Group 37 – Enzo Nunes 106336, Tomás Barros 106588

Machine Learning – Homework2

Ex1 – Bayesian Classifier

a)

1. a)

$$p(x_1 \mid c=A) =$$

$$\mu = \frac{0.6+1+1.6+1.8}{4} = \boxed{\frac{5}{4}}$$

$$\sigma = \left(\frac{1}{4-1}\left(\left(0.6-\frac{5}{4}\right)^2+\left(1-\frac{5}{4}\right)^2+\left(1.6-\frac{5}{4}\right)^2+\left(1.8-\frac{5}{4}\right)^2\right)\right)^{\frac{1}{2}} =$$

$$= \boxed{\frac{\sqrt{273}}{30}}$$

$$p(x_1 \mid c=B)$$

$$\mu = \frac{2+2+3+4}{4} = \boxed{\frac{11}{4}}$$

$$\sigma = \left(\frac{1}{4-1}\left(\left(2-\frac{11}{4}\right)^2+\left(2-\frac{11}{4}\right)^2+\left(3-\frac{11}{4}\right)^2+\left(4-\frac{11}{4}\right)^2\right)\right)^{\frac{1}{2}}$$

$$= \boxed{\frac{\sqrt{33}}{6}}$$

$$p(x_2 \mid c=A) =$$

$$\mu = \frac{0.4+1.1+1.5+1.8}{4} = \boxed{\frac{6}{5}}$$

$$\sigma = \left(\frac{1}{4-1}\left(\left(0.4-\frac{6}{5}\right)^2+\left(1.1-\frac{6}{5}\right)^2+\left(1.5-\frac{6}{5}\right)^2+\left(1.8-\frac{6}{5}\right)^2\right)\right)^{\frac{1}{2}} =$$

$$= \boxed{\frac{\sqrt{330}}{30}}$$

$$p(x_2 \mid c=B) =$$

$$\mu = \frac{0+1+0+1.2}{4} = \boxed{\frac{11}{20}}$$

$$\sigma \left(\frac{1}{4-1}\left(\left(0-\frac{11}{20}\right)^2+\left(1-\frac{11}{20}\right)^2+\left(0-\frac{11}{20}\right)^2+\left(1.2-\frac{11}{20}\right)^2\right)\right)^{\frac{1}{2}} =$$

$$= \boxed{\frac{\sqrt{41}}{10}}$$

$$P(H \mid D) = \frac{P(D \mid H)\,p(H)}{P(D)} \rightarrow \text{Bayes rule}$$

$$P(x_1, x_2 \mid C) = P(x_1 \mid C)P(x_2 \mid C) \rightarrow \text{Naive Bayes Assumption}$$

$$p(c=A \mid x_1=1, x_2=2) = \frac{p(c=A)\,p(x_1=1, x_2=2 \mid c=A)}{p(x_1=1, x_2=2)} =$$

$$= \frac{p(c=A)\,p(x_1=1 \mid c=A)\,p(x_2=2 \mid c=A)}{P(x_1=1, x_2=2)} = \frac{\frac{1}{2} N\left(1 \mid \mu=\frac{5}{4}, \sigma=\frac{\sqrt{273}}{30}\right) N\left(2 \mid \mu=\frac{6}{5}, \sigma=\frac{\sqrt{330}}{30}\right)}{P(x_1=1, x_2=2)}$$

$$= \frac{\frac{1}{2} \times 0.65344 \times 0.27527}{P(x_1=1, x_2=2)} = \frac{0.08994}{P(x_1=1, x_2=2)}$$

$$P(C=B|x_1=1,x_2=2) = \frac{P(C=B)P(x_1=1,x_2=2|C=B)}{P(x_1=1,x_2=2)} =$$

$$= \frac{P(C=B)\,P(x_1=1|C=B)\,P(x_2=2|C=B)}{P(x_1=1,x_2=2)} = \frac{\frac{1}{2} N\left(1 \middle| \mu=\frac{11}{4}, \sigma=\frac{\sqrt{33}}{6}\right) N\left(2 \middle| \mu=\frac{11}{20}, \sigma=\frac{\sqrt{41}}{10}\right)}{P(x_1=1,x_2=2)}$$

$$= \frac{\frac{1}{2} \times 0.07840 \times 0.04797}{P(x_1=1,x_2=2)} = \frac{1.8804 \times 10^{-3}}{P(x_1=1,x_2=2)}$$

$$P(C=A|x_1=1,x_2=2) > P(C=B|x_1=1,x_2=2)$$

The most probable class for the query vector, under the Naive Bayes assumption, using 1-dimensional Gaussians to model the likelihood, is <u>Class A</u>.

