~~Linear Regression~~

* Gradient descent Algorithm *

Let $f(x,y) = x^2 + y^2$ be the ~~Cost func~~ cost function and Convex in nature. so it has only one minimum (that is global minimum).

<u>task</u>: to minimize $f(x,y)$. i.e find (x,y) such that $f(x,y)$ is minimum.

<u>Solution</u>: Let initial solution be $\theta = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$

find ∇f : $\nabla f = \begin{bmatrix} 2x \\ 2y \end{bmatrix}_{2\times 1}$ #of ind. variables = 2

find Hf : to conform convexity $Hf = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}_{2\times 2}$

update θ: $\boxed{\theta_i := \theta_i - \alpha \nabla f}$ $\alpha \rightarrow$ learning rate

Let $\alpha = 0.01$

First iteration: $\theta = \begin{bmatrix} 10 \\ 10 \end{bmatrix} - 0.01 \begin{bmatrix} 2\times 10 \\ 2\times 10 \end{bmatrix}$

$= \begin{bmatrix} 10 \\ 10 \end{bmatrix} - 0.01 \begin{bmatrix} 20 \\ 20 \end{bmatrix}$

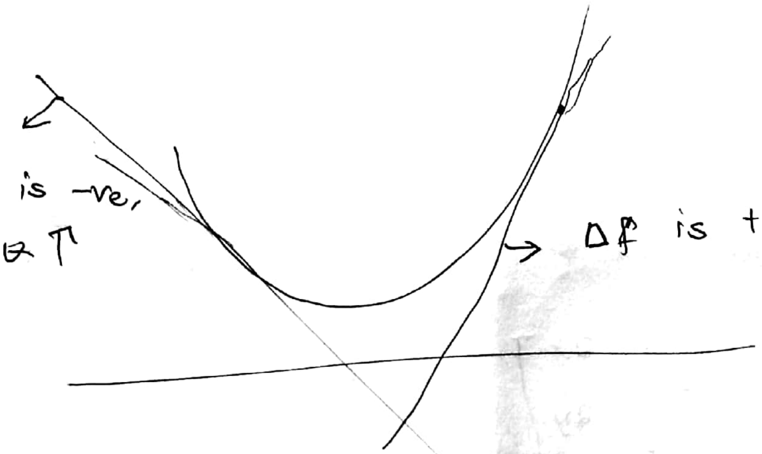
$\theta = \begin{bmatrix} 9.8 \\ 9.8 \end{bmatrix}$ after first iteration θ updated to $\theta = \begin{bmatrix} 9.8 \\ 9.8 \end{bmatrix}$

why α in $\theta := \theta - \alpha \nabla f$?

Δf is Valid and defined in very small interval.

α used to ensure ∇f is used in small interval.

why '$-$ve' sign in $\theta := \theta - \alpha \nabla f$?



∇f is +ve $\rightarrow$ $\theta \downarrow$

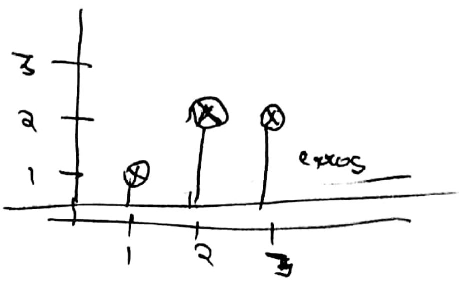
∇f is $-$ve $\rightarrow$ $\theta \uparrow$

10.05.2018

* Linear regression: Gradient descent algorithm *

Data:

x	y
1	1
2	2
3	2



Model/hypothesis: $h_\theta(x) = \theta_0 x_0 + \theta_1 x_1$ where $x_0 = 1$

cost function, $J(\theta) = \frac{1}{2m} \sum_{j=1}^{m} (h_\theta(x^{(j)}) - y^{(j)})^2$

Sum of squared errors

In general m = #. of training examples

$J(\theta)$ is exponential family hence it is convex function and has only one minimum (global minimum)

task: Find minimum of $J(\theta)$, ie find θ's such that $J(\theta)$ is minimum.

$$\boxed{\theta := \theta - \alpha \nabla f}$$ Parameter update rule.

$$h_\theta(x) = \theta^T x$$

$$\nabla f = \frac{1}{2m} \begin{bmatrix} \sum_{j=1}^{m} 2 \times (h_\theta(x^{(j)}) - y^{(j)}) \times \frac{\partial}{\partial \theta_0} h_\theta(x^{(j)}) \\ \vdots \\ \sum_{j=1}^{m} 2 \times (h_\theta(x^{(j)}) - y^{(j)}) \times \frac{\partial}{\partial \theta_n} h_\theta(x^{(j)}) \end{bmatrix}_{(n+1)\times 1}$$

$$= \frac{1}{m} \begin{bmatrix} \sum_{j=1}^{m} (h_\theta(x^{(j)}) - y^{(j)}) x_0^{(j)} \\ \vdots \\ \sum_{j=1}^{m} (h_\theta(x^{(j)}) - y^{(j)}) x_n^{(j)} \end{bmatrix}_{(n+1)\times 1}$$

Initialize θs,

$$\theta_0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

update θs

$$\theta := \theta - \alpha \nabla f$$

α = learning rate

Python tutorial for gradient descent Algorithm :-

$$X = \begin{bmatrix} 1 & x_0^{(1)} & x_1^{(1)} \cdots & x_n^{(1)} \\ 1 & \vdots & & \\ 1 & & & \\ 1 & x_0^{(m)} & x_1^{(m)} \cdots & x_n^{(m)} \end{bmatrix}_{m \times (n+1)} \qquad y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}_{(m \times 1)}$$

$$\nabla f = \frac{1}{m} \begin{bmatrix} \sum_{j=1}^{m} (h_\theta(x^{(j)}) - y^{(j)})\, x_0^{(j)} \\ \vdots \\ \sum_{j=1}^{m} (h_\theta(x^{(j)}) - y^{(j)})\, x_n^{(j)} \end{bmatrix}_{(n+1) \times 1}$$

~~Computed as~~

$$\theta = \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_n \end{bmatrix}_{(n+1) \times 1}$$

$$\nabla f = \left[(X \cdot \theta - y)^T \cdot X \right]^T = X^T (X \cdot \theta - y)$$

$X \cdot \theta \rightarrow [m \times (n+1)] \times [(n+1) \times 1] \Rightarrow m \times 1$

$X \cdot \theta - y \rightarrow (m \times 1)$

$(X \cdot \theta - y)^T \rightarrow 1 \times m$

$(X \cdot \theta - y)^T X \rightarrow (1 \times m)\ (m \times (n+1)) \Rightarrow 1 \times (n+1)$

Scanned by CamScanner

$$\left[(X\cdot\theta - y)^T \cdot X\right]^T \rightarrow \quad (n+1) \times 1$$

$$\underset{(n+1)\times 1}{\theta} := \underset{(n+1)\times 1}{\theta} - \frac{\alpha}{m}\underbrace{\left[(X\cdot\theta - y)^T \cdot X\right]^T}_{(n+1)\times 1}$$

$$\boxed{\theta := \theta - \frac{\alpha}{m}\left[(X\cdot\theta - y)^T \cdot X\right]^T} \quad (AB)^T = B^T A^T$$

$$\boxed{\theta := \theta - \frac{\alpha}{m} X^T (X\cdot\theta - y)}$$

X^T: $(n+1) \times m$ $\quad$ $(X\cdot\theta - y)$: $m \times 1$

$$((n+1)\times m) \times (m \times 1) = (n+1) \times 1$$

Scanned by CamScanner