Aprendizagem 2023
Homework I — Group 003
(ist1107028, ist1107137)

Part I: Pen and paper

We collected four positive (P) observations,

$$\{\boldsymbol{x_1} = (\boldsymbol{A}, 0), \quad \boldsymbol{x_2} = (\boldsymbol{B}, 1), \quad \boldsymbol{x_3} = (\boldsymbol{A}, 1), \quad \boldsymbol{x_4} = (\boldsymbol{A}, 0)\}$$

and four negative (N) observations,

$$\{\boldsymbol{x_5} = (\boldsymbol{B}, 0), \quad \boldsymbol{x_6} = (\boldsymbol{B}, 0), \quad \boldsymbol{x_7} = (\boldsymbol{A}, 1), \quad \boldsymbol{x_8} = (\boldsymbol{B}, 1)\}$$

Consider the problem of classifying observations as positive or negative.

1. **Compute the F1-measure of a kNN with k = 5 and Hamming distance using a leave-one-out evaluation schema. Show all calculus.**

We start by calculating Hamming distance between observations. The Hamming distance is the number of positions at which the corresponding symbols are different.
Since we are workin with $k = 5$, we will consider the 5 nearest neighbors of each observation (written in blue).

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	-	2	**1**	**0**	**1**	**1**	**1**	2
x_2	2	-	**1**	2	**1**	**1**	**1**	**0**
x_3	**1**	**1**	-	**1**	2	2	**0**	**1**
x_4	**0**	2	**1**	-	**1**	**1**	**1**	2
x_5	**1**	**1**	2	**1**	-	**0**	2	**1**
x_6	**1**	**1**	2	**1**	**0**	-	2	**1**
x_7	**1**	**1**	**0**	**1**	2	2	-	**1**
x_8	2	**0**	**1**	2	**1**	**1**	**1**	-

Table 1: Hamming distance between observations

Now that we have the Hamming distance between all observations, we must identify if the prediction is correct or not. We will consider the majority class of the 5 nearest neighbors for each observation.

Example: For x_1, the 5 nearest neighbors are x_3 and x_4 (which are positive), x_5, x_6 and x_7 (which are negative). The majority class is negative, therefore the prediction is incorrect.

We apply the same logic for the rest of the classes, ending up with the following table:

Observation	True Value	Prediction	Confusion Matrix Terminology
x_1	P	N	FP
x_2	P	N	FP
x_3	P	P	TP
x_4	P	N	FN
x_5	N	P	FP
x_6	N	P	FP
x_7	N	P	FP
x_8	N	N	TN

P - Positive observation; N - Negative observation

TP - True Positive; TN - True Negative; FP - False Positive; FN - False Negative

Table 2: Predictions for each observation

With this table, we can know calculate the Precision, Recall and F1-measure using the following formulas:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{1}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{2}$$

$$\text{F1-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

Replacing the corresponding values in the formulas, we get:

for Precision (1) and Recall (2):

$$\text{Precision} = \frac{1}{1+5} \approx 0.1667 \qquad \text{Recall} = \frac{1}{1+1} = 0.5$$

F1-measure (3):

$$\text{F1-measure} = 2 \times \frac{0.1667 \times 0.5}{0.1667 + 0.5} \approx 0.25$$

2. **Propose a new metric (distance) that improves the latter's performance (i.e., the F1-measure) by three fold.**

   To improve the F1-measure by three fold, we can use $k =?$ instead of $k = 5$:

An additional positive observation was acquired, $\boldsymbol{x_9 = (B, 0)}$, and a third variable $\boldsymbol{y_3}$ was independently monitored, yielding estimates,

$$\boldsymbol{y_3|P = \{1.1, 0.8, 0.5, 0.9, 0.8\} \quad and \quad y_3|N = \{1, 0.9, 1.2, 0.9\}}$$

3. **Considering the nine training observations, learn a Bayesian classifier assuming: (i) $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ are dependent; (ii) $\boldsymbol{y_1}$, $\boldsymbol{y_2}$ and $\boldsymbol{y_3}$ variable sets are independent and equally important; and (iii) $\boldsymbol{y_3}$ is normally distributed. Show all parameters.**

   With the nine training observations, we can calculate the parameters for the Bayesian classifier. We will refer to the outcome, which can be positive or negative, as P and N respectively, and 'class' or 'c' when we mean both.

