(x_1=1, x_2=2 \mid C=B)}}{P(x=x)}$$

$$\bar{x}_A = \begin{bmatrix} \bar{x}_{1A} \\ \bar{x}_{2A} \end{bmatrix} = \begin{bmatrix} 1.25 \\ 1.20 \end{bmatrix}$$

$$\bar{x}_B = \begin{bmatrix} \bar{x}_{1B} \\ \bar{x}_{2B} \end{bmatrix} = \begin{bmatrix} 2.75 \\ 0.55 \end{bmatrix}$$

$$\Sigma_A = \begin{array}{c|cc} & x_1 & x_2 \\ x_1 & Var\,x_{1A} & cov(x_{1A}, x_{2A}) \\ x_2 & cov(x_{2A}, x_{1A}) & Var(x_{2A}) \end{array} = \begin{bmatrix} 0.55^2 & 0.326 \\ 0.326 & 0.605^2 \end{bmatrix}$$

$$\Sigma_A^{-1} = \begin{bmatrix} 82.451 & -73.44 \\ -73.44 & 68.136 \end{bmatrix}$$

$$cov(x_{1A}, x_{2A}) = \frac{1}{3}\Big[(0.6-1.25)(0.4-1.2) + (1-1.25)(1.1-1.2) + (1.6-1.25)(1.5-1.2) + (1.8-1.25)(1.8-1.2)\Big]$$

≈ 0.326

$$\det(\Sigma_A) = (0.55^2 \times 0.605^2) - 0.326^2 \approx 0.00444$$

$$N_A = N(\overset{\text{vetores}}{x} \mid \bar{x}_A, \Sigma_A) = \frac{1}{2\pi} \times \frac{1}{\sqrt{\det(\Sigma)}} \cdot \exp\left[-\frac{1}{2}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \bar{x}_{1A} \\ \bar{x}_{2A} \end{bmatrix}\right)^T \times \left[\Sigma_A^{-1}\right] \cdot \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \bar{x}_{1A} \\ \bar{x}_{2A} \end{bmatrix}\right)\right]$$

$$-\frac{1}{2}\begin{pmatrix} 1-1.25 \\ 2-1.2 \end{pmatrix}$$

$$= \frac{1}{2\pi} \times \frac{1}{\sqrt{0.00444}} \exp\Big(\overbrace{[0.125\ \ -0.4] \times \begin{bmatrix} 82.451 & -73.44 \\ -73.44 & 68.136 \end{bmatrix}}^{39} \begin{bmatrix} -0.25 \\ 0.8 \end{bmatrix}\Big)$$

$$= 2{,}75831 \times 10^{-17}$$

$$\Sigma_B = \begin{bmatrix} Var\,x_{1B} & cov(x_{1B}, x_{2B}) \\ cov(x_{1B}, x_{2B}) & Var\,x_{2B} \end{bmatrix} = \begin{bmatrix} 0.957^2 & 0.25 \\ 0.25 & 0.64^2 \end{bmatrix}$$

$$\Sigma_B^{-1} = \begin{bmatrix} 1{,}2896 & -0{,}78828 \\ -0{,}78828 & 2{,}8883 \end{bmatrix}$$

$$cov(x_{1B}, x_{2B}) = \frac{1}{3}\Big[(2-2.75)(0-0.55) + (2-2.75)(1-0.55) + (3-2.75)(0-0.55) + (4-2.75)(1.2-0.55)\Big]$$

$= 0.25$

$$\det(\Sigma_B) = (0.957^2 \times 0.64^2) - (0.25)^2 \approx 0.313$$

$$N_B = N(\overset{\text{vetores}}{x} \mid \bar{x}_B, \Sigma_B) = \frac{1}{2\pi} \times \frac{1}{\sqrt{\det(\Sigma)}} \cdot \exp\left[-\frac{1}{2}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \bar{x}_{1B} \\ \bar{x}_{2B} \end{bmatrix}\right)^T \times \left[\Sigma^{-1}\right] \cdot \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \bar{x}_{1B} \\ \bar{x}_{2B} \end{bmatrix}\right)\right]$$

$$= \frac{1}{2\pi} \times \frac{1}{\sqrt{0.313}} \exp\left[-\frac{1}{2}\underbrace{\left(\begin{bmatrix} 1-2.75 \\ 2-0.55 \end{bmatrix}\right)^T \begin{bmatrix} 1{,}2896 & -0{,}78882 \\ -0{,}78882 & 2{,}8883 \end{bmatrix}}_{[0{,}873\ \ -0{,}725]} \cdot \begin{bmatrix} 1-2.75 \\ 2-0.55 \end{bmatrix}\right]$$

$$= \frac{1}{2\pi} \times \frac{1}{\sqrt{0{,}313}} \times e^{(-2{,}6523)} = 0{,}02004$$

the most probable class is B

it is different from exercise 1a) because it is 2 dimensional, so it doesn't assume independence between x_1 and x_2

c) $P(C=A \mid x_3=1) = \frac{\#\,\text{favoráveis}}{\#\,\text{totais}} = \quad \boxed{\begin{matrix}<\\>\\=\end{matrix}} \quad P(C=B \mid x_3=1)$

$= \frac{2}{\#\,\text{totais}} < \frac{3}{\#\,\text{totais}}$

↓

A classe é a B

d) usar resultados passados e multiplicar (Bayes

x_3 indep. (x_1, x_2) (x_3)

$P(1,2 \mid A) \cdot P(1 \mid A) \cdot p(A) = (N_A) \times \frac{2}{4} \times \frac{4}{8}$

$P(1,2 \mid B) \cdot P(1 \mid B) \cdot p(B) = N_B \times \frac{3}{4} \times \frac{4}{8}$

$P(A, x_{query}) =$

$= 2{,}75831 \times 10^{-17} \times \frac{2}{4} \times \frac{4}{8} = 6{,}896 \times 10^{-18}$

$P(B, x_{query}) =$

$= 0.02004 \times \frac{3}{4} \times \frac{4}{8} \approx 7{,}51 \times 10^{-3}$

R: Classe B

3.

dataset digits

	K = 3	K = 30
Knn accuracy	→ 0.98	" → 0.95
Gauss accuracy	→ 0.79	" → 0.79

now with dataset wine:

	K = 3	K = 30
Knn accuracy	→ 0.67	" → 0.68
Gauss accuracy	→ 0.96	→ 0.96

In the case of the digits dataset, because it deals with images, less neighbours (K's) is better, because the local proximity is more relevant. But for a more general and straightforward dataset, like the wine dataset, having more neighbours is better for accuracy. Assuming gaussian distribution is also better for the specificities mentioned previously. So the accuracy of gaussian for the wine dataset was higher than the one in the digits dataset.