Aprendizagem 2024/25

Homework II – Group 028

(ist1106987, ist1107057)

I. Pen-and-paper

1)

Leaving $\boldsymbol{x_1}$ = (A, 0) out:

Neighbors:

x_2 = (B,1) → Hamming distance = 2

(both features differ)

x_3 = (A,1) → 1

x_4 = (A,0) → 0

x_5 = (B,0) → 1

x_6 = (B,0) → 1

x_7 = (A,1) → 1

x_8 = (B,1) → 2

Closest 5 neighbors: x_4, x_3, x_5, x_6, x_7

(P, P, N, N, N) → Prediction: N (FN)

$\boldsymbol{x_2}$ out: Closest 5 neighbors: x_8, x_3, x_5, x_6, x_7

(N, P, N, N, N) → Prediction: N (FN)

$\boldsymbol{x_3}$ out: Closest 5 neighbors: x_7, x_1, x_2, x_4, x_8

(N, P, P, P, N) → Prediction: P (TP)

$\boldsymbol{x_4}$ out: Closest 5 neighbors: x_1, x_3, x_5, x_6, x_7

(P, P, N, N, N) → Prediction: N (FN)

$\boldsymbol{x_5}$ out: Closest 5 neighbors: x_6, x_1, x_2, x_4, x_8

(N, P, P, P, N) → Prediction: P (FP)

$\boldsymbol{x_6}$ out: Closest 5 neighbors: $x_5, x_1\ x_2, x_4, x_8$

(N, P, P, P, N) → Prediction: P (FP)

$\boldsymbol{x_7}$ out: Closest 5 neighbors: x_3, x_1, x_2, x_4, x_8

(P, P, P, P, N) → Prediction: P (FP)

$\boldsymbol{x_8}$ out: Closest 5 neighbors: $x_2,\ x_3, x_5, x_6, x_7$

(P, P, N, N, N) → Prediction: N (TN)

TP: 1 (x_3)

FP: 3 (x_5, x_6, x_7)

TN: 1 (x_8)

FN: 3 ($x_1\ x_2, x_4$)

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

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

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + R} = 2 \times \frac{\frac{1}{4}\times\frac{1}{4}}{\frac{1}{4}+\frac{1}{4}} = 0.25$$

2)

Using a distance-weighted kNN with $k = 3k$ and a modified distance function similar to the Hamming distance, but counting 2 instead of 1 when y_1 differs, the following results were obtained:

$$D(x_i, x_j) = 2 \times \left(y_1(x_i) \neq y_1(x_j)\right) + 1 \times \left(y_2(x_i) \neq y_2(x_j)\right), \text{where}$$

$$\left(y_m(x_i) \neq y_m(x_j)\right) = \begin{cases} 1, & \text{if True} \\ 0, & \text{if False} \end{cases}, m = \{1, 2\}$$

Leaving $\boldsymbol{x_1}$ = (A, 0) out:

Neighbors:

x_2 = (B,1) → 3 (both features differ)

x_3 = (A,1) → 1

x_4 = (A,0) → 0

x_5 = (B,0) → 2

x_6 = (B,0) → 2

x_7 = (A,1) → 1

x_8 = (B,1) → 3

Closest 3 neighbors: x_3, x_4, x_7

(P, P, N) → Prediction: P (TP)

$\boldsymbol{x_2}$ out: Closest 3 neighbors: x_8, x_5, x_6

(N, N, N) → Prediction: N (FN)

$\boldsymbol{x_3}$ out: Closest 3 neighbors: x_7, x_1, x_4

(N, P, P) → Prediction: P (TP)

$\boldsymbol{x_4}$ out: Closest 3 neighbors: x_1, x_3, x_7

(P, P, N) → Prediction: P (TP)

$\boldsymbol{x_5}$ out: Closest 3 neighbors: x_6, x_2, x_8

(N, P, N) → Prediction: N (TN)

$\boldsymbol{x_6}$ out: Closest 3 neighbors: x_5, x_2, x_8

(N, P, N) → Prediction: N (TN)

$\boldsymbol{x_7}$ out: Closest 3 neighbors: x_3, x_1, x_4
(P, P, P) → Prediction: P (FP)

$\boldsymbol{x_8}$ out: Closest 3 neighbors: x_2, x_5, x_6
(P, N, N) → Prediction: N (TN)

TP: 3 (x_1, x_3, x_4)

FP: 1 (x_7)

TN: 3 (x_5, x_6, x_8)

FN: 1 (x_2,)

