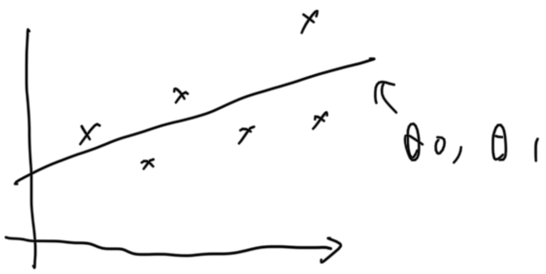
ML w1 Linear regression with one variable

How do we represent "h" ?

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

Cost Function

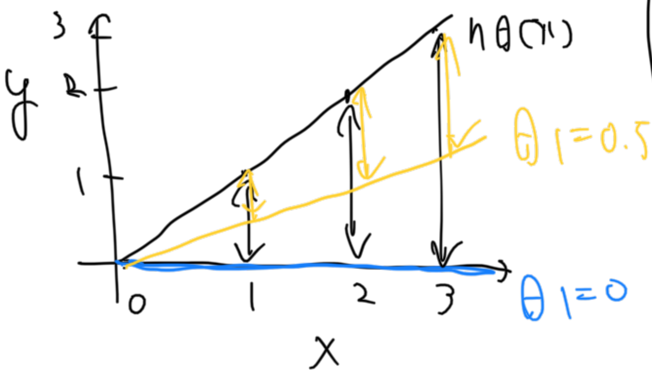


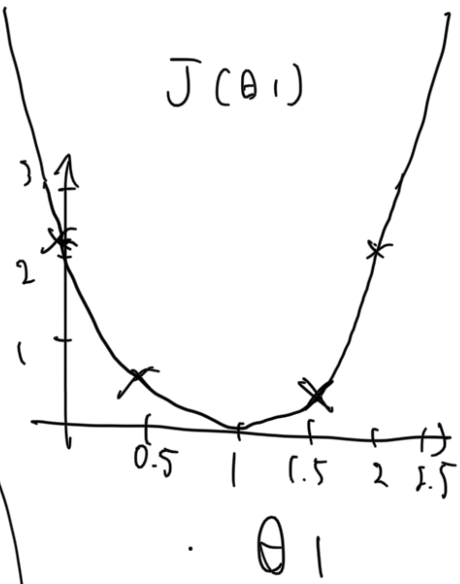
Idea: Choose θ_0, θ_1 so that $h_\theta(x)$ is close to y for training example.

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

minimize $\underbrace{J(\theta_0, \theta_1)}_{\text{cost function}}$

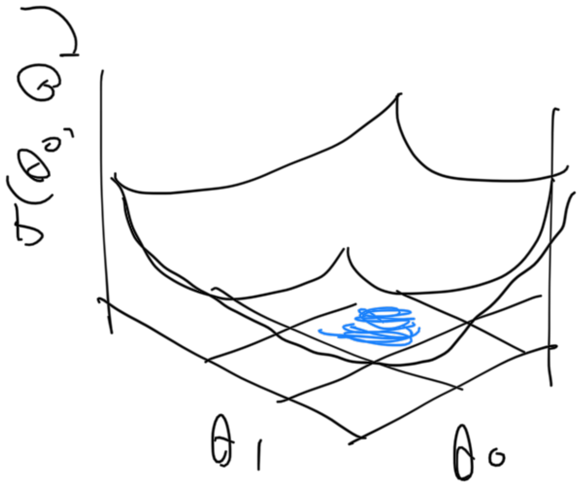
$h_\theta(x)$ (for fixed θ_0)

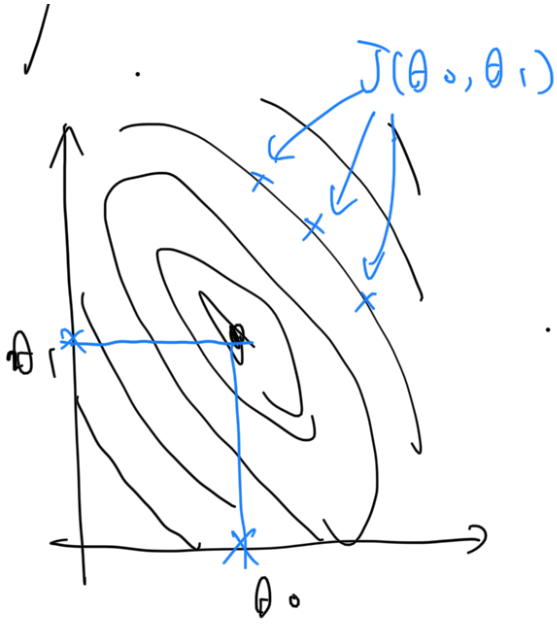




$J(\theta_1) = \frac{1}{2m}(1^2 + 2^2 + 3^2)$
$\theta_1 = 0$
$= \frac{1}{\ }(14) \fallingdotseq 2.3$

