LEIC-T 2023/2024
Aprendizagem - Machine Learning
Homework 2
Deadline 9/10/2024 20:00
Submit on Fenix as pdf

I) Bayesian Classifier (8 pts)

Given a data set describing a sample

X1	X2	Class
0.5	0.5	A
1	1.5	A
1.5	0.8	A
2	1.8	A
2	0	B
2	1	B
3	0	B
5	1.2	B

And the query vector $x = (x_1, x_2)^T = (1,2)^T$

a) (3pts) Compute the most probable class for the query vector, under the Naive Bayes assumption, using 1-dimensional Gaussians to model the likelihoods. (Hint, the likelihood is described of each class is described by two Gaussians (Normal Distributions, each distribution is defined by a mean value and standard deviation.)

p(C=A) = $\frac{1}{2}$

p(C=B) = $\frac{1}{2}$

μ → média; σ → desvio padrão

	p(X1|C=A)	P(X1|C=B)
μ	1.25	1.15
σ	0.6455	0.60277
	p(X2|C=A)	P(X2|C=B)
μ	3	0.55
σ	1.4142	0.6403

$$P(C{=}A|X1{=}1,X2{=}2) = \frac{p(C{=}A)p(X1{=}1,X2{=}2|C{=}A)}{p(X1{=}1,X2{=}2)} = \frac{p(C{=}A)p(X1{=}1|C{=}A)p(X2{=}2|C{=}A)}{p(X1{=}1,X2{=}2)} =$$

$$= \frac{\frac{1}{2}N(1|\mu= 1.25,\sigma =0.6455)N(2|\mu= 1.15,\sigma =0.60277)}{p(X1{=}1,X2{=}2)} = \frac{0.183967}{p(X1{=}1,X2{=}2)}$$

$$P(C{=}B|X1{=}1,X2{=}2) = \frac{p(C{=}B)p(X1{=}1,X2{=}2|C{=}B)}{p(X1{=}1,X2{=}2)} = \frac{p(C{=}B)p(X1{=}1|C{=}A)p(X2{=}2|C{=}B)}{p(X1{=}1,X2{=}2)} =$$

$$= \frac{\frac{1}{2}N(1|\mu= 3,\sigma =1.4142)N(2|\mu= 0.55,\sigma =0.6403)}{p(X1{=}1,X2{=}2)} = \frac{0.025256}{p(X1{=}1,X2{=}2)}$$

Comparando os valores dos numeradores é possível perceber que o valor quando C=A é superior ao de C=B, assim concluímos que A é a classe mais provável para a query vector

b) (3 pts) Compute the most probable class for the query vector assuming that the likelihoods are 2-dimensional Gaussians.

P(C=A) = $\frac{1}{2}$

P(C=B) = $\frac{1}{2}$

P(X1,X2|C=A):

$\mu = \frac{1}{4} \times (\begin{bmatrix}0.5\\0.5\end{bmatrix}+\begin{bmatrix}1\\1.5\end{bmatrix}+\begin{bmatrix}1.5\\0.8\end{bmatrix}+\begin{bmatrix}2\\1.8\end{bmatrix}) = \begin{bmatrix}1.25\\1.15\end{bmatrix}$

$\Sigma 00 = \frac{1}{4-1}((0.5-1.25)^2+(1-1.25)^2+(1.5-1.25)^2+(2-1.25)^2)$ = **0.41667**

$\Sigma 01 = \frac{1}{4-1}((0.5-1.25)(0.5-1.15)+(1-1.25)(1.5-1.15)+(1.5-1.25)(0.8-1.15)+(2-1.25)(1.8-1.15))$ = **0.2667**

$\Sigma 10 = \Sigma 01$ = **0.2667**

$\Sigma 11 = \frac{1}{4-1}((0.5-1.15)^2+(1.5-1.15)^2+(0.8-1.15)^2+(1.8-1.15)^2)$ = **0.3633**

$\Sigma = \begin{bmatrix}0.4167 & 0.2667\\0.2667 & 0.3633\end{bmatrix}$; **det(Σ) = 0.08026**

P(X1,X2|C=B):

$\mu = \frac{1}{4} \times (\begin{bmatrix}2\\0\end{bmatrix}+\begin{bmatrix}2\\1\end{bmatrix}+\begin{bmatrix}3\\0\end{bmatrix}+\begin{bmatrix}5\\1.2\end{bmatrix}) = \begin{bmatrix}3\\0.55\end{bmatrix}$

$\Sigma 00 = \frac{1}{4-1}((2-3)^2+(2-3)^2+(3-3)^2+(5-3)^2)$ = **2**

$\Sigma 01 = \frac{1}{4-1}((2-3)(0-0.55)+(2-3)(1-0.55)+(3-3)(0-0.55)+(5-3)(1.2-0.55))$ = **0.4667**

$\Sigma 10 = \Sigma 01$ = **0.4667**

$\Sigma 11 = \frac{1}{4-1}((0-0.55)^2+(1-0.55)^2+(0-0.55)^2+(1.2-0.55)^2)$ = **0.41**

$\Sigma = \begin{bmatrix}2 & 0.4667\\0.4667 & 0.41\end{bmatrix}$; **det(Σ) = 0.60219**

P(C=A|X1=1,X2=2) = $\frac{p(C=A)p(X1=1,X2=2|C=A)}{p(X1=1,X2=2)} = \frac{\frac{1}{2}N(\begin{bmatrix}1\\2\end{bmatrix}|\mu=\begin{bmatrix}1.25\\1.15\end{bmatrix},\Sigma=\begin{bmatrix}0.4167 & 0.2667\\0.2667 & 0.3633\end{bmatrix})}{p(X1=1,X2=2)} = \frac{0.0257}{p(X1=1,X2=2)}$

