University of Modena e Reggio Emilia
Department of Physics, Informatics and Mathematics

Master's degree in Computer Science

Machine Learning Fundamentals

Author:
Francesco Mecatti

Academic Year 2024-2025

Table of Contents

1. Introduction

The model explored in this course will be SVM - Support Vector Machine. The course is structured in two units: deterministic optimization techniques for ML (gradient descent et sim.) and stochastic optimization techniques for ML.

1.1 Reference Books

- Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond
- Pattern Recognition and Machine Learning
- Nonlinear Programming, Athena Scientific
- Numerical Optimization, Springer-Verlag
- Optimization Methods for Large-Scale Machine Learning, Bottou et. al.

1.2 Mathematical Framework for ML

As a general framework we can define this formal formulation of learning (supervised, unsupervised) problems:

$$x \in X \subseteq R^m \tag{1.1}$$

$$y \in Y \subseteq R \tag{1.2}$$

Data: $D = (x_i, y_i), i = 1, \dots, N, x_i \in X, y_i \in Y$ Goal: model (prediction function F) that makes predictions on new examples. Namely $F : X \to Y$ that gives labels y to new examples x.

See *MNIST database*

1.2.1 Supervised Learning

Binary classification: $D = (x_i, y_i), i = 1, \dots, N, x_i \in X, y_i \in -1, 1$

Multi-class classification: $D = (x_i, y_i), i = 1, \dots, N, x_i \in X, y_i \in 1, 2, \dots, N$

Regression: $D = (x_i, y_i), i = 1, \dots, N, x_i \in X, y_i \in R$

1.2.2 Unsupervised Learning Tasks

We have x vectors without the associated labels:

- Novelty detection: detect an element not compliant to the training set
- Clustering problems: group the training set into classes

1.2.3 Main Steps of Learning Systems

1. Problem formulation
2. Collect the examples. Following sets are disjunct
   - Training set: learn from examples
   - Validation set: model validation and tuning
   - Testing set: performance estimation of the selected model
3. Represent good examples that emphasize the features of the problems
4. Choose learning algo
5. Test and Validate

1.3 Learning Methodologies

Given a training set, try to infer a scheme that models and fits the data, in a general way: **generalization** property is needed to describe (new) examples that are not in the training set.

Some issues can arise: **overfitting** and **underfitting**.

Overfitting: the model minimizes the error on the training set, involving a large error on new examples

1.3.1 Mathematical Setting

The hypothesis space H is the space of the functions suitable to describe relationship between x and y.

The model f^* must solve $min_{f \in H} \sum_{i=1}^{N} V(y_i, f(x_i)) + \lambda \|f\|^2$

The first part ($V(y_i, f(x_i))$), called "rischio empirico [IT]", represents the error of f on the training set. If this was the only part of the above formula, perfect conditions for overfitting would be created.

The second addend ($\|f\|^2$) measures the complexity of the model. It is required to prioritize low complexity solutions (hence avoid overfitting).

λ is a control term ("parametro di regolarizzazione [IT]") to balance the error on the training set and the complexity of the model.

Stochastic optimization pops out from the need to keep evaluation of error function low. For eg. when N is too large.

Model $f*$ depends on several factors:

- Hypothesis space
- Definition of $\|f\|$ on H

- Definition of loss function $V(y, f(x))$

SVM suggets the choice of H and $f^*(x) = min_{f \in H} \sum_{i=1}^{N} c_i^* K(x, x_i) + b^*$.

Where K is a kernel: linear, polynomial, gaussian. The choice of V depends on the type of learning problem.