How SLT offer math basic framework to solve the problem of binary classification in Machine Learning?

Statistical learning theory (SLT) provides the theoretical basis for many modern machine learning algorithms and is arguably one of the most well-developed branches of artificial intelligence in general.

The most well-studied problem in machine learning is the classification problem. To train an algorithm, it is given some training examples $(X_1, Y_1)...(X_n, Y_n)$, that is, pairs of objects with the corresponding category label. The goal is then to find a mapping $f : X \rightarrow Y$ that makes as few errors as possible. That is, among all the elements in X, the number of objects assigned to the wrong category is as small as possible. Classification is an example of a supervised learning problem: the training examples consist of both the instances X_i and the correct labels Y_i on those instances. The goal is to find a functional relationship between the instances and the outputs. This setting is called supervised because, at least on the training examples, the learner can estimate whether the answer is correct, that is, the learner is supervised.

In supervised learning, we are dealing with an input space (instance space, object space) X and an output space (label space) Y. In the case of binary classification, we identify the label space with the set {-1,+1}. That is, each object can belong to one of two classes, and by convention we denote these classes as -1 and 1. The learning question is reduced to the question of estimating a functional relationship of the form $f : X \rightarrow Y$, that is, a relationship between the input and the output. Such a mapping *f* is called a classifier. A classification algorithm (classification rule) is a procedure that takes training data as input and outputs a classifier *f*. We assume that there is a joint probability distribution P over $X \times Y$, and the training examples (X_i, Y_i) are sampled independently from this distribution P. This type of sample is often denoted as an iid (independent and identically distributed) sample.

While SLT itself is not a classification technique, it can play a supportive role in solving binary classification problems by pre-processing data and extracting relevant features. You would still need to combine SLT with a dedicated classification algorithm to achieve the final classification results.

Vadim Ryzhov
5130203/20101