Linear Regression Model (one variable)

Hypothesis H: $h_\theta(x) = \theta_0 + \theta_1 x$

Cost Function $J(\theta)$: $J(\theta_0,\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left[h_\theta(x^{(i)}) - y^{(i)}\right]^2$

training data $(x^{(1)}, y^{(1)}), \cdots, (x^{(m)}, y^{(m)})$

$\Downarrow$

$\underset{\theta_0,\theta_1}{\text{minimize}}\ J(\theta_0,\theta_1)$

$$\frac{\partial}{\partial\theta_j} J(\theta_0,\theta_1) = \frac{\partial}{\partial\theta_j}\frac{1}{2m}\sum_{i=1}^{m}\left[h_\theta(x^{(i)}) - y^{(i)}\right]^2$$

$$\left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2$$

when $j=0$: $\frac{\partial}{\partial\theta_0} J(\theta_0,\theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(\theta_0+\theta_1 x^{(i)} - y^{(i)}\right)$

when $j=1$: $\frac{\partial}{\partial\theta_1} J(\theta_0,\theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(\theta_0+\theta_1 x^{(i)} - y^{(i)}\right)\cdot x^{(i)}$

Gradient descent algorithm

Repeat until convergence {

$\theta_0 := \theta_0 - \alpha\frac{1}{m}\sum_{i=1}^{m}\left[h_\theta(x^{(i)}) - y^{(i)}\right]$

$\theta_1 := \theta_1 - \alpha\frac{1}{m}\sum_{i=1}^{m}\left[h_\theta(x^{(i)}) - y^{(i)}\right]\cdot x^{(i)}$ }

Simultaneously update θ_0, θ_1

Correct:

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

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

$\theta_0 := temp\_\theta_0$

$\theta_1 := temp\_\theta_1$

Incorrect:

$temp\_\theta_0 := \cdots$

$\theta_0 := temp\_\theta_0$

$temp\_\theta_1 := \cdots$

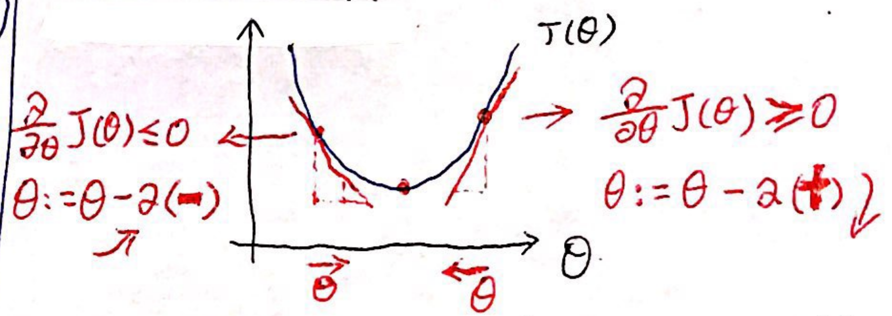
$\theta_1 := temp\_\theta_1$

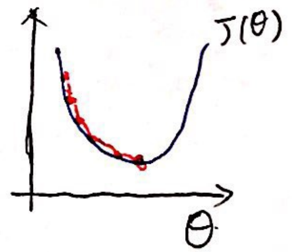
Gradient descent algorithm

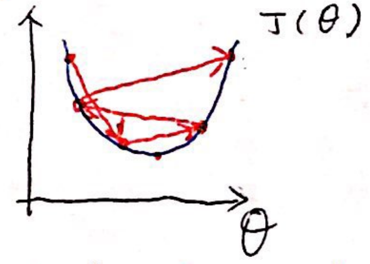
Repeat until convergence {

$$\theta_j := \theta_j - \underbrace{\alpha}_{\text{learning rate}}\underbrace{\frac{\partial}{\partial\theta_j}J(\theta_0,\theta_1)}_{\text{slope / gradient}} \quad (j=0\ \&\ j=1)$$

}







small α, slow converge

big α, overshoot

① how to select α : $0.001 \xrightarrow{10\times} 0.01 \xrightarrow{10\times} 0.1 \xrightarrow{10\times} 1$

② $J(\theta)$ can converge to a local minimum even with a fixed α. (α is small enough)

But fixed α is not always good:

— local minimum.

— slow converge.

①/②

Sep 18, 2018

Scanned with CamScanner

Multivariate Linear Regression

Hypothesis H : $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$

For convenience, define $x_0 = 1$, $x_0^{(i)} = 1$

m training data point $(\cancel{X}^{(i)}, y^{(i)})$

n features : $x_1, x_2, \cdots x_n$. (n+1)

n+1 parameters : $\theta_0, \theta_1, \theta_2 \cdots, \theta_n$

$$X = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n+1} \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n+1}$$

$$\Downarrow$$

$$\underbrace{h_\theta(x)}_{1\times 1} = \underbrace{\theta^T}_{1\times(n+1)} \underbrace{X}_{(n+1)\times 1}$$ inner product.

Polynomial Regression.

why ? $x - y$ cannot use a line to fit.

Hypothesis H : $h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_n x^n$

— map x to the space spanned by $\{x, x^2, \cdots x^n\}$

— it is a linear regression or not ?

— can fit any data if n is big enough

↳ but overfitting issue.

Repeat until converge {

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} [h_\theta(x^{(i)}) - y^{(i)}] x_j^{(i)} \}$$

$\boxed{\text{Gradient descent algorithm for multiple variables.}}$ $(n \geq 1)$

Simultaneously update $\theta_0 \cdots \theta_n$

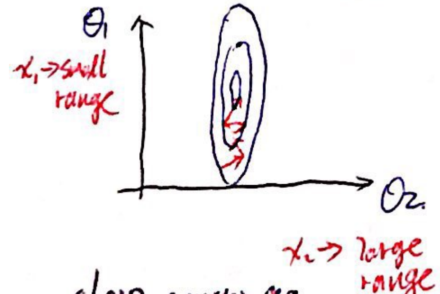
$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} [h_\theta(x^{(i)}) - y^{(i)}]$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} [h_\theta(x^{(i)}) - y^{(i)}] x_1^{(i)}$$

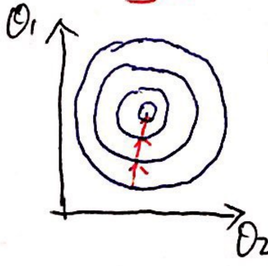
$$\theta_2 := - - - - - - - \cdot x_2^{(i)}$$

$$\theta_3 := - - - - - - - \cdot x_3^{(i)}$$

Feature Scaling , scale every feature x_i in the range $[0, 1]$ or $[-1, 1]$



slow converge



fast converge

— mean normalization

$$x_i := \frac{x_i - Mean(x_i)}{Max(x_i) - Min(x_i)}$$

— Gaussian normalization

$$x_i := \frac{x_i - Mean(x_i)}{Std(x_i)}$$

2/2

Spe 18, 2018

Scanned with CamScanner