

Chapter 2 - Supervised learning

- January 2020
- DOI: [10.1016/B978-0-12-811842-9.00002-9](https://doi.org/10.1016/B978-0-12-811842-9.00002-9)
- In book: Advantages and Pitfalls of Pattern Recognition
- Publisher: Elsevier
- Authors: Horst Langer, Susanna Falsaperla, Conny Hammer

Abstract

In supervised classification, we search criteria allowing us to decide whether a sample belongs to a certain class of patterns. The identification of such decision functions is based on examples where we know a priori to which class they belong. The distinction of seismic signals, produced from earthquakes and nuclear explosions, is a classical problem of discrimination using classification with supervision. We move on from observed data—signals originating from known earthquakes and nuclear tests—and search for criteria on how to assign a class to a signal of unknown origin. We begin with Principal Component Analysis (PCA) and Fisher's Linear Discriminant Analysis (FLDA), identifying a linear element separating groups at best. PCA, FLDA, and likelihood-based approaches make use of statistical properties of the groups. Considering only the number of misclassified samples as a cost, we may prefer alternatives, such as the Multilayer Perceptrons (MLPs). The Support Vector Machines (SVMs) use a modified cost function, combining the criterion of the minimum number of misclassified samples with a request of separating the hulls of the groups with a margin as wide as possible. Both SVMs and MLPs overcome the limits of linear discrimination. A famous example for the advantages of the two techniques is the eXclusive OR (XOR) problem, where we wish to form classes of objects having the same parity—even, e.g., (0,0), (1,1) or odd, e.g., (0,1), (1,0). MLPs and SVMs offer effective methods for the identification of nonlinear decision functions, allowing us to resolve classification problems of any complexity provided the data set used during learning is sufficiently large. In Hidden Markov Models (HMMs), we consider observations where their meaning depends on their context. Observations form a causal chain generated by a hidden process. In Bayesian Networks (BNs) we represent conditional (in)dependencies between a set of random variables by a graphical model. In both HMMs and BNs, we aim at identifying models and parameters that explain observations with a highest possible degree of probability.