   To estimate $P(\text{class}|y_1, y_2, y_3)$, we can use Bayes' theorem:

$$P(\text{class}|y_1, y_2, y_3) = \frac{P(y_1, y_2, y_3|\text{class}) \times P(\text{class})}{P(y_1, y_2, y_3)} \tag{4}$$

   Since we know $\{y_1, y_2\}$ and $\{y_3\}$ are independent, we can rewrite $P(y_1, y_2, y_3)$ as $P(y_1, y_2) \times P(y_3)$. Rewriting (4) with this, results in:

$$P(\text{class}|y_1, y_2, y_3) = \frac{P(y_1, y_2|\text{class})P(y_3|\text{class}) \times P(\text{class})}{P(y_1, y_2) \times P(y_3)} \tag{5}$$

   Given a new observation O, we are able to classify it by calculating $P(\text{class}|O)$ for all classes and selecting the class with the highest probability as our prediction.

$$\begin{aligned} \hat{z} &= \underset{c \in \{P,N\}}{\text{arg max}} \ \{P(\text{c}|O)\} \\ &= \underset{c \in \{P,N\}}{\text{arg max}} \left\{ \frac{P(y_1, y_2|c)P(y_3|c) \times P(c)}{P(y_1, y_2)P(y_3)} \right\} \\ &= \underset{c \in \{P,N\}}{\text{arg max}} \ \{P(y_1, y_2|c)P(y_3|c) \times P(c)\} \quad \text{(we can remove parameters that do not depend on } c\text{)} \end{aligned} \tag{6}$$

   We can now begin to compute these parameters.

$$\text{Priors:} \qquad P(P) = \frac{5}{9} \qquad P(N) = 1 - P(P) = \frac{4}{9}$$

To get the likelihoods, we start by calculating the joint probabilities for the categorical variables y_1 and y_2:

$$\begin{aligned}
P((y_1 = A, y_2 = 0)|P) &= \frac{2}{5} \quad P((y_1 = A, y_2 = 1)|P) &= \frac{1}{5} \\
P((y_1 = A, y_2 = 0)|N) &= \frac{0}{4} \quad P((y_1 = A, y_2 = 1)|N) &= \frac{1}{4} \\
\\
P((y_1 = B, y_2 = 0)|P) &= \frac{1}{5} \quad P((y_1 = B, y_2 = 1)|P) &= \frac{1}{5} \\
P((y_1 = B, y_2 = 0)|N) &= \frac{2}{4} \quad P((y_1 = B, y_2 = 1)|N) &= \frac{1}{4}
\end{aligned}$$

For the third variable, we know that it is normally distributed, so we start by calculating the mean and variance for each class:

$$\textbf{mean:} \quad \mu = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad\qquad \textbf{variance:} \quad \sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \mu)^2$$

Positive class:

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

$$\sigma_P^2 = \tfrac{(1.1-0.82)^2+(0.8-0.82)^2+(0.5-0.82)^2+(0.9-0.82)^2+(0.8-0.82)^2}{4} = 0.47$$

Negative class:

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

$$\sigma_N^2 = \tfrac{(1-1)^2+(0.9-1)^2+(1.2-1)^2+(0.9-1)^2}{3} = 0.02$$

Since y_3 is normally distributed, we can calculate the likelihood for each class using the normal distribution formula:

$$P(y_z|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_z - \mu)^2}{2\sigma^2}\right) \tag{7}$$

We must now replace y_3 with each value in the set $\{1.1, 0.8, 0.5, 0.9, 0.8\}$ and $\{1, 0.9, 1.2, 0.9\}$ to calculate the likelihoods for each class. Having the parameters for the normal distribution, we can calculate the likelihoods for y_3 as per (7):

