Machine Learning – December 16, 2024

Last Name	First Name	Matricola
....	....	....

1. No books, slides, written notes are allowed during the exam.
2. Answers must be explicitly marked with the question they refer to (e.g., **2.1** for question 1 of exercise 2). Cumulative answers which refer to more questions will be evaluated as answering one question only.

Time limit: **1h 45min.**

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.

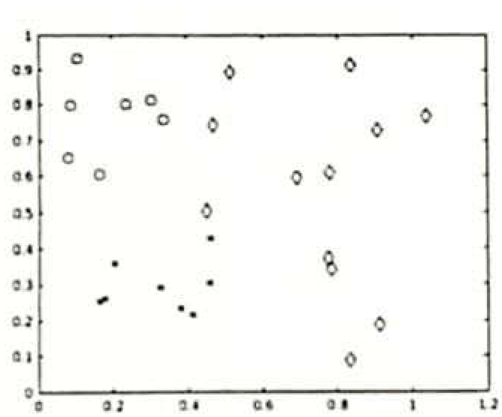
EXERCISE 2

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 $f : A_1 \times A_2 \times A_3 \rightarrow \{T, F\}$, 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

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.

EXERCISE 4

Let $\mathcal{D}$ be a dataset for four-classes classification (with class labels A, B, C, D) containing the following input values $X = \{(3.3, 1.6, -0.72), (7.5, 48.2, 0.01), \ldots, (98.3, 43.5, -2.54), (87.2, 92.4, 0.15)\}$ and target values $T = \{C, A, \ldots, B, D\}$.

Consider designing a Feedforward Neural Network for learning the function $t = f(x)$.

1. Explain what is a valid choice for the activation function of the output layer and for the loss function. Formally describe the suitable activation and loss functions.
2. Describe how the output labels should be encoded, according to the output layer described above.
3. Formally describe the Stochastic Gradient Descent (SGD) algorithm and illustrate its hyper-parameters.

EXERCISE 5

Consider a problem represented with an unsupervised dataset $D = \{\mathbf{x}_n\} \subset \mathbb{R}^2$ generated by an unknown Gaussian Mixture Model (GMM) with $k = 3$

1. Formally define the GMM model, describe the parameters of the model, and determine the size of the model (i.e., the number of independent parameters) for the problem considered above.
2. Draw an example of a dataset for the above problem and show graphically the corresponding parameters of the model.