Chapter 1

Introduction

1.1 Machine Learning

Machine learning is the general task of finding patterns given a set of data. The methods by which these tasks are accomplished range from the simple linear regression to more complex neural networks. Machine learning problems fall into primarily two categories: supervised learning and unsepervised learning.

In the case of supervised learning, we are given a set of inputs, $\{x\}_{i=1}^N \in X$ and outputs, $\{y\}_{i=1}^N \in Y$ to some unknown function, f. The goal is then to determine what that function is. This is usually not feasible due our model for the function not being complex enough, or for being too complex (this is known as overfitting, discussed in section **??**.) The objective is therefore simplified as finding the function, $g \in M$ where M is some set of model functions, and where g minimizes some loss function, $L(g, x, y)$. That is:

$$\arg\min_{g \in M} L(g, x, y) \tag{1.1}$$

In linear regression for example, M is the set of linear functions mapping X to Y and the loss function is $L(g, x, y) = \frac{1}{N} \sum_{i=1}^N ||g(x_i) - y_i||^2$

In the case of unsupervised learning, we are given only some set of data points, $\{x\}_{i=1}^N$ and asked to find some pattern in the data. Finding a patterm can consist

of finding clusters of data points which are "close" together by some measure, or finding some lower dimensional representation for the data, essentially a problem of lossy compression. Usually algorithms work by minimizing some loss function so that techniques can be borrowed from supervised learning, though these are not necessarily measures of the algorithm's success. One well known example of unsupervised learning is principal component analysis and its cousin singular value decomposition. Given some data in the form of a matrix X, singular value decomposition can find a matrix, $\hat{X}$ of rank $r < rank(X)$ such that $||X - \hat{X}||_F$ is minimzed. This has the effect of finding a lower dimensional subspace of X which contains as much information as possible and therefore tells us which dimensions are "important."

1.1.1 Overfitting

Bibliography

[1] G. Cybenko, "Approximation by superpositions of a sigmoidal fucntion," *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303–314, 1989.