Aprendizagem 2024
Homework II – Group 053
(ist106930, ist107059)

Part I: Pen and paper

1. F1-Measure

Iteration 1: Leave out $x_1 = (A, 0)$

Distances:

$$x_2 = (B, 1) : \text{Hamming distance} = 2$$
$$x_3 = (A, 1) : \text{Hamming distance} = 1$$
$$x_4 = (A, 0) : \text{Hamming distance} = 0$$
$$x_5 = (B, 0) : \text{Hamming distance} = 1$$
$$x_6 = (B, 0) : \text{Hamming distance} = 1$$
$$x_7 = (A, 1) : \text{Hamming distance} = 1$$
$$x_8 = (B, 1) : \text{Hamming distance} = 2$$

The five nearest neighbors are x_4, x_3, x_5, x_6, x_7.

Positives: x_3, x_4

Negatives: x_5, x_6, x_7

Result: The majority is Negative, so x_1 is misclassified as Negative.

Iteration 2: Leave out $x_2 = (B, 1)$

Distances:

$$x_1 = (A, 0) : \text{Hamming distance} = 2$$
$$x_3 = (A, 1) : \text{Hamming distance} = 1$$
$$x_4 = (A, 0) : \text{Hamming distance} = 2$$
$$x_5 = (B, 0) : \text{Hamming distance} = 1$$
$$x_6 = (B, 0) : \text{Hamming distance} = 1$$
$$x_7 = (A, 1) : \text{Hamming distance} = 1$$
$$x_8 = (B, 1) : \text{Hamming distance} = 0$$

The five nearest neighbors are x_8, x_3, x_5, x_6, x_7.

Positives: x_3

Negatives: x_5, x_6, x_7, x_8

Result: The majority is Negative, so x_2 is misclassified as Negative.

Iteration 3: Leave out $x_3 = (A, 1)$

Distances:

$$x_1 = (A, 0) : \text{Hamming distance} = 1$$
$$x_2 = (B, 1) : \text{Hamming distance} = 1$$
$$x_4 = (A, 0) : \text{Hamming distance} = 1$$
$$x_5 = (B, 0) : \text{Hamming distance} = 2$$
$$x_6 = (B, 0) : \text{Hamming distance} = 2$$
$$x_7 = (A, 1) : \text{Hamming distance} = 0$$
$$x_8 = (B, 1) : \text{Hamming distance} = 1$$

The five nearest neighbors are x_7, x_1, x_2, x_4, x_8.

Positives: x_1, x_2, x_4

Negatives: x_7, x_8

Result: The majority is Positive, so x_3 is correctly classified as Positive.

Iteration 4: Leave out $x_4 = (A, 0)$

Distances:

$$x_1 = (A, 0) : \text{Hamming distance} = 0$$
$$x_2 = (B, 1) : \text{Hamming distance} = 2$$
$$x_3 = (A, 1) : \text{Hamming distance} = 1$$
$$x_5 = (B, 0) : \text{Hamming distance} = 1$$
$$x_6 = (B, 0) : \text{Hamming distance} = 1$$
$$x_7 = (A, 1) : \text{Hamming distance} = 1$$
$$x_8 = (B, 1) : \text{Hamming distance} = 2$$

The five nearest neighbors are x_1, x_3, x_5, x_6, x_7.

Positives: x_1, x_3

Negatives: x_5, x_6, x_7

Result: The majority is Negative, so x_4 is misclassified as Negative.

Iteration 5: Leave out $x_5 = (B, 0)$

Distances:

$$\begin{aligned}
x_1 &= (A, 0) : \text{Hamming distance} = 1 \\
x_2 &= (B, 1) : \text{Hamming distance} = 1 \\
x_3 &= (A, 1) : \text{Hamming distance} = 2 \\
x_4 &= (A, 0) : \text{Hamming distance} = 1 \\
x_6 &= (B, 0) : \text{Hamming distance} = 0 \\
x_7 &= (A, 1) : \text{Hamming distance} = 2 \\
x_8 &= (B, 1) : \text{Hamming distance} = 1
\end{aligned}$$

The five nearest neighbors are x_6, x_1, x_2, x_4, x_8.

Positives: x_1, x_2, x_4

Negatives: x_6, x_8

Result: The majority is Positive, so x_5 is misclassified as Positive.

