I)

a)

p(C=A) =

p(C=B) =

µ -> média; σ -> desvio padrão

|  |  |  |
| --- | --- | --- |
|  | p(X1|C=p 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) = = = = =

P(C=B|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)

P(C=A) =

P(C=B) =

P(X1,X2|C=A):

µ = x (+++) =

Σ00 = () = 0.41667

Σ01 = (+ ) = 0.2667

Σ10 == 0.2667

Σ11 = () = 0.3633

Σ = ; det(Σ) = 0.08026

P(X1,X2|C=B):

µ = x (+++) =

Σ00 = () = 2

Σ01 = (+ ) = 0.4667

Σ10 == 0.4667

Σ11 = () = 0.41

Σ = ; det(Σ) = 0.60219

P(C=A|X1=1,X2=2) = = =

P(C=B|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)

P(C=A) =

P(C=B) =

p(C = A|X3 = 1) = **÷**

p(C = A|X3 = 0) = **÷**

p(C = B|X3 = 1) = **÷**

p(C = B|X3 = 0) = **÷**

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)

P(C=A) =

P(C=B) =

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

p((1,2)|A)= = 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)= = 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 assim sendo é possível inferir que A é a classe mais provável. A probabilidade relativa de A será = 3.6269

III

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 devido à sua capacidade de capturar padrões complexos em dados de alta dimensão, enquanto Gaussian Naive Bayes tem melhor desempenho no conjunto de dados Wine, porque os atributos do conjunto são condicionalmente independentes.