Report about Statistical Learning Theory (SLT)

Explanation of the binary classification problem in a formal mathematical way

In supervised learning, binary classification involves estimating a function $f: X \to Y$, where X is the input space and Y represents the label space, identified as Y = {−1,+1}. The primary goal is to derive a classifier that accurately predicts the label for new instances based on training data, denoted as $(X_1, Y_1), \dots, (X_n, Y_n) \in X \times Y$. This classifier is constructed from a set of training examples sampled independently from an unknown joint probability distribution P over X×Y, following the independent and identically distributed (iid) assumption.

Key Challenges in Binary Classification.

1. Uncertainty in Labels: the labels Y_i are not deterministic functions of the instances X_i. This non-determinism can arise from two primary sources: label Noise and overlapping classes.

The conditional likelihood $\eta(x) = P(Y = 1|X = x)$ captures the probability of an instance $X = x$ being classified as +1. If $\eta(x)$ approaches 0.5, it indicates a high degree of uncertainty in the classification, resulting in increased prediction errors.

2. Independent Sampling: the assumption of iid sampling is crucial for the validity of many learning algorithms. However, this assumption may not hold in all contexts, such as in drug discovery, where training examples are often hand-selected, violating the independence requirement.

3. Unknown Distribution: the probability distribution PPP is unknown during training, making it impossible to directly compute the optimal classifier, known as the Bayes classifier. The Bayes classifier is defined as:

$$f_{Bayes(x)} = \begin{cases} +1 \; if \; P(Y = 1|X = x) \geq 0{,}5 \\ -1 \; otherwise \end{cases}$$

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

Statistical Learning Theory (SLT) provides a mathematical framework to address these challenges in binary classification by focusing on the following aspects.

Risk Minimization. SLT introduces the concept of risk R(f), defined as the expected loss over the distribution P:

$$R(f) = E_{(X,Y)\sim P}[l(X, Y, f(x))]$$

where l is a loss function measuring classification errors. The objective is to construct a classifier f that minimizes this risk, ideally approximating the Bayes classifier.

Learning Bounds. SLT establishes theoretical bounds on the generalization error, which quantifies the performance of the classifier on unseen data. By analyzing the sample size and the complexity of the hypothesis class, SLT offers insights into how well a classifier trained on a finite sample can generalize to the broader input space.

Empirical Risk Minimization. In practice, SLT suggests using empirical risk minimization, where the risk is approximated using training data. The challenge lies in choosing a suitable hypothesis space F from which the function f is selected, balancing complexity and fit to avoid overfitting.

Agnostic Learning. SLT operates in an agnostic setting, allowing for flexibility in assuming the distribution P. This agnostic approach is advantageous as it does not impose strict assumptions about the nature of the data, making SLT applicable across various domains.