Iteration 6: Leave out $x_6 = (B, 0)$

Distances:

$$\begin{aligned}
x_1 &= (A, 0) : \text{Hamming distance} = 1 \\
x_2 &= (B, 1) : \text{Hamming distance} = 1 \\
x_3 &= (A, 1) : \text{Hamming distance} = 2 \\
x_4 &= (A, 0) : \text{Hamming distance} = 1 \\
x_5 &= (B, 0) : \text{Hamming distance} = 0 \\
x_7 &= (A, 1) : \text{Hamming distance} = 2 \\
x_8 &= (B, 1) : \text{Hamming distance} = 1
\end{aligned}$$

The five nearest neighbors are x_5, x_1, x_2, x_4, x_8.

Positives: x_1, x_2, x_4

Negatives: x_5, x_8

Result: The majority is Positive, so x_6 is misclassified as Positive.

Iteration 7: Leave out $x_7 = (A, 1)$

Distances:

$$x_1 = (A, 0) : \text{Hamming distance} = 1$$
$$x_2 = (B, 1) : \text{Hamming distance} = 1$$
$$x_3 = (A, 1) : \text{Hamming distance} = 0$$
$$x_4 = (A, 0) : \text{Hamming distance} = 1$$
$$x_5 = (B, 0) : \text{Hamming distance} = 2$$
$$x_6 = (B, 0) : \text{Hamming distance} = 2$$
$$x_8 = (B, 1) : \text{Hamming distance} = 1$$

The five nearest neighbors are x_3, x_1, x_2, x_4, x_8.

Positives: x_1, x_2, x_3, x_4

Negatives: x_8

Result: The majority is Positive, so x_7 is misclassified as Positive.

Iteration 8: Leave out $x_8 = (B, 1)$

Distances:

$$x_1 = (A, 0) : \text{Hamming distance} = 2$$
$$x_2 = (B, 1) : \text{Hamming distance} = 0$$
$$x_3 = (A, 1) : \text{Hamming distance} = 1$$
$$x_4 = (A, 0) : \text{Hamming distance} = 2$$
$$x_5 = (B, 0) : \text{Hamming distance} = 1$$
$$x_6 = (B, 0) : \text{Hamming distance} = 1$$
$$x_7 = (A, 1) : \text{Hamming distance} = 1$$

The five nearest neighbors are x_2, x_3, x_5, x_6, x_7.

Positives: x_2, x_3

Negatives: x_5, x_6, x_7

Result: The majority is Negative, so x_8 is correctly classified as Negative.

Summary:
True Positives (TP): x_2
True Negatives (TN): x_8
False Positives (FP): x_5, x_6, x_7
False Negatives (FN): x_1, x_3, x_4

Precision:

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{1}{1+3} = \frac{1}{4} = 0.25$$

Recall:

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{1}{1+3} = \frac{1}{4} = 0.25$$

F1-Measure:

$$\text{F1-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times 0.25 \times 0.25}{0.25 + 0.25} = \frac{0.125}{0.50} = \boxed{0.25}$$

2. New metric

Our new metric is:

$$d(x_i, x_j) = 1 \times d_1(x_i, x_j) + 0.1 \times d_2(x_i, x_j)$$

We also altered **k to 3** (on exercise 1 k was 5).

Iteration 1: Leave out $x_1 = (A, 0)$

Distances:

$x_2 = (B, 1)$: Hamming distance = 1.1
$x_3 = (A, 1)$: Hamming distance = 0.1
$x_4 = (A, 0)$: Hamming distance = 0
$x_5 = (B, 0)$: Hamming distance = 1
$x_6 = (B, 0)$: Hamming distance = 1
$x_7 = (A, 1)$: Hamming distance = 0.1
$x_8 = (B, 1)$: Hamming distance = 1.1

The three nearest neighbors are x_4, x_3, x_7.

Positives: x_3, x_4

Negatives: x_7

Result: The majority is Positive, so x_1 is correctly classified as Positive.

