2021-11-20

Information Theoretic Error Bounds on NISQ Learning Systems

B.Tech Project I

Sankalp Gambhir

Email   sgambhir@iitb.ac.in

Abstract   Bla bla.

Contents

1 Introduction

There has been long standing interest in constructing systems capable of learning from experience since even before computers in their modern form have existed. In the last few decades, with computing power skyrocketing exponentially coupled with leaping advances in theory of learning systems and statistical inference, these problems became tractable and eventually came into use ubiquitously. With applications ranging from facial detection systems for surveillence to identifying cosmic objects for cosmology, they have found widespread adoption in industry and academia. These systems circumvent the need to produce a precise mathematical model for the problem at hand, exploiting general techniques to instead infer a model from available data. With their advent, however, has come an ever rising need for computing power to facilitate their operation. This has found data centers of unprecendented scales consuming enormous amounts of power to provide the instant predictions we've come to rely on.

With snowballing energy and space requirements of classical computers in the form of GPU clusters and Application Specific Integrated-Circuits (ASICs), there has been a spark of interest in offloading this computation onto quantum computers, which, till recently, have largely remained a rare species spotted only in labs surrounded by helium-cooled superconductors and white-coated predators. Current scales of available quantum computers, however, still lack the power required to fully tackle these challenges while main-

taining reliable error-levels or adding their own error checking and correction. This has motivated using quantum computers to run bottlenecked computational subroutines with classical control systems. These systems generally lack error correction, and thus earn themselves the title of 'noisy'. These form the basis of computation considered in this thesis, Noisy Intermediate-Scale Quantum (NISQ) computers.

Connecting a quantum computer to a classical puppeteer is not expected to come without its own issues either. It constrains the architecture and is itself bottlenecked on both ends, first by the parameter transfer and configuration from the classical to the quantum, then finally by the detectors on the quantum side to the classical. In this thesis, we focus on the former, discussing the limits of computation and computational precision achievable with this hybrid architecture.

1.1 Structure

In section 2, definitions and relevant results in classical computing, physics, and quantum information are presented. [Note: Extend this.]

1.2 Outline of New Results

[Note: Add summary of results at the end.]

2 Preliminaries

2.1 Classical Computing

2.1.1 Learning Problem

Learning [1] can be broadly defined as attempting to learn the input-output pattern given sample data. For this thesis, we consider three major categories of learning problems:

- Binary Classification — input points in a chosen domain, and a binary output label for each point.

- Multi-Label Classification — input points in a chosen domain, and one of n labels as output for each point.
- Regression — input points in a chosen domain with real-valued output.

2.1.2 Classification Problem

We take as input elements $\{x_i\}$, generally called *feature vectors*, in a chosen domain X called the *feature space* and output an element from a finite set $L = l_i$ of labels. [Note: Add a nice picture]

The problem is called binary classification if $|L| = 2$.

Formally, we attempt to learn a function $f : X \to L$ given a set of inputs in the domain, and possibly paired output labels.

The problem proceeds in two manners given the form of inputs: if provided input-output pairs, the problem is called a *supervised learning problem*, while attempting to learn a set of labels given just (clustered) inputs is called *unsupervised* learning. We focus on supervised classification here.

The set of input-output pairs provided is called the *training data*.

Given the difficulty of working with discretized domains, the input domain is generally converted to be a subset of a Euclidean space, using a suitable *embedding function*.

2.1.3 Embedding

An embedding of X in Y is a function $f : X \to Y$ that is injective and structure-preserving. The exact restrictions on the map to be structure-preserving depend on the structures of the domain and the codomain [2]. It is denoted here as $f : X \hookrightarrow Y$.

For example, a topological embedding, i.e., the embedding of a topological space, will be restricted to preserve its associated structure of open sets. A field embedding, similarly, will be restricted to preserve the field operations $+$ and $\times$.

For a given feature space X, it is generally embedded into $\mathbb{R}^n$ for some n.

2.1.4 Linear Classification

Classification generally proceeds by producing linear functions as candidate (supplemented with a discretization function) labelling functions and fitting them to the training data. For simplicity, we first restrict the discussion to binary classifiers.
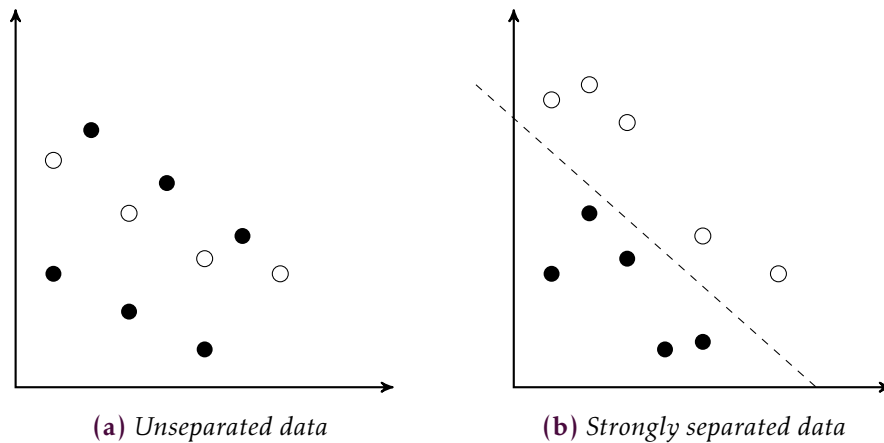

(a) *Unseparated data*   **(b)** *Strongly separated data*

Figure 1: *The magic of the (strong) separation axiom.*

Given a set of points which, due to an embedding, may be assumed to be in $X \subseteq \mathbb{R}^n$, attempting to classify them may still be an arduous task if the spatial regions corresponding to the labels are intertwined. Thus, to make the problem tractable, we restrict the data to be *strongly* separated, i.e., any labelling function $f : X \to L$ that agrees with the training data [Note: try to write separation properly].

2.2 Quantum Computing

References

[1]   Nello Cristianini, John Shawe-Taylor et al. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[2] Hanamantagouda P Sankappanavar and Stanley Burris. 'A course in universal algebra'. In: *Graduate Texts Math* 78 (1981).