Machine Learning – January 19, 2024

Matricola Last Name First Name

[redacted]

Notes

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: **2 hours**.

EXERCISE 1

1. Provide the definition of *Confusion matrix* for a multi-class classification problem, formally explain the content of component $C_{i,j}$ of the matrix.
2. Provide a numerical example of a confusion matrix for a 3-classes classification problem with a balanced data set including 200 samples for each class (600 samples in total) and an average accuracy around 70% for class 1, 80% for class 2 and 90% for class 3. The matrix **must not be simmetric**. 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 the accuracy of the classifier for the numerical example provided above.

EXERCISE 2

Consider a binary classification problem $X \to \{T, F\}$, with $X = \{T, F\}^3$, i.e. $(x_1, x_2, x_3) \in X$ and $x_i \in \{T, F\}$, and the dataset $D = \{\langle(F,F,F),F\rangle, \langle(F,T,T),T\rangle, \langle(T,T,F),T\rangle, \langle(T,F,T),T\rangle\}$. Consider the two hypothesis $h_1 = (x_1 \wedge \neg x_2 \wedge x_3) \vee x_2$ and $h_2 = (\neg x_1 \wedge x_2 \wedge x_3) \vee x_1$.

1. Determine whether h_1 and h_2 are consistent with D, showing all the passages needed to answer.
2. Assuming the likelihood probabilities $P(D|h_1) = 0.6$ and $P(D|h_2) = 0.8$ and the prior probabilities $P(h_1) = 0.2$ and $P(h_2) = 0.1$, determine the higher a poteriori hypothesis between h_1 and h_2.

EXERCISE 3

Consider a regression problem $f : \Re^d \to \Re$ with a dataset $D = \{(\mathbf{x}_n, t_n)_{n=1}^N\}$, where f is known to be non-linear in x.

1. Describe a linear model for this problem and determine the trainable parameters and the size of the model (i.e., number of trainable parameters).
2. Describe a solution of the problem in terms of least square error minimization. Define the error function corresponeding to the model given above and illustrate a method to find a solution of the optimization problem.

EXERCISE 4

1. Describe the principle of soft margins used by SVM classifiers. Illustrate the concept also with a geometric example.
2. Draw a dataset for binary classification of 2D samples and show two solutions based on SVM with and without soft margin constraints. Choose a proper example that illustrates well the concept, i.e., in which the two solutions are significantly different.

EXERCISE 5

Let $\mathcal{D}$ be a dataset containing the following input values $X = \{(3.3, 1.6), (7.5, 48.2), \ldots, (98.3, 43.5), (87.2, 92.4)\}$ and target values $T = \{0, 2, \ldots, 4, 3\}$.
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.
2. Provide some valid options for the activation functions of the hidden units.
3. Formally describe the Stochastic Gradient Descent (SGD) algorithm and illustrate its hyper-parameters.

EXERCISE 6

1. Describe the K-means algorithm in a **formal** way (i.e., with precise mathematical formulas and equations), including: input and output of the algorithm, its main steps, and the termination condition.
2. Draw a suitable 2-D data set for K-means.
3. **Qualitatively** simulate the execution of K-means in such 2-D data, showing at least three steps of the algorithm and the final output.