minimize $J(\theta_1)$





Gradient Decent

- Start with some θ_0, θ_1
- Keep changing θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$ untill we hopefully end up at a minimum

algorithm

repeat until convergence

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad (\text{for } j=0 \text{ and } j=1)$$

learning rate

simultaneosly update θ_0 and θ_1

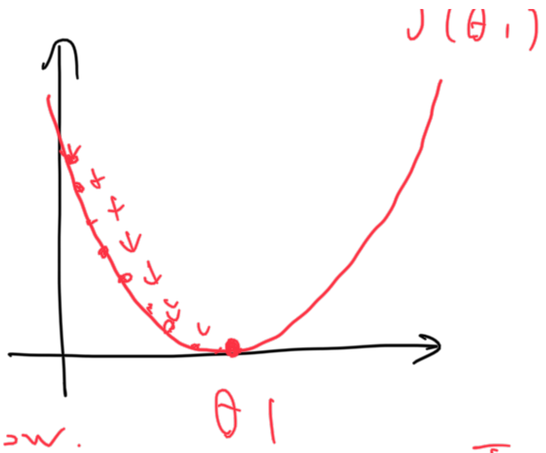
$$\text{temp}0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

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

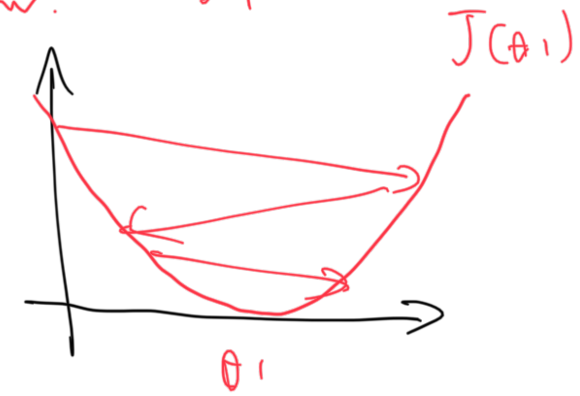
$$\theta_0 := \text{temp}0$$

$$\theta_1 := \text{temp}1$$

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



If α is too small,
gradient descent can be slow.



If α is too large,
gradient descent can
overshoot the minimum.
It may fail to converge,
or even diverge.

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

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

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

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

$\downarrow$

∂

Gradient descent algorithm

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

repeat until convergence {

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

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

}

$$\frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

Linear Algebla review

Matrix

$$A = \begin{bmatrix} 123 & 416 \\ 8 & 10 \\ 9 & 7 \\ 30 & 2 \end{bmatrix}$$

$$A_{11} = 123$$

$$A_{32} = 7$$

$\underbrace{4}_{\text{row}} \times \underbrace{2}_{\text{col}}$ matrix

Vector : An $n \times 1$ matrix

$$y = \begin{bmatrix} 410 \\ 320 \\ 30 \\ 2 \end{bmatrix} \quad n = 4$$

matrix adition

$$\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0.5 \\ 2 & 5 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0.5 \\ 4 & 10 \\ 5 & 2 \end{bmatrix}$$

$$\cancel{\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 5 \\ 0 & 1 \end{bmatrix}}$$

Scalar Multiplication

$\uparrow$
real number

$$3 \times \begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 6 & 15 \\ 9 & 3 \end{bmatrix}$$

Matrix - vector Multiplication

$$\begin{bmatrix} 1 & 3 \\ 4 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 16 \\ 4 \\ 7 \end{bmatrix}$$