Iteration 2: Leave out $x_2 = (B, 1)$

Distances:

$$
\begin{aligned}
x_1 &= (A, 0) : \text{Hamming distance} = 1.1 \\
x_3 &= (A, 1) : \text{Hamming distance} = 1 \\
x_4 &= (A, 0) : \text{Hamming distance} = 1.1 \\
x_5 &= (B, 0) : \text{Hamming distance} = 0.1 \\
x_6 &= (B, 0) : \text{Hamming distance} = 0.1 \\
x_7 &= (A, 1) : \text{Hamming distance} = 1 \\
x_8 &= (B, 1) : \text{Hamming distance} = 0
\end{aligned}
$$

The three nearest neighbors are x_8, x_5, x_6.

Positives: None

Negatives: x_5, x_6, x_8

Result: The majority is Negative, so x_2 is misclassified as Negative.

Iteration 3: Leave out $x_3 = (A, 1)$

Distances:

$$
\begin{aligned}
x_1 &= (A, 0) : \text{Hamming distance} = 0.1 \\
x_2 &= (B, 1) : \text{Hamming distance} = 1 \\
x_4 &= (A, 0) : \text{Hamming distance} = 0.1 \\
x_5 &= (B, 0) : \text{Hamming distance} = 1.1 \\
x_6 &= (B, 0) : \text{Hamming distance} = 1.1 \\
x_7 &= (A, 1) : \text{Hamming distance} = 0 \\
x_8 &= (B, 1) : \text{Hamming distance} = 1
\end{aligned}
$$

The three nearest neighbors are x_7, x_1, x_4.

Positives: x_1, x_4

Negatives: x_7

Result: The majority is Positive, so x_3 is correctly classified as Positive.

Iteration 4: Leave out $x_4 = (A, 0)$

Distances:

$$\begin{aligned}
x_1 &= (A, 0) : \text{Hamming distance} = 0 \\
x_2 &= (B, 1) : \text{Hamming distance} = 1.1 \\
x_3 &= (A, 1) : \text{Hamming distance} = 0.1 \\
x_5 &= (B, 0) : \text{Hamming distance} = 1 \\
x_6 &= (B, 0) : \text{Hamming distance} = 1 \\
x_7 &= (A, 1) : \text{Hamming distance} = 0.1 \\
x_8 &= (B, 1) : \text{Hamming distance} = 1.1
\end{aligned}$$

The three nearest neighbors are x_1, x_3, x_7.

Positives: x_1, x_3

Negatives: x_7

Result: The majority is Positive, so x_4 is correctly classified as Positive.

Iteration 5: Leave out $x_5 = (B, 0)$

Distances:

$$\begin{aligned}
x_1 &= (A, 0) : \text{Hamming distance} = 1 \\
x_2 &= (B, 1) : \text{Hamming distance} = 0.1 \\
x_3 &= (A, 1) : \text{Hamming distance} = 1.1 \\
x_4 &= (A, 0) : \text{Hamming distance} = 1 \\
x_6 &= (B, 0) : \text{Hamming distance} = 0 \\
x_7 &= (A, 1) : \text{Hamming distance} = 1.1 \\
x_8 &= (B, 1) : \text{Hamming distance} = 0.1
\end{aligned}$$

The three nearest neighbors are x_6, x_2, x_8.

Positives: x_2

Negatives: x_6, x_8

Result: The majority is Negative, so x_5 is correctly classified as Negative.

Iteration 6: Leave out $x_6 = (B, 0)$

Distances:

$$\begin{aligned}
x_1 &= (A, 0) : \text{Hamming distance} = 1 \\
x_2 &= (B, 1) : \text{Hamming distance} = 0.1 \\
x_3 &= (A, 1) : \text{Hamming distance} = 1.1 \\
x_4 &= (A, 0) : \text{Hamming distance} = 1 \\
x_5 &= (B, 0) : \text{Hamming distance} = 0 \\
x_7 &= (A, 1) : \text{Hamming distance} = 1.1 \\
x_8 &= (B, 1) : \text{Hamming distance} = 0.1
\end{aligned}$$

The three nearest neighbors are x_5, x_2, x_8.

Positives: x_2

Negatives: x_5, x_8

Result: The majority is Negative, so x_6 is correctly classified as Negative.

Iteration 7: Leave out $x_7 = (A, 1)$

Distances:

$$\begin{aligned}
x_1 &= (A, 0) : \text{Hamming distance} = 0.1 \\
x_2 &= (B, 1) : \text{Hamming distance} = 1 \\
x_3 &= (A, 1) : \text{Hamming distance} = 0 \\
x_4 &= (A, 0) : \text{Hamming distance} = 0.1 \\
x_5 &= (B, 0) : \text{Hamming distance} = 1.1 \\
x_6 &= (B, 0) : \text{Hamming distance} = 1.1 \\
x_8 &= (B, 1) : \text{Hamming distance} = 1
\end{aligned}$$

The three nearest neighbors are x_3, x_1, x_4.

