

Learning on pairs with tree-based methods

Pierre Geurts (Montefiore Institute, ULg)

April 16, 2010

In many machine learning applications, one actually wants to learn a classification or regression model on pairs of objects, each object of the pair being potentially drawn from a different universe of objects and described by a distinct input feature vector. In this talk, we will show how to leverage tree-based methods to learn in this context. We will first consider the case when both objects in the pair are coming from the same universe (described by the same set of input features), and the output can be assimilated to a kernel or an euclidean distance defined on the pairs. In this case, it is possible to formulate the problem as a learning problem on single objects that can be addressed elegantly with tree-based methods. We will then discuss several possible extensions of tree-based methods for the case of arbitrary relations, possibly defined on two different sets of objects. Illustration of these approaches will be given on bioinformatics applications.