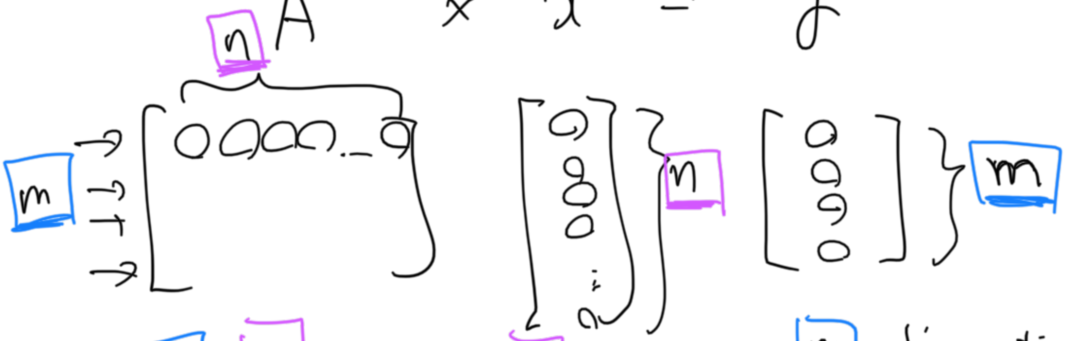
3×2 $\quad$ 2×1

$1 \times 1 + 3 \times 5 = 16$

$4 \times 1 + 0 \times 5 = 4$

$2 \times 1 + 1 \times 5 = 7$

$$A \times x = y$$



$m \times n$ matrix $\quad n \times 1$ vector $\quad m$-dimentional vector

$$\begin{bmatrix}1&2&1&5\\0&3&0&4\\-1&-2&0&0\end{bmatrix}\begin{bmatrix}1\\3\\2\\1\end{bmatrix}=\begin{bmatrix}14\\13\\-7\end{bmatrix}$$

$$3\times\boxed{4} \qquad \boxed{4}\times 1 \qquad 3\times 1$$

$$1\times1+2\times3+1\times2+5\times1=14$$

$$0\times1+3\times3+0\times2+4\times1=13$$

$$-1\times1+-2\times3+0\times2+0\times1=-7$$

House sizes $\qquad \underline{h_\theta(x)} = -40 + 0.25x$

2104
1416
1534
852

$$\begin{bmatrix}1&2104\\1&1416\\1&1534\\1&852\end{bmatrix}\times\begin{bmatrix}-40\\0.25\end{bmatrix}=\begin{bmatrix}-40\times1+2104\times0.25\\-40\times1+1416\times0.25\\-40\times1+1534\times0.25\\-40\times1+852\times0.25\end{bmatrix}$$

Matrix - matrix multiplication

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 11 & 10 \\ 9 & 14 \end{bmatrix}$$

3×2

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} = \boxed{\begin{bmatrix} 11 \\ 9 \end{bmatrix}}$$

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} = \boxed{\begin{bmatrix} 10 \\ 14 \end{bmatrix}}$$

$$A \times B = C$$

$$\underset{m \times n}{[\;\;]} \times \underset{n \times o}{[\;\;]} = \underset{m \times o}{[\;\;]}$$

eg.

House sizes   3 competing hypotheses

$$\begin{cases} 2104 \\ 1406 \\ 1534 \\ 852 \end{cases}$$

$\rightarrow h_\theta(x) = -40 + 0.25x$

$\rightarrow h_\theta(x) = 200 + 0.1x$

$\rightarrow h_\theta(x) = -150 + 0.4x$

Matrix

$$\begin{bmatrix} 1 & 2104 \\ 1 & 1406 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix} \times \begin{bmatrix} -40 \\ 0.25 \end{bmatrix} \begin{bmatrix} 200 \\ 0.1 \end{bmatrix} \begin{bmatrix} -150 \\ 0.4 \end{bmatrix}$$

$$= \begin{bmatrix} 486 \\ 314 \\ 344 \\ 173 \end{bmatrix} \begin{bmatrix} 410 \\ 342 \\ 353 \\ 285 \end{bmatrix} \begin{bmatrix} 692 \\ 416 \\ 464 \\ 191 \end{bmatrix}$$

Inverse and Transpose

$$3 \cdot (\underbrace{3^{-1}}_{\text{inverse}}) = 1 \qquad 3^{-1} = \frac{1}{3}$$

$$12 (12^{-1}) = 1 \qquad 12^{-1} = \frac{1}{12}$$

- Not all number have an inverse.

- **Matrix inverse**

If A is an $m \times m$ (square) matrix, and it has an inverse,

$$A(A^{-1}) = (A^{-1})A = I$$

e.g.

$$\underbrace{\begin{bmatrix} 3 & 4 \\ 2 & 16 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 0.4 & -0.1 \\ -0.05 & 0.075 \end{bmatrix}}_{A^{-1}} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{A^{-1}A}$$

$$\begin{matrix} \underset{1.2}{3 \times 0.4} + \underset{-0.2}{4 \times (-0.05)} \\ \underset{0.8}{2 \times 0.4} + \underset{-0.8}{16 \times (-0.05)} \end{matrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\begin{matrix} \underset{-0.3}{3 \times (-0.1)} + 4 \times 0.075 \\ 2 \times (-0.1) + 16 \times 0.075 \end{matrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Matrix Transpose

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 5 & 9 \end{bmatrix} \qquad B = A^T = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 0 & 9 \end{bmatrix}$$

$$m \times n \qquad\qquad n \times m$$