Positive class:

$$
\begin{aligned}
P(y_3 = 1.1|\mu_P, \sigma_P^2) &= \frac{1}{\sqrt{2\pi \times 0.47}} \exp\left(-\frac{(1.1-0.82)^2}{2 \times 0.47}\right) \approx 0.535 \\
P(y_3 = 0.8|\mu_P, \sigma_P^2) &= \frac{1}{\sqrt{2\pi \times 0.47}} \exp\left(-\frac{(0.8-0.82)^2}{2 \times 0.47}\right) \approx 0.582 \\
P(y_3 = 0.5|\mu_P, \sigma_P^2) &= \frac{1}{\sqrt{2\pi \times 0.47}} \exp\left(-\frac{(0.5-0.82)^2}{2 \times 0.47}\right) \approx 0.522 \\
P(y_3 = 0.9|\mu_P, \sigma_P^2) &= \frac{1}{\sqrt{2\pi \times 0.47}} \exp\left(-\frac{(0.9-0.82)^2}{2 \times 0.47}\right) \approx 0.578
\end{aligned}
$$

Negative class:

$$
\begin{aligned}
P(y_3 = 1|\mu_N, \sigma_N^2) &= \frac{1}{\sqrt{2\pi \times 0.02}} \exp\left(-\frac{(1-1)^2}{2 \times 0.02}\right) \approx 2.821 \\
P(y_3 = 0.9|\mu_N, \sigma_N^2) &= \frac{1}{\sqrt{2\pi \times 0.02}} \exp\left(-\frac{(0.9-1)^2}{2 \times 0.02}\right) \approx 2.197 \\
P(y_3 = 1.2|\mu_N, \sigma_N^2) &= \frac{1}{\sqrt{2\pi \times 0.02}} \exp\left(-\frac{(1.2-1)^2}{2 \times 0.02}\right) \approx 1.038
\end{aligned}
$$

Now we have all the parameters to apply the Bayesian classifier to new observations.

Consider now three testing observations,

$$\boldsymbol{\{(A, 1, 0.8), (B, 1, 1), (B, 0, 0.9)\}}$$

4. **Under a MAP assumption, classify each testing observation showing all your calculus.**

MAP (Maximum A Posteriori) is defined as:

$$
\begin{aligned}
\hat{z} &= \arg\max_{c_i} \{P(c_i|x)\} \\
&= \arg\max_{c_i} \left\{\frac{P(x|c_i) \times P(c_i)}{P(x)}\right\} \\
&= \arg\max_{c_i} \{P(x|c_i) \times P(c_i)\}
\end{aligned}
\tag{8}
$$

In order to apply this assumption we use the likelihoods and the priors calculated in the beginning of this report.

$$P(P) = \frac{5}{9} \qquad P(N) = 1 - P(P) = \frac{4}{9}$$

$$\begin{aligned}
P((y_1 = A, y_2 = 1)|P) &= \frac{1}{5} & P((y_1 = A, y_2 = 1)|N) &= \frac{1}{4} \\
P((y_1 = B, y_2 = 1)|P) &= \frac{1}{5} & P((y_1 = B, y_2 = 1)|N) &= \frac{1}{4} \\
P((y_1 = B, y_2 = 0)|P) &= \frac{1}{5} & P((y_1 = B, y_2 = 0)|N) &= \frac{2}{4}
\end{aligned}$$

$$\begin{aligned}
&P(y_3 = 0.8|\mu_P, \sigma_P^2) \approx 0.582 \quad P(y_3 = 1.0|\mu_N, \sigma_N^2) \approx 2.821 \\
&P(y_3 = 0.9|\mu_N, \sigma_N^2) \approx 2.197 \quad P(y_3 = 0.9|\mu_P, \sigma_P^2) \approx 0.578
\end{aligned}$$

Before using MAP assumption, we must calculate the missing values needed - $P(y_3 = 0.8|\mu_N, \sigma_N^2)$ and $P(y_3 = 1|\mu_P, \sigma_P^2)$.

$$\begin{aligned}
P(y_3 = 0.8|\mu_N, \sigma_N^2) &= \frac{1}{\sqrt{2\pi \times 0.02}} \exp\left(-\frac{(0.8 - 1)^2}{2 \times 0.02}\right) \approx 2.821 \\
P(y_3 = 1|\mu_P, \sigma_P^2) &= \frac{1}{\sqrt{2\pi \times 0.47}} \exp\left(-\frac{(1 - 0.82)^2}{2 \times 0.47}\right) \approx 0.535
\end{aligned}$$

With this, we can just replace the values in the MAP formula (8) to get the predictions for each observation.

