

Chapter 1 - Patterns, objects, and features

- January 2020
- DOI: [10.1016/B978-0-12-811842-9.00001-7](https://doi.org/10.1016/B978-0-12-811842-9.00001-7)
- In book: Advantages and Pitfalls of Pattern Recognition
- Publisher: Elsevier
- Authors: Horst Langer, Susanna Falsaperla, Conny Hammer

Abstract

Patterns and objects are described by a variety of characteristics, namely features and feature vectors. Features can be numerical, ordinal, and categorical. Patterns can be made up of a number of objects, such as in speech processing. In geophysics, numerical features are the most common ones and we focus on those. The choice of appropriate features requires a priori reasoning about the physical relation between patterns and features. We present strategies for feature identification and procedures suitable for pattern recognition. In time series analysis and image processing, the direct use of raw data is not feasible. Procedures of feature extraction, based on locally encountered characteristics of the data, are applied. Here we present the problem of delineating segments of interest in time series and textures in image processing. In transformations, we “translate” our raw data to a form suitable for learning. In Principal Component Analysis, we rotate the original features to a system of uncorrelated variables, limiting redundancy. Independent Component Analysis follows a similar strategy, transforming our data into variables independent of each other. Fourier transform and wavelet transform are based on the representation of the original data as a series of basis functions—sines and cosines or finite-length wavelets. Redundancy reduction is achieved considering the contributions of the single basis functions. Even though a large number of features help to solve a classification problem, feature vectors with high dimensions pose severe problems. Besides the computational burden, we encounter problems known under the term “curse of dimensionality.” The curse of dimensionality entails the necessity of feature selection and reduction, which includes a priori considerations as well as redundancy reduction. The significance of features may be evaluated with tests, such as Student’s *t* or Hotelling’s *T*², and, in more complex problems, with cross-validation methods.