Galitskikh Darya, 5130203/20102

Binary Classification

Binary classification is a fundamental problems in supervised learning, where the objective is to categorize instances into one of two predefined classes, typically labeled as +1 and -1. In machine learning, the goal is to learn a function, called a classifier, that map inputs from a given input space X to labels in a label space Y={-1, +1}. This function is derived from a set of training examples consisting of pairs(X1, Y1),…,(Xi,Yi) where each Xi is an instance, and each Yi is the corresponding label.

Key Concepts:

• Training Data: The examples used to learn the classifier are assumed to come from an unknown joint probability distribution P(X,Y). The examples are drawn independently from this distribution, a condition referred to as independent and identically distributed.

• Classifier: The classifier f: X→ Y aims to minimize classification errors on both seen (training) and unseen (test) data. The challenge is to generalize from the training data to make accurate predictions on new data.

• Label Noise and Overlapping Classes: In real-world scenarios, the labels may not always be deterministic. Noise in the data, such as human errors in labeling, or natural overlaps between classes (e.g., predicting gender based on height), complicates classification.

• Conditional Probability: The key quantity of interest in binary classification is the conditional probability $\eta(X) = P(Y = 1|X = x)$, which represents the likelihood that an input belongs to class +1. If $\eta(x) \geq 0{,}5$, the classifier should assign label +1 to x; otherwise, it should assign -1.

• Bayes Classifier: The optimal classifier under the assumption that the probability distribution is known is the Bayes classifier, which minimizes the risk (expected loss). However, since the probability distribution is unknown in practice, constructing the Bayes classifier directly impossible.

SLT Framework for Solving Binary Classification:

Statistical Learning Theory (SLT) offers a mathematical framework for solving the binary classification problem. SLT works based on several key assumptions:

1. Unknown distribution: We don't know the true distribution of the data. Our task is to get it out of the training sample.

2. Risk minimization: The goal is to minimize the risk that represents the expected classification error across the entire input space. However, since we do not know the true distribution, we approximate the risk from the training data.

3. Empirical Risk Minimization (ERM): The key principle of SLT is empirical risk minimization (ERM). We choose a classifier that minimizes the error in the training sample. However, this can lead to overfitting when the model works well on the training sample, but poorly on new data.

In conclusion, SLT forms a mathematical framework for understanding and solving binary classification problems. SLT principles such as ERM provide the theoretical guarantees needed to create algorithms that generalize well to new data.