# GARVEN KALULU, 5130203/20101, SECOND ASSIGNMENT.

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

Binary classification refers to the task where the goal is to assign a label from a set $\{-1,1\}$ to an input object. Formally, given an input space X (the space of instances) and a label space Y= $\{-1,1\}$ (the set of possible outputs), the goal is to learn a function $f:X\rightarrow Y$ that maps an input instance to one of the two labels. The objective is to find a classifier f that minimizes the number of classification errors.

Mathematically, given a set of training data $\{(X1, Y1), (X2, Y2),\ldots,(Xn,Yn)\}$, where each $Xi \in X$ and $Yi \in \{-1,1\}$, the task is to construct a function f that generalizes well to new, unseen data.

SLT offers a way to analyse and solve the problem of binary classification by building a framework that ensures the chosen classifier generalizes well to unseen data. It does so through several key concepts:

1. **Agnostic Learning**: In SLT, no assumptions are made about the underlying distribution P(X,Y)P(X, Y)P(X,Y). Instead, it considers any possible distribution, allowing the framework to be general and flexible.
2. **Independent and Identically Distributed (IID) Assumption**: SLT assumes that the training examples are sampled independently from the same distribution. This allows the empirical risk to be used as a good estimate of the true risk, under certain conditions.
3. **The VC Dimension and Uniform Convergence**: One of the central ideas of SLT is **uniform convergence**, which ensures that the empirical risk converges to the true risk as the sample size grows. The **Vapnik-Chervonenkis (VC) dimension** is a measure of the complexity of the function class from which the classifier is chosen. SLT shows that if the function class has a small VC dimension, then uniform convergence can be guaranteed, leading to generalization.
4. **Generalization Bounds**: SLT provides theoretical **generalization bounds** that quantify how the empirical risk relates to the true risk. These bounds ensure that, with high probability, the classifier learned from the training data will perform well on unseen data.

Lastly, SLT provides the mathematical foundation for understanding and solving the problem of binary classification in machine learning. It offers tools like the VC dimension and generalization bounds that ensure the classifier not only fits the training data but also generalizes well to new, unseen data. This is crucial for constructing robust machine learning models.