Introduction to Machine Learning

Louis Becker

March 21, 2018

This markdown is meant to complement the available slides, and will have the odd link here and there to some related material.

1. Introduction

Examples

Links to some stuff mentioned:

- FiveThirtyEight
- Watson

Examples of statistical learning problems:

- Identify the risk factors for prostate cancer.
- Classify a recorded phoneme based on a log-periodogram.
- Predict whether someone will have a heart attack on the basis of demographic, diet and clinical measurements.
- Customize an email spam detection system.
- Identify the numbers in a handwritten zip code.
- Classify a tissue sample into one of several cancer classes, based on a gene expression profile.
- Establish the relationship between salary and demographic variables in population survey data.
- Classify the pixels in a LANDSAT image, by usage.

The Supervised Learning Problem

Where do we start? With defintitions, obviously:

- Y - Outcome variable (dependent variable, response variable, target variabel, etc.)
- $\boldsymbol{X}$ - vector of p predictor measurements (inputs, regressors, covariates, features, independent variables)
- *Regression Problem* - Problem where Y is defeined as a **quantitative** variable
- *Classification Problem* - where Y is defined as a **qualitative** or **indicator** variable, taking values in a finite, unordered set
- *Training Data* - Observations $(x_1,y_1),\ldots,(x_N,y_N)$