$\theta_0 : 1 + \sqrt{2}$

$\theta_1 : 2 + \sqrt{2}$

Cost Function → training example

$m = 47$

$h_\theta(x) = \theta_0 + \theta_1 x$ θ_0, θ_1 → Parameters

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

$$h_\theta(x^i) = \theta_0 + \theta_1 x^{(i)}$$

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

Minimize $J(\theta_0, \theta_1)$

θ_0, θ_1 cost func

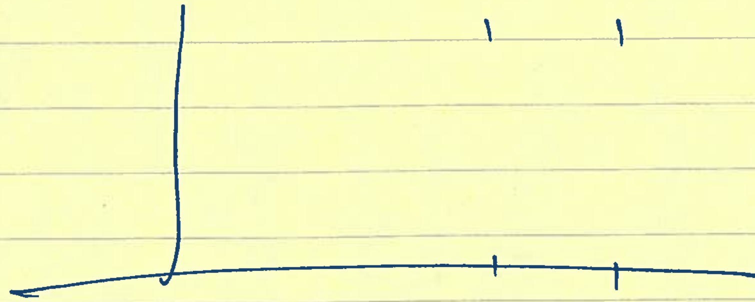
Squared error function

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

$$\frac{1}{2(3)} \sum \left[1^2 + 2^2 + 3^2\right] = \frac{10}{6} = 5/3$$

1 4 9

15/6



$$\frac{1}{\cancel{10}}\left[(5-4)^2+(3-4)^2+(0-1)^2+(4-3)^2\right]$$

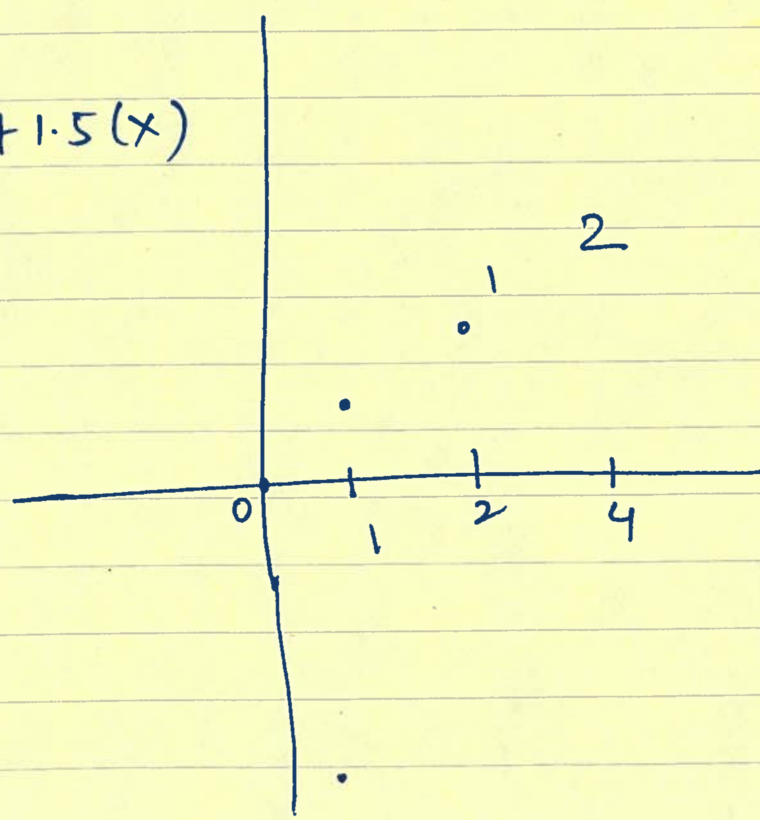
$$\frac{1}{8}[1+1+1+1] = 4/8 = 0.5.$$

$h_0(x) = 1x$

$h_0(x) = -2 + 0.5(x)$ $h_\theta(x) = 0 + 1.5(x)$

$h_0(6) = -2 + 3 = 1$ $h(2) = 1.5(2)$

$h(2) = 0 + 1.5(x)$



$x^{(2)} \xrightarrow{} 2nd$ Row

$$x^{(2)} = \begin{bmatrix} 1416 \\ 3 \\ 2 \\ 40 \end{bmatrix} \in \mathbb{R}^4$$

$x_j^{(i)}$ = value feature j in i^{th} training ex

$$x_3^2 = 2$$

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

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \cdots \theta_n x$$

E.g. $h_\theta(x) = 80 + 0.1 x_1 + 0.00 x_2 + 3 x_3 - 2 x_4$.

