CS181P Final Project Proposal #1

Santi Santichaivekin, Jingyi Liu, Jacob Boerma, Lorraine Zhao
Harvey Mudd College and Claremont McKenna College
Claremont, California, USA

1   Team

The team includes Santi Santichaivekin, Jingyi Liu, Jacob Boerma, and Lorraine Zhao. The team name is Go Pika Pika! The team logo is provided below.

2   Project

We will attempt the second project listed on the project ideas:

> ML, Search, and Compression. It has been shown that ML is a form of search (we will see this during the course, for example in Montañez (Chapter 3)), and also shown that ML can be viewed as a form of compression (as in minimum description length approaches). Can we show compression is a form of search, search a form of compression, search a form of ML, or compression a form of ML, to make a triangle of two-way equivalences? Or taking a subset of these equivalences, can we apply a set of results from one domain to the other, such as forming no free lunch theorems for compression?

We hypothesize that all the three problems can be phrased as one another. If we prove the three-way equivalency among the three problems, then we can apply different theories from one field to another. Also, this problem is applicable to the real world, where we can solve actual real world problems by directing them as either ML, search, or compression.

3   Initial Plan of Attack

We will start with understanding the concepts of ML, data compression, and search individually, and then trying to prove the No Free Lunch theorem for data compression among other theorems of search that might be applied to data compression. We can then look for algorithms for data compression and obtain insights about how it might be framed as a machine learning problem and search.

4   Separation of Work

For now, we will each read the papers and get on the same level of understanding. Once we have a good understanding of ML, compression, and search, we might split up work based on sub-problems that appear. We have divided the readings to different team members, so that one person can focus on one or two readings, and then update the rest of the team members on what he/she learned.

- Jacob will read about No Free Lunch on a high level, and will try to understand the Famine of Forte paper and how we might apply the same concepts to compression schemes.
- Santi will be trying to understand the No Free Lunch Theory and apply it to compression.
- Lorraine will go over her part of the readings and get a better understanding of Machine Learning and Searches, and try to link the two by establishing an equivalent relationship.
- Rose will be reading about data compression algorithms and see how to generalize them as search problems.

5   Literature Review

We have found some papers and textbooks that will be useful for our research. We will read the suggested readings for the next few lectures in class, as well as the following papers/articles.

Why ML Works `http://www.cs.cmu.edu/~gmontane/montanez_dissertation.pdf` [Lorraine]

Famine of Forte `https://arxiv.org/pdf/1609.08913.pdf` [Jacob]

Probabilistic machine learning and artificial intelligence `https://www.nature.com/articles/nature14541` [Rose]

Generalization as Search `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.5764&rep=rep1&type=pdf` [Lorraine]

Data Compression Explained `http://mattmahoney.net/dc/dce.html#Section_13` [Rose]

No Free Lunch For Optimization `https://ti.arc.nasa.gov/m/profile/dhw/papers/78.pdf` [Santi] [Jacob]