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

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

$F1 = 2 \times \frac{Precision \times Recall}{Precision + Re} = 2 \times \frac{\frac{3}{4}\times\frac{3}{4}}{\frac{3}{4}+\frac{3}{4}} = 0.75$

3)

$$p(h|\mathbb{X}) = \frac{p(\mathbb{X}|h) \times p(h)}{p(\mathbb{X})}$$

$p(h)$: **Priors:** $p(P) = \frac{5}{9}$, $p(N) = \frac{4}{9}$

$p(\mathbb{X}|h)$: **Probability Mass Functions:**

$$p(A,0|P) = \frac{2}{5}\ ,\ p(A,1|P) = \frac{1}{5},\ p(B,0|P) = \frac{1}{5},\ p(B,1|P) = \frac{1}{5}$$

$$p(A,0\ |N) = 0, p(A,1|N) = \frac{1}{4}, p(B,0|N) = \frac{2}{4}, p(B,1|N) = \frac{1}{4}$$

Probability Density Functions:

$$N(u_3 = 0.82, \sigma_3^2 = 0.047|P), N(u_3 = 1, \sigma_3^2 = 0.02|N)\ *$$

$$p(A,0) = \frac{2}{9}, p(A,1) = \frac{2}{9}\ , p(B,0) = \frac{3}{9}\ , p(B,1) = \frac{2}{9},\ N(u_3 = 0.9, \sigma_3^2 = 0.2)*$$

* Mean and Variance were automatically calculated on https://www.calculator.net/standard-deviation-calculator.html

4)

$$h_{\text{MAP}} = \text{argmax}_h\ p(h|\mathbb{X}_{\text{new}}) = \text{argmax}_h\ p(\mathbb{X}_{\text{new}}|h)\ p(h)$$

$$\text{Normal: } f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

$$\mathbb{X}_1: p(A, 1, 0.8\ |P)\ p(P) = p(A, 1|P)\ p(0.8|P)\ p(P) = \frac{1}{5} \times 1.806 \times \frac{5}{9} \approx 0.201$$

$$p(0.8|P) = f_X(0.8) \approx 1.806, \text{where } X \sim N(u_3 = 0.82, \sigma_3^2 = 0.047|P)$$

$$p(A, 1, 0.8\ |N)\ p(N) = p(A, 1|N)\ p(0.8|N)\ p(N) = \frac{1}{4} \times 1.207 \times \frac{4}{9} \approx 0.134$$

$$p(0.8|N) = f_X(0.8) \approx 1.207\ , \text{where } X \sim N(u_3 = 1, \sigma_3^2 = 0.02|N)$$

$\mathbb{X}_1 = \{A, 1, 0.8\}$ is classified as Positive.

$$\mathbb{X}_2: p(B, 1, 1\ |P)\ p(P) = p(B, 1|P)\ p(1|P)\ p(P) = \frac{1}{5} \times 1.298 \times \frac{5}{9} \approx 0.144$$

$$p(1|P) = f_X(1) \approx 1.298, \text{where } X \sim N(u_3 = 0.82, \sigma_3^2 = 0.047|P)$$

$$p(B, 1, 1\ |N)\ p(N) = p(B, 1|P)\ p(1|P)\ p(P) = \frac{1}{4} \times 2.850 \times \frac{4}{9} \approx 0.317$$

$$p(1|N) = f_X(1) \approx 2.850, \text{where } X \sim N(u_3 = 1, \sigma_3^2 = 0.02|N)$$

$\mathbb{X}_2 = \{B, 1, 1\}$ is classified as Negative.

$$\mathbb{X}_3: p(B, 0, 0.9\ |P)\ p(P) = p(B, 0|P)\ p(0.9|P)\ p(P) = \frac{1}{5} \times 1.697 \times \frac{5}{9} \approx 0.189$$

$$p(0.9|P) = f_X(0.9) \approx 1.697, \text{where } X \sim N(u_3 = 0.82, \sigma_3^2 = 0.047|P)$$

$$p(B, 0, 0.9\ |N)\ p(N) = p(B, 0|N)\ p(0.9|N)\ p(N) = \frac{2}{4} \times 2.208 \times \frac{4}{9} \approx 0.491$$

$$p(0.9|N) = f_X(0.9) \approx 2.208, \text{where } X \sim N(u_3 = 1, \sigma_3^2 = 0.02|N)$$

$\mathbb{X}_3 = \{B, 0, 0.9\}$ is classified as Negative.

