CSCI 632 Notes

Clay L. McLeod

January 25, 2016

1 Machine Learning Overview

1.1 Supervised Learning

An **observation** is a d-dimensional vector X such that $X \in \mathbb{R}^d$.

The unknown nature of observation is called a **class**. We denote it by Y where $y \in \{1, 2, ..., M\}$. For the purpose of this course, only discrete classes are considered (no regression).

The goal is to create a function $g(x) : \mathbb{R}^d \to \{1, ..., M\}$ $g(x)$ one's guess of y given x. The classifier is $g(x)$. If $g(x) \neq y$.

Questions:

1. How does one construct a good classifier?
2. How good can a classifier be?
3. Is classifier A better than classifier B?
4. Can we estimate how good a classifier can be?
5. What is the best classifier?

The answer to all of these questions is yes: there are ways to find an upper bound on the performance of each algorithm and evaluate it empircally.

1.2 Unsupervised Learning

Same definition for an observation, except we don't have labels for the class in X. What approaches might this help us tackle?

- Clustering

- Dimensionality reduction

Clustering

Unsupervised learning is directly related supervised learning. For instance: feature selection is probably the most important part of designing Machine Learning algorithms. Unsupervised learning helps us find good features for supervised learning algorithms.

Dimensionality reduction

As you increase the number of dimensions, you loss the ability to distinguish between two examples. Also, run time increases exponentially.

1.3 Semisupervised Learning

Partially labelled data where we try to gain some intuition. Usually involves a cost function instead of a solution set.

1.4 References

1. *A Probability Theory of Pattern Recognition* for Theoretical Design
2. *Machine Learning* for History of ML
3. *The Elements of Statistical Learning* for Statistical Vantagepoint
4. *Pattern Recognition and Machine Learning* (Textbook)
5. *Kernel Methods for Pattern Analysis* for Kernel Methods