Positives: x_1, x_3, x_4

Negatives: None

Result: The majority is Positive, so x_7 is misclassified as Positive.

Iteration 8: Leave out $x_8 = (B, 1)$

Distances:

$$
\begin{aligned}
x_1 &= (A, 0) : \text{Hamming distance} = 1.1 \\
x_2 &= (B, 1) : \text{Hamming distance} = 0 \\
x_3 &= (A, 1) : \text{Hamming distance} = 1 \\
x_4 &= (A, 0) : \text{Hamming distance} = 1.1 \\
x_5 &= (B, 0) : \text{Hamming distance} = 0.1 \\
x_6 &= (B, 0) : \text{Hamming distance} = 0.1 \\
x_7 &= (A, 1) : \text{Hamming distance} = 1
\end{aligned}
$$

The three nearest neighbors are x_2, x_5, x_6.

Positives: x_2

Negatives: x_5, x_6

Result: The majority is Negative, so x_8 is correctly classified as Negative.

Summary:
True Positives (TP): x_1, x_3, x_4
True Negatives (TN): x_5, x_6, x_8
False Positives (FP): x_7
False Negatives (FN): x_2

Precision:

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{3}{1 + 3} = \frac{3}{4} = 0.75$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{3}{1 + 3} = \frac{3}{4} = 0.75$$

F1-Measure:

$$\text{F1-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times 0.75 \times 0.75}{0.75 + 0.75} = \frac{1.125}{1.50} = \boxed{0.75}$$

3. Bayesian classifier

   To calculate the prior probabilities $P(P)$ and $P(N)$:

   Total observations = 9
   Number of positive observations = 5
   Number of negative observations = 4

$$P(P) = \frac{5}{9} \qquad P(N) = \frac{4}{9}$$

   For positive observations P:

   Values of y_3: $\{1.1, 0.8, 0.5, 0.9, 0.8\}$
   Mean μ_P:

$$\mu_P = \frac{1.1 + 0.8 + 0.5 + 0.9 + 0.8}{5} = \frac{4.1}{5} = 0.82$$

   Variance σ_P^2:

$$\sigma_P^2 = \frac{(1.1 - 0.82)^2 + (0.8 - 0.82)^2 + (0.5 - 0.82)^2 + (0.9 - 0.82)^2 + (0.8 - 0.82)^2}{4}$$

$$= \frac{0.188}{4} \quad = 0.047$$

   For negative observations N:

   Values of y_3: $\{1, 0.9, 1.2, 0.9\}$
   Mean μ_N:

$$\mu_N = \frac{1 + 0.9 + 1.2 + 0.9}{4} = \frac{4}{4} = 1.0$$

   Variance σ_N^2:

$$\sigma_N^2 = \frac{(1 - 1.0)^2 + (0.9 - 1.0)^2 + (1.2 - 1.0)^2 + (0.9 - 1.0)^2}{3}$$

$$= \frac{0.06}{3} \quad = 0.02$$

   Summary

   Prior Probabilities:

$$P(P) = \frac{5}{9}, \quad P(N) = \frac{4}{9}$$

Conditional Probabilities for y_3:

For P:

$$\mu_P = 0.82, \quad \sigma_P^2 = 0.047$$

For N:

$$\mu_N = 1.0, \quad \sigma_N^2 = 0.02$$

Conclusion

These parameters allow us to construct a Bayesian classifier based on the values of y_3 and the class labels P and N. This classifier will use the calculated means and variances for each class, along with the prior probabilities, to classify new observations based on their likelihood of belonging to each class.