P(C=B|X1=1,X2=2) = $\frac{p(C=B)p(X1=1,X2=2|C=B)}{p(X1=1,X2=2)} = \frac{\frac{1}{2}N(\begin{bmatrix}1\\2\end{bmatrix}|\mu=\begin{bmatrix}3\\0.55\end{bmatrix},\Sigma=\begin{bmatrix}2 & 0.4667\\0.4667 & 0.41\end{bmatrix})}{p(X1=1,X2=2)} = \frac{0.004724}{p(X1=1,X2=2)}$

Comparando os valores dos numeradores é possível perceber que o valor quando C=A é superior ao de C=B, assim concluímos que A é a classe mais provável para a query vector

c) (1 pts) Given a data set

X3	Class
0	A
1	A
1	A
0	A
1	B
1	B
0	B
1	B

And the query vector *x3 =True=1*

Compute the most probable class, with x3 being a categorial class 1=True, 0=False.

P(C=A) = $\frac{1}{2}$

P(C=B) = $\frac{1}{2}$

P(X3 = 1) = $\frac{5}{8}$

P(X3 = 0) = $\frac{3}{8}$

p(C = A, X3 = 1) = $\frac{1}{2}$

p(C = A|X3 = 1) = $\frac{\mathbf{1}}{\mathbf{2}} \div \frac{\mathbf{5}}{\mathbf{8}} = \frac{\mathbf{4}}{\mathbf{5}}$

p(C = A, X3 = 0) = $\frac{1}{2}$

p(C = A|X3 = 0) = $\frac{\mathbf{1}}{\mathbf{2}} \div \frac{\mathbf{3}}{\mathbf{8}} = \frac{\mathbf{4}}{\mathbf{3}}$

p(C = B, X3 = 1) = $\frac{3}{4}$

p(C = B|X3 = 1) = $\frac{\mathbf{3}}{\mathbf{4}} \div \frac{\mathbf{5}}{\mathbf{8}} = \frac{\mathbf{6}}{\mathbf{5}}$

p(C = B, X3 = 0) = $\frac{1}{4}$

p(C = B|X3 = 0) = $\frac{\mathbf{1}}{\mathbf{4}} \div \frac{\mathbf{3}}{\mathbf{8}} = \frac{\mathbf{2}}{\mathbf{3}}$

Quando X3 = 1(True) a classe mais provável é a classe B.

Quando X3 = 0(False) a classe mais provável é a classe A.

d) (1pts) Given a data set describing a sample combining the data set before

X1	X2	X3	Class
0.5	0.5	0	A
1	1.5	1	A
1.5	0.8	1	A
2	1.8	0	A
2	0	1	B
2	1	1	B
3	0	0	B
5	1.2	1	B

x1 and x2 are dependable and x3 is independent of x1 and x2. x3 is a categorial class. And the query vector $x = (1,2,1)^T$ Compute the most probable class and indicate the estimated relative probability.

Hint,

p(A, xquery) = p((1,2)|A) ·P(1|A) ·p(A)

p(B, xquery) = p((1,2)|B) ·P(1|B) ·p(B)

you have already computed the values in b) and in c)

P(1|A) =card(A.1)/card(A)=2/4

P(1|B) =card(A.1)/card(B)=3/4

P(C=A) = $\frac{1}{2}$

P(C=B) = $\frac{1}{2}$

p(A, xquery) = p((1,2)|A)p(1|A)p(A)

p((1,2)|A)= $\frac{0.0257}{\frac{1}{2}}$ = 0.0514

p(1|A) = 0.5

probabilidade posterior da classe A = p((1,2)|A) p(1|A)p(A)=0.01285

p(B, xquery) = p((1,2)|B)p(1|B)p(B)

p((1,2)|A)= $\frac{0.004724}{\frac{1}{2}}$ = 0.009448

p(1|B) = 0.75

probabilidade posterior da classe B = p((1,2)|B) p(1|B)p(B)=0.003543

Como é possível perceber A tem uma probabilidade posterior superior assim sendo é possível inferir que A é a classe mais provável.

A probabilidade relativa de A será $\frac{0.01285}{0.003543}$ = 3.6269 .

III Software Experiments (2pts)

Download the jupyter notebook HM2_kB.ipynb.

Split the data using the command (in the notebook)

```
digits = datasets.load_digits()
X, y = digits.data, digits.target
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7,stratify=y,random_state=your_group number)
```

And do the experiments with kNN with k=1, k=3, and GaussNB as indicated in the file and indicate the accuracy results.

Load the wine data set wine = datasets.load_wine() and redo the experiments, indicate the new accuracy values.

Which method gives better result for which data set? Do you know why? Pls indicate in one sentence.

Digits data set:
train size: 1257
test size: 540
KNN (k=1) accuracy on testing set: 0.98
KNN (k=3) accuracy on testing set: 0.99
Gauss NB accuracy on testing set: 0.84

Wine data set:
Train size: 124
Test size: 54
KNN (k=1) accuracy on testing set: 0.7
KNN (k=3) accuracy on testing set: 0.7
Gaussian NB accuracy on testing set: 0.98

KNN com k=3 dá melhores resultados para o conjunto de dados Digits pois é eficaz com dados de dimensões reduzidas e com pouca variância neste caso os pixéis são pretos ou brancos, enquanto Gaussian Naive Bayes tem melhor desempenho no conjunto de dados Wine, porque os atributos do conjunto são condicionalmente independentes e contínuos.