Machine Learning (Andrew Ng) Notes #1

标签（空格分隔）： Coursera ML DL AI Andrew_Ng Stanford

A. Supervised Learning

定义: 已知数据有确定对应标签，通过学习现有对应关系预测新对应关系。例如：线性回归

1. Classification: 数据对应**diecrete 类别**，如鉴别良性肿瘤或恶性肿瘤
2. Regression: 数据对应**continous 数据**，如通过房子大小预测售价

B. Unsupervised Learning

定义： 已知数据并无确认标签，需通过分析找出现有数据规律。例如：将新闻按照内容相似性归类 (clustering)

C. Class notation:

m: number of training examples
x's: 'input' variable / features
y's: 'output' variable / 'target' variable
(x,y): one training example
(x^{i}, y^{i}): the i^{th} training example

D. Sample cost function

Hypothesis Function:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Parameters:

$$\theta_0, \theta_1$$

Cost Function:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2$$

Goal:
Minimize:

$$J(\theta_0, \theta_1)$$

E. Gradient Descent algorithm

Definition:

The idea is to use a fixed jumping distance * slope ($\frac{\partial}{\partial \theta_j}$) to itereate to the local minimum of θ_j.

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

$$\alpha = learning\_rate$$

Notes: Need to simultaneously update both θ_0 and θ_1

F. Applying gradient descent into cost function

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_0(x^i) - y^i)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_0(x^i) - y^i) * x^i$$

The cost function can also be directly solved with **normal equation method**, however, gradient descent will scale better at large data sets.