4. MAP assumption

$$P(y_3 \mid P) = \frac{1}{\sqrt{2\pi\sigma_P^2}} \exp\left(-\frac{(y_3 - \mu_P)^2}{2\sigma_P^2}\right) \qquad P(y_3 \mid N) = \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left(-\frac{(y_3 - \mu_N)^2}{2\sigma_N^2}\right)$$

<u>Observation 1: $(A, 1, 0.8)$</u>

For P:

$$P(0.8 \mid P) = \frac{1}{\sqrt{2\pi \cdot 0.047}} \exp\left(-\frac{(0.8 - 0.82)^2}{2 \cdot 0.047}\right) \times \frac{1}{5} = 0.366$$

For N:

$$P(0.8 \mid N) = \frac{1}{\sqrt{2\pi \cdot 0.02}} \exp\left(-\frac{(0.8 - 1.0)^2}{2 \cdot 0.02}\right) \times \frac{1}{4} = 0.260$$

$$P(P \mid (A, 1, 0.8)) = P(0.8 \mid P) \times P(P) \approx 0.366 \cdot \frac{5}{9} \approx 0.203$$

$$P(N \mid (A, 1, 0.8)) = P(0.8 \mid N) \times P(N) \approx 0.260 \cdot \frac{4}{9} \approx 0.116$$

Classify: Since $P(P \mid (A, 1, 0.8)) > P(N \mid (A, 1, 0.8))$, the class is **Positive**.

Observation 2: $(B, 1, 1)$

For P:

$$P(1 \mid P) = \frac{1}{\sqrt{2\pi \cdot 0.047}} \exp\left(-\frac{(1-0.82)^2}{2 \cdot 0.047}\right) \times \frac{1}{5} = 0.26046$$

For N:

$$P(1 \mid N) = \frac{1}{\sqrt{2\pi \cdot 0.02}} \exp(0) \times \frac{1}{4} = 0.7074$$

$$P(P \mid (B, 1, 1)) = P(1 \mid P) \times P(P) \approx 0.26046 \cdot \frac{5}{9} \approx 0.145$$

$$P(N \mid (B, 1, 1)) = P(1 \mid N) \times P(N) \approx 0.7074 \cdot \frac{4}{9} \approx 0.314$$

Classify: Since $P(N \mid (B, 1, 1)) > P(P \mid (B, 1, 1))$, the class is **Negative**.

Observation 3: $(B, 0, 0.9)$

For P:

$$P(0.9 \mid P) = \frac{1}{\sqrt{2\pi \cdot 0.047}} \exp\left(-\frac{(0.9-0.82)^2}{2 \cdot 0.047}\right) \times \frac{1}{5} = 0.34344$$

For N:

$$P(0.9 \mid N) = \frac{1}{\sqrt{2\pi \cdot 0.02}} \exp\left(-\frac{(0.9-1.0)^2}{2 \cdot 0.02}\right) \times \frac{1}{4} = 1.1018$$

$$P(P \mid (B, 0, 0.9)) = P(0.9 \mid P) \times P(P) \approx 0.34344 \cdot \frac{5}{9} \approx 0.191$$

$$P(N \mid (B, 0, 0.9)) = P(0.9 \mid N) \times P(N) \approx 1.1018 \cdot \frac{4}{9} \approx 0.490$$

Classify: Since $P(N \mid (B, 0, 0.9)) > P(P \mid (B, 0, 0.9))$, the class is **Negative**.

5. Classifying 'i like to run'

   Vocabulary size (V) = 8
   N_P = 5 (amazing, run, i, like, it) N_N = 4 (too, tired, bad, run)
   4 sentences (2 pos, 2 neg)

$$P(P) = \frac{1}{2} \qquad P(N) = \frac{1}{2}$$

   Now, using the formula given, we will calculate the likehoods of the words in the sentence 'i like to run':

Positive Class

"I":

$$p(i \mid P) = \frac{1+1}{5+8} = \frac{2}{13}$$

"like":

$$p(like \mid P) = \frac{1+1}{5+8} = \frac{2}{13}$$

"to":

$$p(to \mid P) = \frac{0+1}{5+8} = \frac{1}{13}$$

"run":

$$p(run \mid P) = \frac{1+1}{5+8} = \frac{2}{13}$$

Negative Class

"I":

$$p(i \mid P) = \frac{0+1}{4+8} = \frac{1}{12}$$

"like":

$$p(like \mid P) = \frac{0+1}{4+8} = \frac{1}{12}$$

"to":

$$p(to \mid P) = \frac{0+1}{4+8} = \frac{1}{12}$$

"run":

