L8: Linear Regression

Prof. Xun Jiao

Before class

- HW1 feedback: check with TA dma2@Villanova.edu (Dongning Ma)
- HW 2 due by Wed

Review

- Machine learning
  - supervised learning
    - regression
      - Linear Regression
      - ……
    - classification
  - unsupervised learning

What is a linear equation?

Linear regression assumes a predictor of the form

$$X\theta = y$$

matrix of features
(data)

unknowns
(which features are relevant)

vector of outputs
(labels)

(or $Ax = b$ if you prefer)

Linear
Regression

Linear regression assumes a predictor of the form

$$X\theta = y$$

Q: Solve for theta

How to find Θ?

- In other words, how to find a line that can fit the data?

Cost Function

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

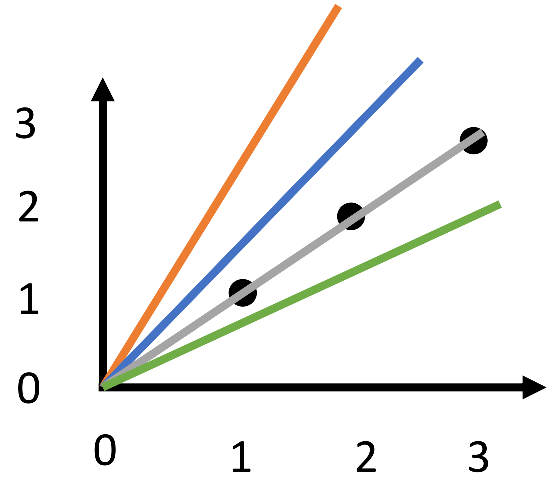
Parameters: θ_0, θ_1

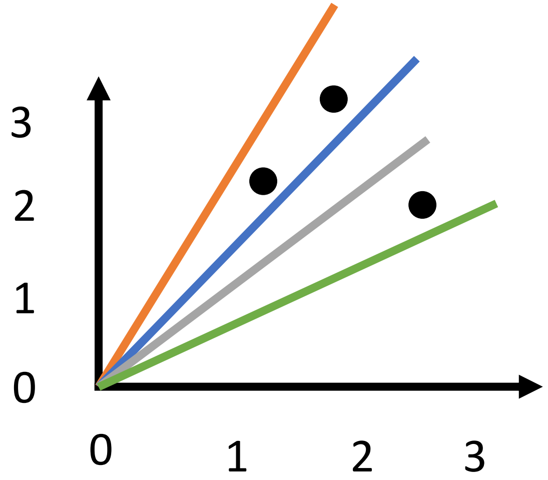
Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

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

Squared error function: most-widely used cost function.

Let us use an example with one parameter



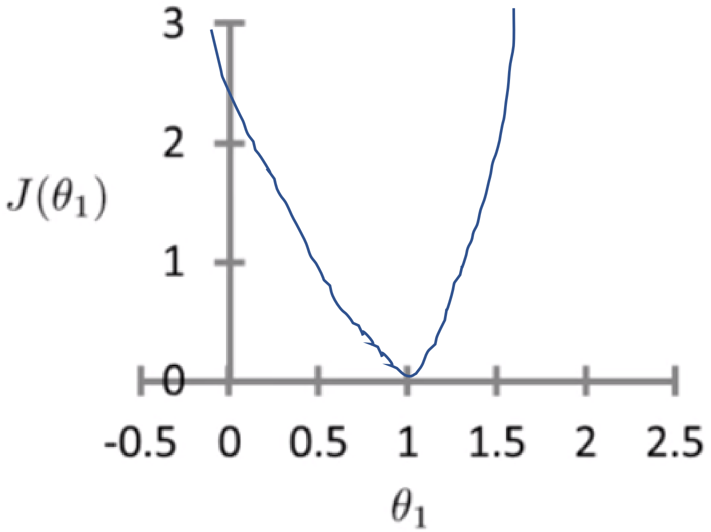


Hypothesis: $h_\theta(x) = \cancel{\theta_0} \cancel{+} \theta_1 x$

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

$J(\theta_1)$

(function of the parameter θ_1)

