Data driven Optimization techniques for a single cost optimization problem

Qazi Zarif Ul Islam

November 19, 2024

1 Abstract

2 Introduction

2.1 The central tenets of Machine Learning

Machine learning (ML) is a field that primarily focuses on predicting some outcome, given some data. The central tenet of using such a framework is that the input variables that are used are somehow correlated with the outcome. For example, in the popular iris dataset, the input variables are the sepal and petal characterisitcs and the outcome is one of three species of the iris flower. The secondary tenet is that no observation made in this world is by itself the true representation of what is being observed and so *all observations* are noisy (contain noisy information). Thus, it follows that a single prediction made from a single noisy observation is also noisy (Noisy input gives noisy output). However, time and time again, such predictions have proven helpful. For example, a $\geq 90\%$ accurate transformer is deemed capable enough to complete incomplete passages or poems, computer vision systems are used to identify objects, understand human sentiments, etc. These wonders are due to the fact that all phenomena are generated by probability distributions and a random sample approaches a limiting probability distribution as the size of the sample increases. Thus, even though the learner is unaware of the underlying limiting distribution, as we increase the sample size, the particular error-correction algorithm of the learner leads it towards the true prediction.

2.2 Quantitative decisions in socal policy making

ML techniques have been used primarily to make tasks more efficient, not to induce social policy and decision. We trust qualitative arguments when deciding where to build a hospital or where to put more health care funding. One explanation for this is that we can determine biases in an argument instantly and doing the same for a learner is much harder, simply due to the secondary tenet (see 2.1). Another reason for the insuitability of ML techniques in this area is that the predictions do not take into account *its own effects* on downstream decisions. But where we have not trusted predictions, we have trusted optimization and statistical testing. Optimization techniques have been heavily relied on when the answer to a problem is deemed to be deterministic. Examples of such problems include The Travelling Salesman proble, The Knapsack problem, etc. Optimization problems differ from Machine Learning problems in that the problem is not solved using data but rather a fixed set of conditions. This fixed set contains a function relating the inputs to the output and equations that give the 'feasible answer region'. But there are numerous problems where we don't have a function that gives the input-output relationship but yet we want to optimize the inputs for either maximizing or minimizing the output based on some constraints on the inputs. For example,

Another area that contains techniques that lead to decisions from data as opposed to predictions from data is statistical hypothesis testing.