EXERCISE 1

1. Provide the definition of *Confusion matrix* for a binary classification problem, formally explain the content of the matrix.
2. Provide a numerical example of a **non-symmetric** confusion matrix for an **unbalanced** dataset with 2 classes (e.g., about 90% of samples from the negative class). Show the confusion matrix in two formats: with absolute values and with the corresponding percentage values. (Hint: use simple numerical values, so that you do not need to make complex calculations.)
3. Compute accuracy, precision and recall according to the numerical example provided above.

The Recall is: $\frac{TP}{TP+FN}$. The matrix C is the conf. matrix where

C_{ij} : number of class i element classified as j where $i, j \in \{0,1\}$

$$C: \begin{array}{c|cc} \text{Pre.} \backslash \text{Real} & 0 & 1 \\ \hline 0 & 85 & 3 \\ 1 & 5 & 7 \end{array}$$

$$\text{precision} = \frac{TP}{TP+FP} = 58\%$$

$$\text{recall} = \frac{TP}{TP+FN} = 70\%$$

$$\text{accuracy} = 92\%$$

EXERCISE 2

1. Formally describe the *Bayes Optimal Classifier* and the *Naive Bayes Classifier* and highlight their differences. Explain all the terms of the formulas.
2. Consider a classification problem $f : A_1 \times A_2 \times A_3 \to \{T, F\}$, with $A_1 = \{a, b, c\}$, $A_2 = \{h, k\}$, $A_3 = \{u, v, w\}$ and the data set in the table on the right. Use Naive Bayes to predict the output for the input value (a, k, u), showing all the steps needed to provide the answer.

A_1	A_2	A_3	f
a	h	v	F
b	k	w	T
c	h	v	T
b	k	u	F
a	k	w	T
c	h	u	T

The Bayes Optimal Classifier classify x as follow:

$$t(x) = \underset{c}{\operatorname{argmax}}\, \mathbb{P}(c \mid x, \underset{\text{dataset}}{D}) = \underset{c}{\operatorname{argmax}} \sum_{h} \mathbb{P}(c \mid h, x)\,\mathbb{P}(h \mid D)$$

all hypothesis

The Naïve assumes that the attributes are independent

$$\Rightarrow \underset{c}{\operatorname{argmax}}\, \mathbb{P}(c \mid x, D) = \underset{c}{\operatorname{argmax}}\, \mathbb{P}(x \mid c, D)\,\mathbb{P}(c \mid D)$$

$$= \underset{c}{\operatorname{argmax}}\, \mathbb{P}(c \mid D) \prod_{a_i \in x} \mathbb{P}(a_i \mid c, D)$$

$$\mathbb{P}(c \mid D) = \frac{|\{(x, b) \in D : b = c\}|}{|D|} \qquad \mathbb{P}(a_i \mid c, D) = \frac{|\{(x, b) \in D : b = c \wedge a_i \in x\}|}{|\{(x, b) \in D : b = c\}|}$$

For the Dataset, we have:

$\mathbb{P}(F \mid D) = 1/3$ $\mathbb{P}(a \mid F, D) = 1/2$ $\mathbb{P}(k \mid F, D) = 1/2$ $\mathbb{P}(u \mid F, D) = 1/2$

$\mathbb{P}(T \mid D) = 2/3$ $\mathbb{P}(a \mid T, D) = 1/4$ $\mathbb{P}(k \mid T, D) = 1/2$ $\mathbb{P}(u \mid T, D) = 1/4$

$$\Rightarrow \mathbb{P}(F \mid (a,k,u), D) = \frac{1}{3} \cdot \left(\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}\right) = 1/24$$

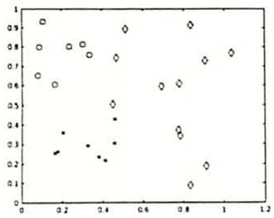
$$\mathbb{P}(T \mid (a,k,u), D) = \frac{2}{3} \cdot \left(\frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{4}\right) = 1/48$$

$\Rightarrow$ output: F

EXERCISE 3

Consider a 3-classes classification problem and the data shown in the figure on the right and classification based on support vector machines (SVMs):

1. Describe a linear model for this problem (3-classes classification)
2. Explain if the data in the figure are linearly separable and motivate your answer
3. Explain what type of kernel function for SVM you would use for this dataset and provide the formal definition of the kernelized model.



We can use 3 vectors W_1, W_2, W_3 and assign the class C_i where $i = \operatorname{argmax}_i W_i^T x$. The dataset is separable, so a linear kernel is sufficient.