b)

1. b) $p(x_1, x_2 | C=A)$

$$\mu = \begin{bmatrix} \frac{0,6+1+1,6+1,8}{4} \\ \frac{0,4+1,1+1,5+1,8}{4} \end{bmatrix} = \begin{bmatrix} \frac{5}{4} \\ \frac{6}{5} \end{bmatrix} \quad \Sigma_{ab} = \frac{1}{N-1} \sum_{i=1}^{m} (x_{ia} - \bar{x}_a)$$

$$\Sigma_{00} = \frac{1}{4-1}\left[\left(0,6-\frac{5}{4}\right)\left(0,6-\frac{5}{4}\right) + \left(1-\frac{5}{4}\right)\left(1-\frac{5}{4}\right) + \left(1,6-\frac{5}{4}\right)\left(1,6-\frac{5}{4}\right) + \left(1,8-\frac{5}{4}\right)\left(1,8-\frac{5}{4}\right)\right] = \frac{91}{300}$$

$$\Sigma_{01} = \Sigma_{10} = \frac{1}{4-1}\left[\left(0,6-\frac{5}{4}\right)\left(0,4-\frac{6}{5}\right) + \left(1-\frac{5}{4}\right)\left(1,1-\frac{6}{5}\right) + \left(1,6-\frac{5}{4}\right)\left(1,5-\frac{6}{5}\right) + \left(1,8-\frac{5}{4}\right)\left(1,8-\frac{6}{5}\right)\right] = \frac{49}{150}$$

$$\Sigma_{11} = \frac{1}{4-1}\left[\left(0,4-\frac{6}{5}\right)\left(0,4-\frac{6}{5}\right) + \left(1,1-\frac{6}{5}\right)\left(1,1-\frac{6}{5}\right) + \left(1,5-\frac{6}{5}\right)\left(1,5-\frac{6}{5}\right) + \left(1,8-\frac{6}{5}\right)\left(1,8-\frac{6}{5}\right)\right] = \frac{11}{30}$$

$$\Sigma = \begin{bmatrix} \frac{91}{300} & \frac{49}{150} \\ \frac{49}{150} & \frac{11}{30} \end{bmatrix} \quad det(\Sigma) = \left(\frac{91}{300} \times \frac{11}{30}\right) - \left(\frac{49}{150} \times \frac{49}{150}\right) = \frac{203}{45000} \quad \Sigma^{-1} = \frac{45000}{203}\begin{bmatrix} \frac{11}{30} & -\frac{49}{150} \\ -\frac{49}{150} & \frac{91}{300} \end{bmatrix} = \begin{bmatrix} \frac{16500}{203} & -\frac{2100}{29} \\ -\frac{2100}{29} & \frac{1950}{29} \end{bmatrix}$$

$p(x_1, x_2 | C=B)$

$$\mu = \begin{bmatrix} \frac{2+2+3+4}{4} \\ \frac{0+1+0+1,2}{4} \end{bmatrix} = \begin{bmatrix} \frac{11}{4} \\ \frac{11}{20} \end{bmatrix}$$

$$\Sigma_{00} = \frac{1}{4-1}\left[\left(2-\frac{11}{4}\right)\left(2-\frac{11}{4}\right) + \left(2-\frac{11}{4}\right)\left(2-\frac{11}{4}\right) + \left(3-\frac{11}{4}\right)\left(3-\frac{11}{4}\right) + \left(4-\frac{11}{4}\right)\left(4-\frac{11}{4}\right)\right] = \frac{11}{12}$$

$$\Sigma_{01} = \Sigma_{10} = \frac{1}{4-1}\left[\left(2-\frac{11}{4}\right)\left(0-\frac{11}{20}\right) + \left(2-\frac{11}{4}\right)\left(1-\frac{11}{20}\right) + \left(3-\frac{11}{4}\right)\left(0-\frac{11}{20}\right) + \left(4-\frac{11}{4}\right)\left(1,2-\frac{11}{20}\right)\right] = \frac{1}{4}$$