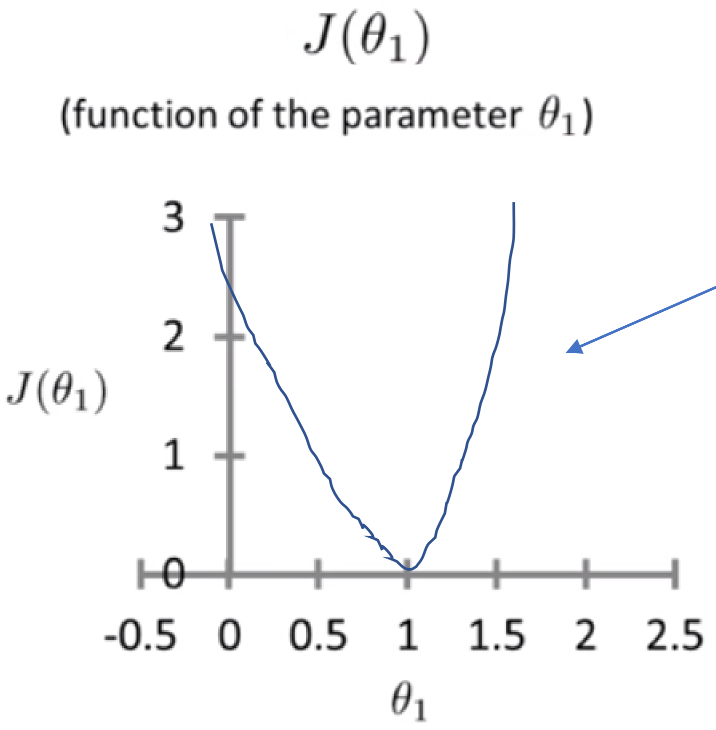


What is the J() value under different theta?

- Find theta that can minimize J()

What is our final goal?

- Our final goal is to find a parameter theta that can minimize the cost function J.
  - Did we find it?
  - How can we validate it?



Bowl-shape function

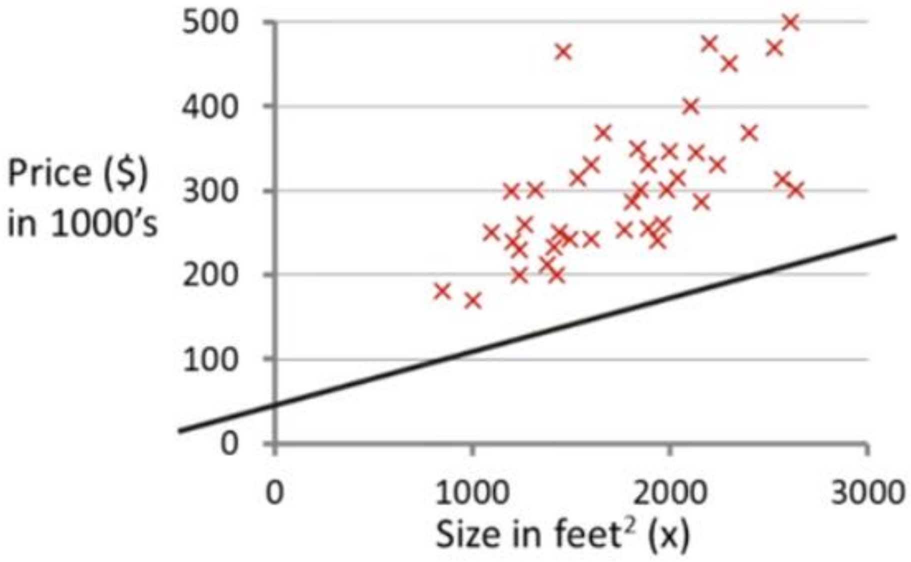
Now, what if we have two parameters?

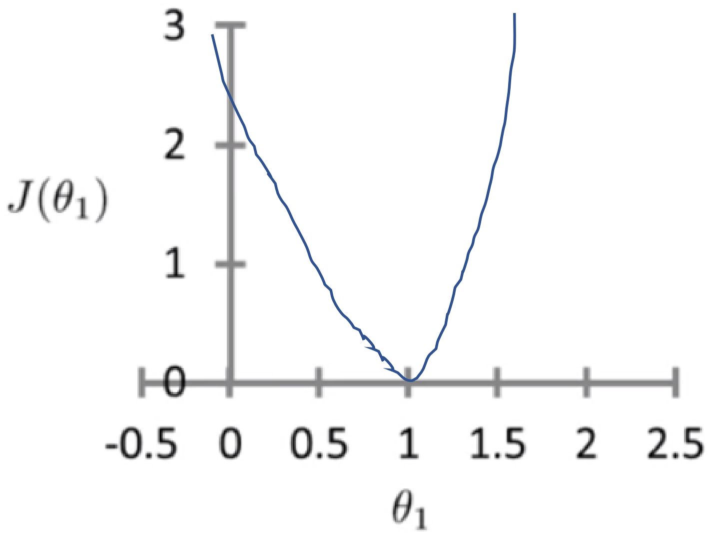
What would our J() look like???