5) Vocabulary = {amazing, run, I, like, it, too, tired, bad} $\rightarrow V = 8$

$$N_c = \begin{cases} 5, c = P \\ 4, c = N \end{cases}$$

$$p(\text{"I"}|P) = \frac{1+1}{5+8} = \frac{2}{13} \qquad p(\text{"I"}|N) = \frac{0+1}{4+8} = \frac{1}{12}$$

$$p(\text{"like"}|P) = \frac{1+1}{5+8} = \frac{2}{13} \qquad p(\text{"like"}|N) = \frac{0+1}{4+8} = \frac{1}{12}$$

$$p(\text{"to"}|P) = \frac{0+1}{5+8} = \frac{1}{13} \qquad p(\text{"to"}|N) = \frac{0+1}{4+8} = \frac{1}{12}$$

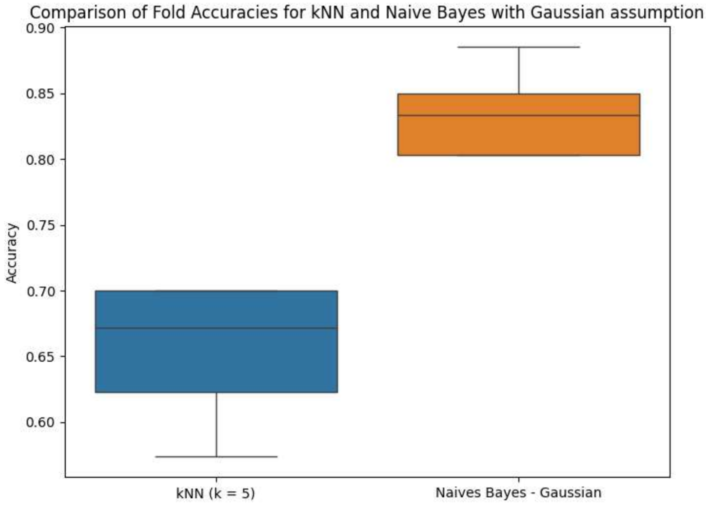
$$p(\text{"run"}|P) = \frac{1+1}{5+8} = \frac{2}{13} \qquad p(\text{"run"}|N) = \frac{1+1}{4+8} = \frac{2}{12}$$

$$p(\text{"I like to run"}|P) = \frac{2}{13} \times \frac{2}{13} \times \frac{1}{13} \times \frac{2}{13} \approx 0.00028 \qquad p(\text{"I like to run"}|N) = \frac{1}{12} \times \frac{1}{12} \times \frac{1}{12} \times \frac{2}{12} \approx 0.000096$$

$p(\text{"I like to run"}|P) > p(\text{"I like to run"}|N)$, so "I like to run" is classified as Positive.

II. Programming and critical analysis

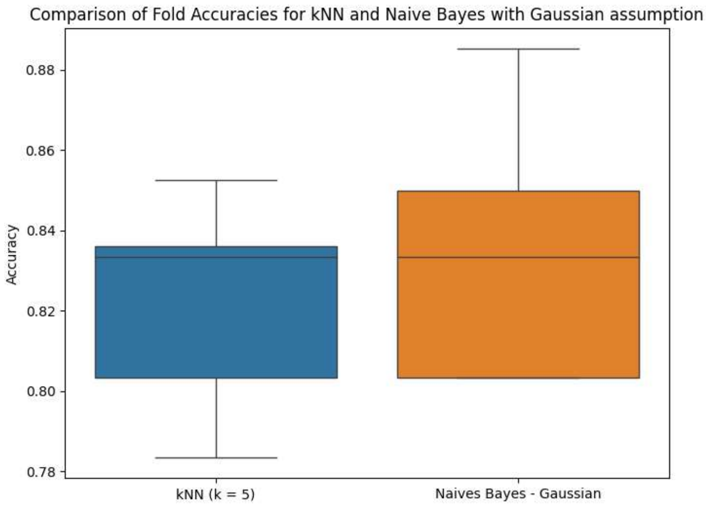
1)
a.



Naïve Bayes is more stable, as shown by the smaller variation in accuracy values across the folds, in comparison to kNN.

This might be because Naïve Bayes tends to perform well even with smaller datasets, such as the given one, due to its independence assumptions, which smooth out the impact of data variability. On the other hand, kNN relies on distance metrics, which are sensitive to changes in the dataset and lead to instability.

b.



