In this Chapter, we introduce the main concepts and types of learning, classification, and regression, as well as elaborate on generic properties of classifiers and regression models (regressors) along with their architectures, learning, and assessment (performance evaluation) mechanisms.

1. Introductory Comments

In data mining, we encounter a diversity of concepts that support the creation of models of data. Here we elaborate in detail on learning, classification and regression as being the most dominant categories of developments of a variety of models.

1.1. Main Modes of Learning from Data-Problem Formulation

To deal with huge databases, we first describe several fundamental ways to complete their analyses, build underlying models, and deliver major findings. We present the fundamental approaches to such analyses by distinguishing between supervised and unsupervised learning. We stress that such a dichotomy is not the only taxonomy available since a number of interesting and useful alternatives lie somewhere in-between the two thus forming a continuum of options that could be utilized. The relevance and usefulness of such alternatives are described below.

1.2. Unsupervised Learning

The paradigm of unsupervised learning, quite often referred to as clustering involves a process that automatically reveals (discovers) structure in data and does not involve any supervision. Given an $N$-dimensional dataset $X = \{x_1, x_2, \ldots, x_N\}$, where each $x_k$ is characterized by a set of attributes, we want to determine the structure of $X$, i.e., identify and describe groups (clusters) present within it. To illustrate the essence of the problem and build some conceptual prerequisites, let us consider the examples of two-dimensional data shown in Figure 4.1. What can we say about the first one, shown in Figure 4.1(a)? Without any hesitation, we can distinguish three well-separated spherical groups of points. These are the clusters expressing the structure in the data.

The clusters could also exhibit a very different geometry. For instance, in Figure 4.1(b), we see two elongated structures and one ellipsoidal cluster. All the clusters are well separated and clearly visible. In Figure 4.1(c), the structure is much less apparent; the clusters overlap significantly. Perhaps the two that are close to each other could be considered to form a single cluster. In
Figure 4.1. Examples of two-dimensional data (all in the two-dimensional feature space $x_1 - x_2$) and the search for their structures: geometry of clusters and their distribution in the two-dimensional data space.

Figure 4.1(d), the shapes are even more complicated: a horseshoe exhibits far higher geometric complexity than the two remaining spherical clusters.

The practicality of clustering is enormous. In essence, we perform clustering almost everywhere. Clusters form aggregates or, to put it differently, build an abstraction of the dataset. Rather than dealing with millions of data points, we focus on a few clusters and doing so is evidently very convenient. Note, however, that clusters do not have a numeric character; instead we perceive “clouds” of data and afterwards operate on such structures. Hopefully, each cluster comes with well-defined semantics that capture some dominant and distinguishable parts of the data. Consider, for instance, a collection of transactions in a supermarket; the data sets generated daily are enormous. What could we learn from them? Who are the customers? Are there any clusters – segments of the market we should learn about and study for the purpose of an advertising campaign? Discovering these segments takes place through clustering. A concise description of the clusters leads to an understanding of structure within the collection of customers. Obviously, these data are highly dimensional and involve a significant number of features that describe the customers. A seamless visual inspection of data (as we have already seen in the case of examples in Figure 4.1) is not feasible. We need a powerful “computer eye” that will help us to explore the structure in any space even highly-dimensional one. Clustering delivers an algorithmic solution to this problem.

In spite of the evident diversity of clustering mechanisms and their algorithmic underpinnings, the underlying principle of grouping is evident and quite intuitive. We look for the closest data points and put them together. The clusters start to grow as we expand them by bringing more points together. This stepwise formation of clusters is the crux of hierarchical clustering. We could look at some centers (prototypes) and request that the data points be split so that the given distance function assumes its lowest value (i.e., their similarity or dissimilarity is highest).

Here we note that all strategies rely heavily on the concept of similarity or distance between the data. Data that are close to each other are likely to be assigned to the same cluster. Distance impacts clustering in the sense that it predefines a character of the “computer eye” we use when searching for structure. Let us briefly recall that for any elements – patterns $x, y, z$ (treated formally