$$p(run \mid P) = \frac{1+1}{4+8} = \frac{2}{12}$$

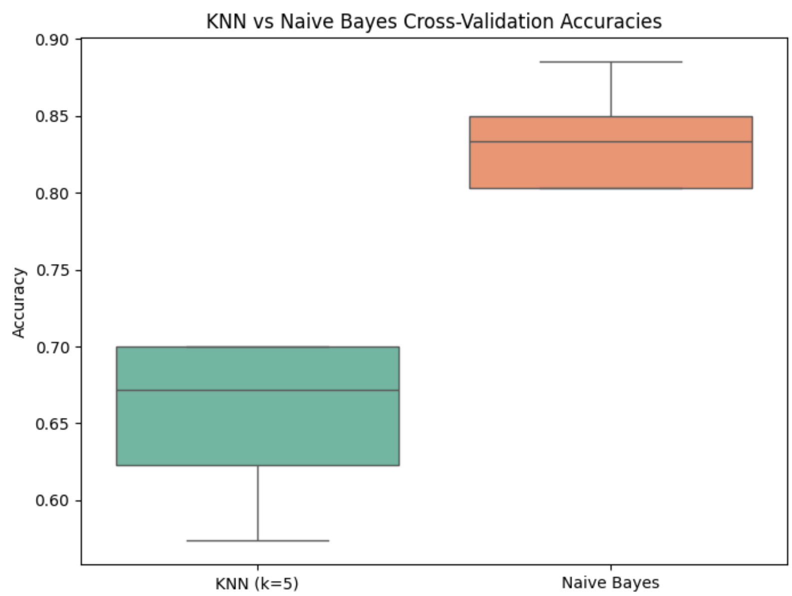
$$P(P \mid \text{"}I \text{ like } to \text{ run"}) = 0.5 \times \frac{2}{13} \times \frac{2}{13} \times \frac{1}{13} \times \frac{2}{13} \approx 0.00014$$

$$P(N \mid \text{"}I \text{ like } to \text{ run"}) = 0.5 \times \frac{1}{12} \times \frac{1}{12} \times \frac{1}{12} \times \frac{2}{12} \approx 0.000048$$

R:Since $P(P \mid \text{"}I \text{ like } to \text{ run"}) > P(N \mid \text{"}I \text{ like } to \text{ run"})$, we classify the sentence "I like to run" as Positive.

Part II: Programming

1. Graphic for exercise 1.
   alínea a):



KNN accuracies:

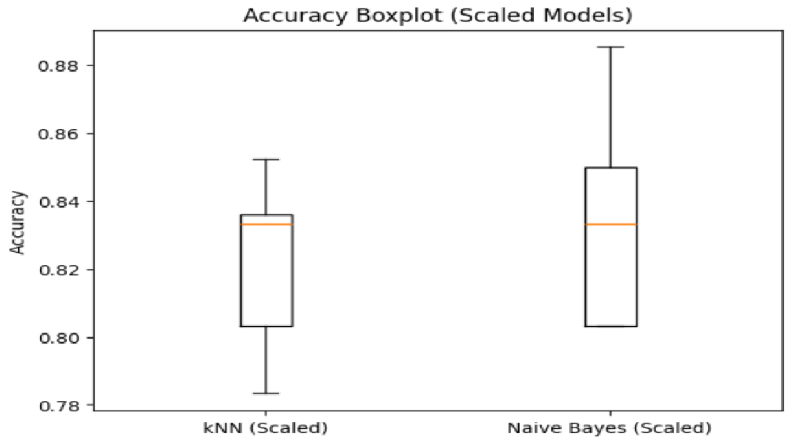
$$[0.62295082, 0.57377049, 0.67213115, 0.7, 0.7]$$

Naive Bayes accuracies:

$$[0.8852459, 0.80327869, 0.80327869, 0.85, 0.83333333]$$

Gaussian Naive Bayes (GNB) shows a higher average accuracy compared to k-Nearest Neighbors (kNN) with $k = 5$. GNB appears more stable as the interquartile range is much narrower compared to kNN, which indicates less variability in the accuracy across different folds.
On the other hand, kNN shows more variability, with a wider box and longer whiskers, suggesting that its performance fluctuates more across the cross-validation folds. The Naive Bayes classifier is not only more accurate but also more stable than kNN in this scenario. The higher stability could be attributed to the probabilistic nature of Naive Bayes, which is less affected by small variations in the data compared to kNN, which relies on distances and local neighborhoods.

alínea b):



Mean accuracy of KNN (k=5) after Min-Max scaling: 0.8217

Mean accuracy of Gaussian Naive Bayes after Min-Max scaling:0.8350

K-Nearest Neighbors (kNN):

The mean accuracy of 0.8217 indicates a significant improvement in performance compared to the previous accuracy when the data wasn't scaled. As noted earlier, kNN's performance relies heavily on the distance between points in the feature space. By scaling to the same range, we allow the kNN algorithm to function more effectively, resulting in the observed accuracy.