$\boxed{(A, 1, 0.8)}$

$$\begin{aligned}
P(y_1 = A, y_2 = 1, y_3 = 0.8|P) &= P(y_1 = A, y_2 = 1|\text{P}) \times P(y_3 = 0.8|\text{P}) \times P(P) \\
&= \frac{1}{5} \times 0.582 \times \frac{5}{9} \approx 0.065
\end{aligned}$$

$$\begin{aligned}
P(y_1 = A, y_2 = 1, y_3 = 0.8|N) &= P(y_1 = A, y_2 = 1|\text{N}) \times P(y_3 = 0.8|\text{N}) \times P(N) \\
&= \frac{1}{4} \times 2.197 \times \frac{4}{9} \approx 0.488
\end{aligned}$$

$$\begin{aligned}
\hat{z}_{(A,1,0.8)} &= \underset{c \in \{P,N\}}{\arg\max} \; \{P(y_1 = A, y_2 = 1|c)P(y_3 = 0.8|c) \times P(c)\} \\
&= \arg\max \; \{P(y_1, y_2|P)P(y_3|P) \times P(P); P(y_1, y_2|N)P(y_3|N) \times P(N)\} \\
&= N
\end{aligned}$$

$\boxed{(B,1,1)}$

$$
\begin{aligned}
P(y_1 = B, y_2 = 1, y_3 = 1|P) &= P(y_1 = B, y_2 = 1|\text{P}) \times P(y_3 = 1|\text{P}) \times P(P) \\
&= \frac{1}{5} \times 2.821 \times \frac{5}{9} \approx 0.313
\end{aligned}
$$

$$
\begin{aligned}
P(y_1 = B, y_2 = 1, y_3 = 1|N) &= P(y_1 = B, y_2 = 1|\text{N}) \times P(y_3 = 1|\text{N}) \times P(N) \\
&= \frac{1}{4} \times 0.535 \times \frac{4}{9} \approx 0.059
\end{aligned}
$$

$$
\begin{aligned}
\hat{z}_{(B,1,1)} &= \underset{c \in \{P,N\}}{\arg\max} \ \{P(y_1 = B, y_2 = 1|c)P(y_3 = 1|c) \times P(c)\} \\
&= \arg\max \ \{P(y_1, y_2|P)P(y_3|P) \times P(P); P(y_1, y_2|N)P(y_3|N) \times P(N)\} \\
&= P
\end{aligned}
$$

$\boxed{(B,0,0.9)}$

$$
\begin{aligned}
P(y_1 = B, y_2 = 0, y_3 = 0.9|P) &= P(y_1 = B, y_2 = 0|\text{P}) \times P(y_3 = 0.9|\text{P}) \times P(P) \\
&= \frac{1}{5} \times 0.578 \times \frac{5}{9} \approx 0.064
\end{aligned}
$$

$$
\begin{aligned}
P(y_1 = B, y_2 = 0, y_3 = 0.9|N) &= P(y_1 = B, y_2 = 0|\text{N}) \times P(y_3 = 0.9|\text{N}) \times P(N) \\
&= \frac{2}{4} \times 2.197 \times \frac{4}{9} \approx 0.488
\end{aligned}
$$

$$
\begin{aligned}
\hat{z}_{(B,0,0.9)} &= \underset{c \in \{P,N\}}{\arg\max} \ \{P(y_1 = B, y_2 = 0|c)P(y_3 = 0.9|c) \times P(c)\} \\
&= \arg\max \ \{P(y_1, y_2|P)P(y_3|P) \times P(P); P(y_1, y_2|N)P(y_3|N) \times P(N)\} \\
&= N
\end{aligned}
$$

Therefore, the predictions for each observation are N, P and N respectively.

At last, consider only the following sentences and their respective connotations,

$$\boldsymbol{\{(\text{"}\textit{Amazing run}\text{"}, P), (\text{"}\textit{I like it}\text{"}, P), (\text{"}\textit{To tired}\text{"}, N), (\text{"}\textit{Bad run}\text{"}, N)\}}$$

5. **Using a naïve Bayes under a ML assumption, classify the new sentence "I like to run". For the likelihoods calculation consider the following formula,**

$$\boldsymbol{P(T_i|c) = \frac{freq(t_i) + 1}{N_c + V}}$$

where $\mathbf{t_i}$ represents a certain term i, V the number of unique terms in the vocabulary, and $\mathbf{N_c}$ the total number of terms in class c. Show all calculus.

Part II: Programming