Coursera- Machine Learning
May 2019
Taught by Prof. Andrew Ng

Janeshi99

Summary

Supervised learning

- linear regression, logistic regression, neural network, SVMs

Unsupervised learning

- k-means, PCA, Anomaly detection

Special applications/special topics

- Recommender systems, large scale machine learning

Advice for building a machine learning system

- bias/variance, regularization, deciding what to work next, evaluation of a learning algorithm, learning curves, error analysis, ceiling analysis

Week 1

<u>Intro</u>
Definition of ML

- A program learns from experience (E) w.r.t task(T) and performance measure (P) if its performance on T improves with more E.

- With supervised learning, we know what our answers are as a relation of input and output. But with unsupervised learning, we have little idea about the result.

Cost function

- $$h_\theta(x) = \theta_0 + \theta_1 x$$
  our goal is to minimize the cost function, which is calculated as square error
  $$\min_{\theta_0,\theta_1} J(\theta_0, \theta_1)$$
- where the error function is defined as
  $$J(\theta_0, \theta_1) = \frac{1}{2m} \sum (h_\theta(x^{(i)} - y^{(i)})^2)$$

Linear regression

- Repeat until converge{ $\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$ for $j = 0, 1$ }
- Note that the update is simultaneous :
- $$\text{temp}_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$
  $$\text{temp}_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$
  $$\theta_0 := \text{temp}_0$$
  $$\theta_1 := \text{temp}_1$$
- if we compute the derivative we get Repeat until converge{
  $$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)$$
  $$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} ((h_\theta(x_i) - y_i) x_i)$$
  }
- α is the learning rate.
- we use linear regression algorithm to updates the parameters until we arrive at the minimal cost.

Week 2

Multi-feature linear regression

- Hypothesis
$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \ldots \theta_n x_n$$
- convenience $\forall x$, $x_0 = 1$, so that $h_\theta = \sum_{i=0}^{n} \theta_i x_i$
- $$x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n+1} \text{ and that } \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$
- Hypothesis can be represented