Gaussian Naive Bayes (GNB):

The mean accuracy of 0.8350 shows that GNB performs slightly better than kNN in this case. GNB's performance benefits from Min-Max scaling in terms of ensuring that the numerical range of the features does not lead to computational issues. However, because GNB operates under different assumptions about the distribution of data, its performance is inherently more stable regardless of scaling.

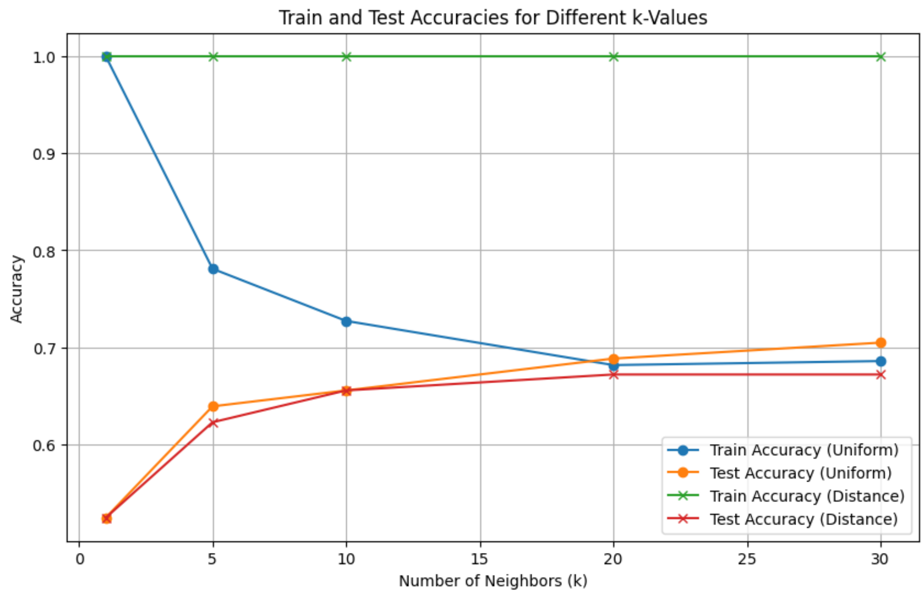
alínea c):

T-statistic:-6.6903

P-value: 0.9987

Conclusion: Fail to reject the null hypothesis. There is no significant difference between KNN and Naive Bayes.

2. Graphic for exercise 2.
   alínea a):



alínea b):

The number of neighbors (k) in a kNN classifier has a profound impact on its generalization ability. As k increases, the model shifts from being sensitive to individual training samples to a more generalized perspective that considers broader patterns in the data.

When k is low, the model may memorize the training data, resulting in high training accuracy but poor performance on unseen data due to overfitting. As k increases, the model starts to average predictions over more neighbors, which can reduce the impact of noise and improve generalization, leading to higher test accuracy. This demonstrates that moderate values of k can help the model better reflect the true underlying distribution of the data.

Beyond a certain k, the model can become too generalized, losing the ability to discern important patterns. This underfitting occurs because averaging predictions across too many neighbors can dilute the influence of relevant data points, decreased accuracy.

3. Two difficulties of the naïve Bayes model

1. Handling Continuous Features: Naïve Bayes models assume that features are categorical or follow a specific distribution (Gaussian, for example). However, several features in this dataset are continuous (e.g., age, trestbps, chol, thalach, oldpeak). The model may need to make assumptions for these variables, which may not hold true and could reduce its accuracy.

2. Correlated Features: Naïve Bayes assumes that all features are independent, but in medical datasets like this one, many features are likely to be correlated. For instance, blood pressure (trestbps) and cholesterol (chol) might be related, and the presence of such correlations can violate the independence assumption, leading to suboptimal model performance.