

MACHINE LEARNING

EXAM - CONTRÔLE DE CONNAISSANCES (DURATION 1H30)

No document is authorized, nor computers. The answers must be precise and short. French or English is accepted.

1 - INTRODUCTION TO SUPERVISED CLASSIFICATION

Notations. We consider the probabilistic and statistical framework of supervised classification where X is a random vector on \mathbb{R}^d , $d \geq 1$ and Y is a binary random variable with values in $\{-1, +1\}$. A random sample $S_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$, with n independent copies of the pair (X, Y) of joint probability distribution P .

1. What is a classifier?
2. Give a definition of the theoretical problem of supervised binary classification relying upon the definition of risk.
3. Define the empirical risk of a classifier calculated using S_n . Explain the principle of *Empirical risk minimization*.
4. Why the empirical risk minimization may rise issues? Which approach to propose to address this issue?

2. Support Vector Machines

We consider the framework of binary supervised classification.

- ① What optimization problem do we need to solve in the primal space to find the Optimal Margin Hyperplane, e.g. a linear SVM when data are noisy?
[You should not solve the problem in the dual space, only describe the problem in the primal space and indicate the role of each term].
2. Give the definition of a positive definite kernel and explain its key property used in SVM to deal with data non linearly separable.

3. Decision trees and ensemble methods

1. Describe the construction algorithm for a decision tree.
2. When do we find a null training error when building a decision tree?
3. Explain in this framework the notion of overfitting
4. Which hyperparameter do you advise to tune in order to control overfitting?
5. Describe Random Forest [give the algorithm and briefly explain it]

6. Give the definition of importance feature

4 - INTRODUCTION TO DEEP LEARNING

Let $f : \mathbb{R}^d \rightarrow [0, 1]^C$ be the function computed by a one-hidden layer perceptron for a classification problem with C classes.

1. Propose an architecture for f (describe each layer) and define the required functions.
2. Define the optimization problem for the architecture you defined for multiclass classification.
3. Without giving all the details, express which gradients need to be computed for each kind of parameters in the network.

5 - TOWARDS LARGE SCALE

- Among the approaches that you have studied during the lectures or the practical session for supervised classification, which one would you use when you have a very large training set (big data)? Justify briefly your choice.

↳ Deep learning