$$\Sigma_{11} = \frac{1}{4-1}\left[\left(0-\frac{11}{20}\right)\left(0-\frac{11}{20}\right) + \left(1-\frac{11}{20}\right)\left(1-\frac{11}{20}\right) + \left(0-\frac{11}{20}\right)\left(0-\frac{11}{20}\right) + \left(1,2-\frac{11}{20}\right)\left(1,2-\frac{11}{20}\right)\right] = \frac{41}{100}$$

$$\Sigma = \begin{bmatrix} \frac{11}{12} & \frac{1}{4} \\ \frac{1}{4} & \frac{41}{100} \end{bmatrix} \quad det(\Sigma) = \frac{11}{12} \times \frac{41}{100} - \left(\frac{1}{4} \times \frac{1}{4}\right) = \frac{47}{150} \quad \Sigma^{-1} = \frac{150}{47}\begin{bmatrix} \frac{41}{100} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{11}{12} \end{bmatrix} = \begin{bmatrix} \frac{123}{94} & -\frac{75}{94} \\ -\frac{75}{94} & \frac{275}{94} \end{bmatrix}$$

$$P(C=A \mid X_1=1, X_2=2) =$$

$$\frac{P(C=A)\,P(X_1=1, X_2=2 \mid C=A)}{P(X_1=1, X_2=2)} = \frac{\frac{1}{2} N\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} \middle| \mu = \begin{bmatrix} \frac{5}{4} \\ \frac{6}{5} \end{bmatrix}, \Sigma = \begin{bmatrix} \frac{91}{300} & \frac{49}{150} \\ \frac{49}{150} & \frac{11}{30} \end{bmatrix}\right)}{P(X_1=1, X_2=2)} = *$$

$$N\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} \middle| \mu = \begin{bmatrix} \frac{5}{4} \\ \frac{6}{5} \end{bmatrix}, \Sigma = \begin{bmatrix} \frac{91}{300} & \frac{49}{150} \\ \frac{49}{150} & \frac{11}{30} \end{bmatrix}\right) = \frac{1}{(2\pi)^{\frac{2}{2}} \sqrt{\frac{203}{45000}}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} \frac{5}{4} \\ \frac{6}{5} \end{bmatrix}\right)^T \begin{bmatrix} \frac{16500}{203} & -\frac{2100}{29} \\ -\frac{2100}{29} & \frac{1950}{29} \end{bmatrix} \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} \frac{5}{4} \\ \frac{6}{5} \end{bmatrix}\right)\right)$$

$$* = \frac{\frac{1}{2} \times 4.334693 \times 10^{-17}}{P(X_1=1, X_2=2)}$$

$$P(C=B \mid X_1=1, X_2=2) =$$

$$\frac{P(C=B)\,P(X_1=1, X_2=2 \mid C=B)}{P(X_1=1, X_2=2)} = \frac{\frac{1}{2} N\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} \middle| \mu = \begin{bmatrix} \frac{11}{4} \\ \frac{11}{20} \end{bmatrix}, \Sigma = \begin{bmatrix} \frac{11}{12} & \frac{1}{4} \\ \frac{1}{4} & \frac{41}{100} \end{bmatrix}\right)}{P(X_1=1, X_2=2)} **$$

$$N\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} \middle| \mu = \begin{bmatrix} \frac{11}{4} \\ \frac{11}{20} \end{bmatrix}, \Sigma = \begin{bmatrix} \frac{11}{12} & \frac{1}{4} \\ \frac{1}{4} & \frac{41}{100} \end{bmatrix}\right) = \frac{1}{(2\pi)^{\frac{2}{2}} \sqrt{\frac{47}{150}}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} \frac{11}{4} \\ \frac{11}{20} \end{bmatrix}\right)^T \begin{bmatrix} \frac{123}{94} & -\frac{75}{94} \\ -\frac{75}{94} & \frac{275}{94} \end{bmatrix} \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} \frac{11}{4} \\ \frac{11}{20} \end{bmatrix}\right)\right)$$

$$** \simeq \frac{\frac{1}{2} \times 0.000234}{P(X_1=1, X_2=2)}$$

$$P(C=A \mid X_1=1, X_2=2) < P(C=B \mid X_1=1, X_2=2)$$

that the likelihoods are 2-dimensional Gaussian

The most probable class for the query vector, assuming that the likelihoods are 2-dimensional Gaussians, is <u>Class B</u>. Since the results are different, we can deduce that there is a dependence between $X_1$and X_2. Given that, we can conclude that the Naive Bayes assumption does not make sense in this context.