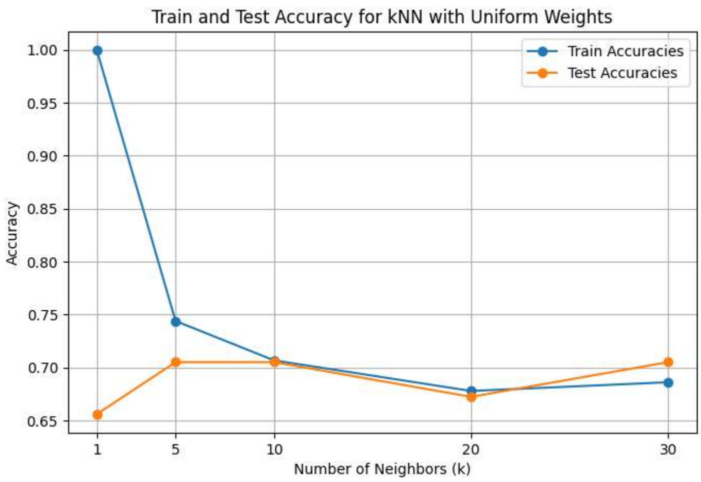
Concerning kNN, the range of accuracies is narrower and the overall accuracy is higher, [0.78; 0.85], in comparison to the previous exercise. Previously, features with larger ranges were more important in measuring distances. Therefore, when applying Min-Max scaler, which assigns the feature values to a common scale, all features contribute equally to the distance metric. This results in more balanced distance calculations, improving the performance consistency and accuracy of the model.

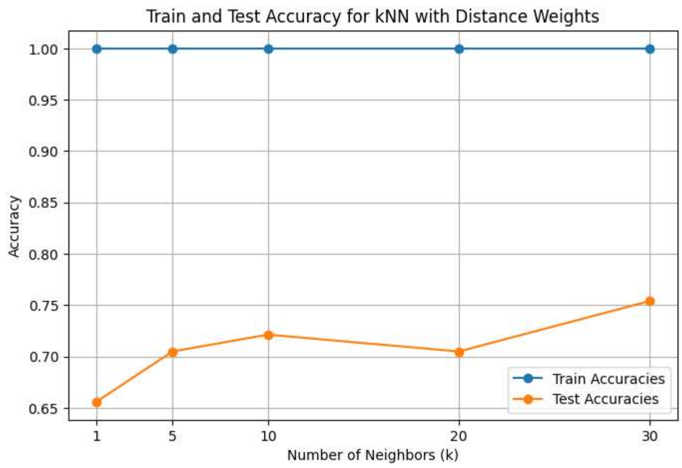
Concerning Naïve Bayes, the performance is similar to previously, [0.80, 0.89]. Since Gaussian Naïve Bayes is based in the distribution for each feature, the relative positions of the data points within each feature's distribution are more important than the range of the feature values. Therefore, the Min-Max scaler, while it changes the range of the feature values, does not change the shape of their distribution, which translates to a minimal impact of the scaler in this model.

c. Using the models described in question 1.a, p-value $= 0.9987$.
The null hypothesis can not be rejected at any significance level $\alpha \leq 99.87\%$, including common significance levels (1%, 5% and 10%). Therefore, it is not possible to assert the given hypothesis as true. However, considering the absence of additional statistical tests, the given hypothesis can not be taken as rejected neither.

2)

a.





b.

Concerning the kNN with uniform weights, the large difference between training and testing accuracies in $k = 1$ suggests the model is overfitting. From $k = 1$ to $k = 5$, the training accuracy drops significantly, as the model becomes more generalized. From that point on, it continues to decrease at a slower rate. On the other hand, from $k = 1$ to $k = 5$ the testing accuracy increases, but afterwards begins to decline. This suggests that the model underfits with higher k values, as it becomes less sensitive to individual data points, which translates in a poorer performance.

Concerning the kNN with distance weights, the train accuracy is constant at 1.0, which means that using distance weights allows the model to correctly classify all training points. On the other hand, the testing accuracy fluctuates, demonstrating that the model is more consistent when it comes to generalization, since closer points are weighted more heavily, maintaining relevance even as k increases.

3)

One possible difficulty is handling continuous variables such as age, trestbps, chol, thalach, and oldpeak. In the naïve Bayes model, continuous variables are assumed to follow a normal (Gaussian) distribution. However, in the real world, these variables could follow a very different distribution. As a result, the model might make predictions that have little connection to the patterns in the data. If the continuous variables have asymmetric distributions or outliers, the assumption of this distribution can lead to unrealistic predictions.

Other possible difficulty might be that naïve Bayes assumes that all features are conditionally independent given the class. However, in this dataset, certain features may be correlated (for instance, trestbps(resting blood pressure) and chol(cholesterol) may be linked, as they are cardiovascular health indicators. This assumption of independence when there is correlation can lead to overestimating their combined impact on the prediction, affecting the model's accuracy.