$x_0 = 1$

$(x_0^{(i)} = 1)$

θ^T is 1 by $(n+1)$ matrix not $(n+1)$

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n+1} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n+1}$$

$$\theta^T = [\theta_0 \; \theta_1 \; \theta_2 + \cdots \theta_n$$

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$
$$= \theta^T x$$

Gradient Descent for Multiple Variables.

$$x_1 = 1$$

Hyp: $h_\theta(x) = \theta^T x = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$

Parameters $\underline{\theta}$ $n+1$ dimensional vector

$$\underbrace{J(\theta_0 + \theta_1 + \cdots \theta_n)}_{J(\underline{\theta})} = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^{(i)})^2$$

Repeat

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \underbrace{\boxed{\sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})}}_{\frac{\partial}{\partial \theta_0} J(\theta)}$$

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

$n \geq 1$

$$\theta_j = \theta_j - \alpha \frac{1}{m} \underbrace{\boxed{\sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}}}_{\frac{\partial}{\partial \theta_j} J(\theta)}$$

$$J_0 = \frac{1}{2m} \sum_{i=1}^{m} \sum h_\theta(x^i) - y^i$$

$x_1 \Leftarrow \frac{x_1 - \mu_1}{s_1}$ ⇐ avg value of x_1 in traingset

$s_1 \rightarrow$ Range $\boxed{\text{Max} - \text{Min}}$ or std dev.

$y = 28 \times 1$

4×4 x

$-0.5 \le x_1 \le 0.5 \quad -0.5 \le x_2 \le 0.5$

$14 \times 4 \quad [14 \times 1]$

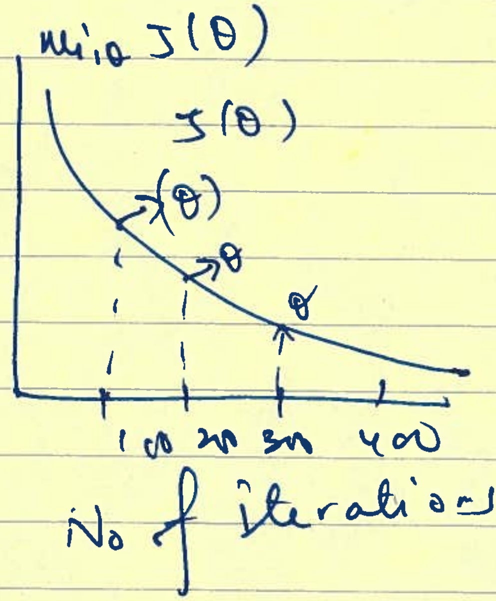
Learning Rate

$\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j}$

6675.5 -

$\frac{4761 - 6675.5}{3160}$

How to choose learning rate α.



$h_\theta(x) = \theta_1$

$$h_\theta(x) = \theta_0 + \theta_1 \times \underbrace{\text{frontage}}_{x_1} + \theta_2 \times \underbrace{\text{depth}}_{x_2}$$

Area =

x = frontage × depth

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

land area.