c)

Calculate the count for each class:

$$Count(A) = Count(B) = 4$$

Calculate the prior probability for each class:

$$P(A) = P(B) = \frac{4}{8} = \frac{1}{2}$$

Calculate the likelihoods for each class (where $x_3 = 1$):

$$P(X_3 = 1|A) = \frac{2}{4} = \frac{1}{2}$$

$$P(X_3 = 1|B) = \frac{3}{4}$$

Calculate the posterior probability for each class using the Bayes' theorem:

$$P(A|X_3 = 1) = \frac{P(X_3 = 1|A) \cdot P(A)}{P(X_3 = 1)}$$

$$P(B|X_3 = 1) = \frac{P(X_3 = 1|B) \cdot P(B)}{P(X_3 = 1)}$$

Calculate the ratio:

$$\frac{P(A|X_3 = 1)}{P(B|X_3 = 1)} = \frac{\frac{\frac{2}{4} \times \frac{4}{8}}{P(X_3 = 1)}}{\frac{\frac{3}{4} \times \frac{4}{8}}{P(X_3 = 1)}} = \frac{0.5}{0.75} = \frac{2}{3} \Rightarrow P(A|X_3 = 1) < (B|X_3 = 1)$$

We can conclude that the most probable class for this query vector is <u>Class B</u>.

d)

From exercise **b)**:

$$P\big((1,2)\big|A\big) \cong 4.33 \times 10^{-17}$$

$$P\big((1,2)\big|B\big) = 2.34 \times 10^{-4}$$

From exercise **c)**:

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

$$P(1|B) = \frac{3}{4}$$

From the prior probabilities:

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

Calculate the posterior probabilities:

$$P\big(A\big|X_{query}\big) = P\big((1,2)\big|A\big) \cdot P(1|A) \cdot P(A) = 4.33 \times 10^{-17} \times \frac{1}{2} \times \frac{1}{2} = 1.0825 \times 10^{-17}$$

$$P\big(B\big|X_{query}\big) = P\big((1,2)\big|B\big) \cdot P(1|B) \cdot P(B) = 2.34 \times 10^{-4} \times \frac{3}{4} \times \frac{1}{2} = 8.775 \times 10^{-5}$$

Calculate the ratio:

$$\frac{P\big(A\big|X_{query}\big)}{P\big(B\big|X_{query}\big)} = \frac{1.0825 \times 10^{-17}}{8.775 \times 10^{-5}} \cong 1.2336 \times 10^{-13} \Rightarrow P\big(A\big|X_{query}\big) \ll P\big(B\big|X_{query}\big)$$

The relative probability between both classes is 1.2336×10^{-13}. Meaning, Class B is $\frac{1}{1.2336 \times 10^{-13}} = 8.11 \times 10^{12}$ times more likely than Class A. We can conclude that <u>Class B</u> is overwhelmingly more probable than Class A.

Ex2 – Software Experiments

With the digits dataset:

- kNN accuracy:
  - $0.98,\ k = 3$
  - $0.95,\ k = 30$
- Gaussian accuracy:
  - 0.85

With the Wine dataset:

- kNN accuracy:
  - $0.70,\ k = 3$
  - $0.71,\ k = 30$
- Gaussian accuracy:
  - 0.94

For the digits dataset, the kNN classifier performs best, with an accuracy of 0.98 for k=3. This suggests that kNN is better suited for this dataset, likely because it captures the fine distinctions between different handwritten digit patterns. For the Wine dataset, Gaussian Naive Bayes gives the best result, with an accuracy of 0.94. This implies that the wine dataset aligns better with the assumptions of Gaussian Naive Bayes, such as the features being approximately Gaussian-distributed and independent.