Machine Learning Notes

Spring 2019

Learning Materials

- Andrew Ng's Machine Learning Course - coursera.org
- The Hundred-Page Machine Learning Book - Andriy Burkov
- Machine Learning Guice Podcast - OCDevel LLC
- ....

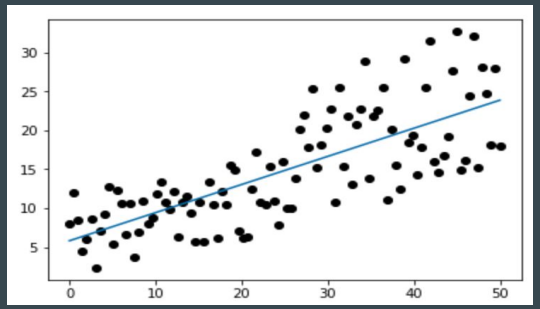
Machine Learning

- Subfield of CS and AI
- Solving a practical problem with a statistical model
  1. Gathering dataset
  2. Algorithmically building a statistical model of the data
- Supervised
  - $Y = f(x)$
- Unsupervised
  - Model data structure
- Semi-supervised
  - Mixture
- Reinforcement

Supervised Learning

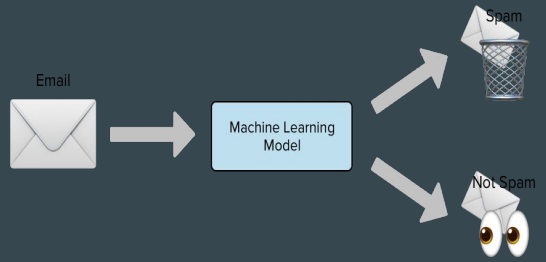
Regression

- **Target** Prediction.
- Map input variables (**features**) to some continuous function (**hypothesis**).



Classification

- **Label** assignment.
- Map **features** into discrete categories or **classes**.



Multivariate Linear Regression

- Dataset comprised of several features.
- Use the features to model a function describing the data.
  - Similar goal of univariate regression
  - Numeric guess.
- Requirements:
  - Cost function, gradient descent OR normal equation

What's Needed

Hypothesis function:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n$$

- Continuous function to model the data

Loss function:

$$J_{\theta_n}(x, y) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

- Prediction & actual deviation

Gradient Descent or Normal Equation:

$$\frac{\partial J_{\theta_n}}{\partial \theta_0}, \frac{\partial J_{\theta_n}}{\partial \theta_{n \neq 0}}$$

- Minimize loss function
- Updates Θ parameters
- Learning rate and epochs

$$\theta = (X^T X)^{-1} X^T Y$$

Gradient Descent vs. Normal Equation

- Normal has no epochs, less math, one step, good up to 1000 features, fast
- Gradient Descent for more sophisticated models, more control (number of epochs, learning rate), good up to hundreds of thousands of features