$x_1 = \frac{\text{size}}{1000}$, $x_2 = \frac{\sqrt{\text{size}}}{32}$ Range 1000

$\sqrt{1000} \approx 32$

0.47

0.37

$$x_1 = \frac{\text{size}}{1000}, \quad x_2 = \frac{\sqrt{\text{size}}}{32}$$

$x_2 = \frac{4761}{\quad}$

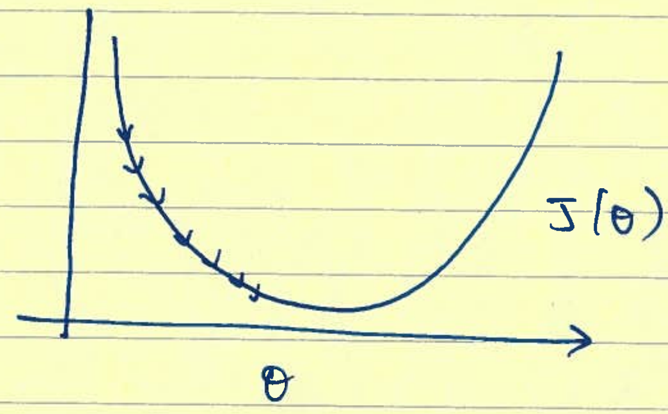
Feature Scaling / Mean Normalization with Polynomials.

$\frac{4761 - 6679}{(25)^2}$

$\frac{\text{size}}{1000}, \frac{(\text{size})^2}{(1000)^2}$

$\frac{\text{midterm}}{\text{Range}}$

Normal Equation



$$1D \; (\theta \in \mathbb{R})$$

$$J(\theta) = a\theta^2 + d\theta + c$$

$m = 4.$

$m = \cancel{4}\ 28$

$n = 5$

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

$X = 28 \times 6$

$y = 28 \times 1 \qquad \theta = 6 \times 1$

$(x^{(1)}, y^{(1)}), \ldots (x^{(m)}, y^{(m)})$: n features

$$x^{(i)} = \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix} \in \mathbb{R}^{n+1}$$

X (design matrix) $= \begin{bmatrix} x^{(1)T} \\ x^{(2)T} \\ \vdots \\ x^{(m)T} \end{bmatrix}$

$m \times (n+1)$

if $x^{(i)} = \begin{bmatrix} 1 \\ x_1^{(i)} \end{bmatrix}$

$$(X^T X)^{-1}$$

set $A = X^T X$

$$(\cancel{X}^T X)^{-1} = A^{-1}$$

Gradient Descent

→ Need to Choose α

· Needs many iteration

n & m.

↓

works very well n is large.

Normal Equation

→ No need to choose α

→ Don't need to iterate

Need to compute

$(X^T X)^{-1}$ $n \times n$ $O(n^3)$

Slow if n is very large.

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

$$= \frac{\partial}{\partial\theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2$$

$j=0:$ $\frac{\partial}{\partial\theta_0} J(\theta_0,\theta_1) = \frac{1}{m} \sum_{i=1}^{m} \cancel{h_0}\ \theta_0 + \theta_1 x^{(i)} - y^{(i)}$

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

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \frac{\partial}{\partial\theta_0} (J(\theta_0,\theta_1)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \frac{\partial}{\partial\theta_1} J(\theta_0,\theta_1)$$

```
i = 1:m

j1 = sum((theta(1) + theta(2).*x(i,2) - y(i));
j2 = sum(((theta(1)+ theta(2).*x(i,2) - y(i))
         .* X(i,2));

theta(1) = theta(1) - (alpha/m) * (j1);
theta(2) = theta(2) - (alpha/m) * (j2);

theta = theta - alpha * 1/m * (((X * theta) - y)' * X)'
```

Cost Function

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$$

$$Cost(h_\theta(x), y) = -\log(h_\theta(x)) \quad \text{if } y = 1$$

$$Cost(h_\theta(x), y) = -\log(1 - h_\theta(x) \quad \text{if } y = 0$$

$$h = g(X\theta)$$

$\downarrow$

sigmoid

$$J(\theta) = \frac{1}{m} \cdot (-y^T \log(h) - (1-y)^T \log(1-h))$$

$$J = 1/m * sum(-y .* log(h) - (1-y) .* log(1-h))$$

Grad Descent

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

$$\theta : \theta - \frac{\alpha}{m} X^T (g(X\theta) - \vec{y})$$

grad = ~~theta~~ - $\frac{alpha}{m}$ * sum((h - y) .* X);