Binary classification

Binary classification is a fundamental problem in the field of machine learning and statistics, focusing on the task of categorizing data points into one of two distinct classes. In binary classification, we deal with an input space (space of instances) X and and output space (label space) Y. We identify the label space with the set $\{-1, +1\}$. The task involves assigning each object from space of instances to one of these two classes. The issue of learning can be simplified to estimating a functional relationship represented as $f : X \to Y$. This type of mapping f is referred to as a classifier. In order to do this, we get access to some training points $(X_1, Y_1), ..., (X_n, Y_n) \in X \times Y$, drawn from an unknown probability distribution $P(X, Y)$, the goal is to find function f that generalizes well to unseen data, minimizing misclassification errors.

The risk of a function is the average loss over data points generated according to the underlying distribution $P : R(f) := E(l(X, Y, (X)))$. In other words, the risk of a classifier f is the expected loss of the function f across all points $X \in X$. This risk measures the number of elements in the instance space X that are misclassified by the function f. Of course, a function f is a better classifier than another function g if its risk is smaller, that is if $R(f) < R(g)$. To find a good classifier f we need to find one for which $R(f)$ is as small as possible. The best classifier is the one with the smallest risk value $R(f)$.

Statistical learning theory

Statistical Learning Theory (SLT) provides a foundational framework for understanding and addressing problems in machine learning, including binary classification. Here's how SLT contributes to this area:

1. Understanding Learning as a Statistical Problem:

   SLT frames the learning process as a statistical problem, where the goal is to infer a function from input features X to output classes Y.

   Binary classification specifically involves determining whether an instance belongs to class 0 or class 1.

2. Hypothesis Space:

   In SLT, a hypothesis space is defined, which consists of all possible functions that can be used to map inputs to their corresponding outputs.

   The choice of hypothesis space is critical; it should be complex enough to capture the underlying data distribution but not so complex that it leads to overfitting.

3. The Empirical Risk Minimization Principle

   SLT motivates the use of the Empirical Risk Minimization (ERM) principle, which aims to minimize the average loss (error) on training data.

For binary classification, this often involves using loss functions such as the binary cross-entropy or 0-1 loss, which track how well the classifier performs on the labeled training examples.