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

$J(\theta_1)$

(function of the parameter θ_1)

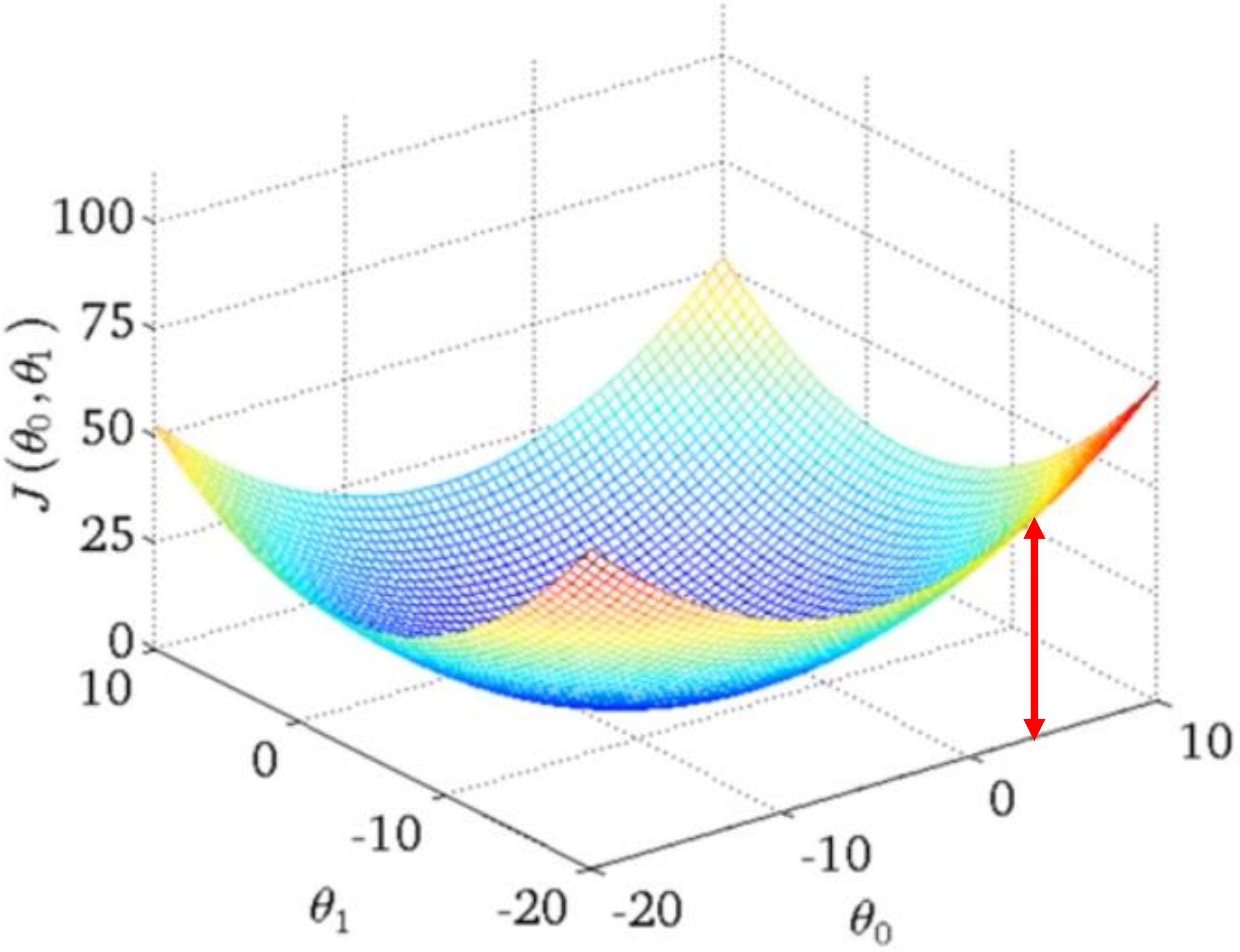




still this?

θ_0 =50 $\quad h_\theta(x) = 50 + 0.06x$

θ_1 =0.06



100
75
50
25
0
$J(\theta_0, \theta_1)$
10
0
-10
-20
θ_1
-20
-10
0
10
θ_0

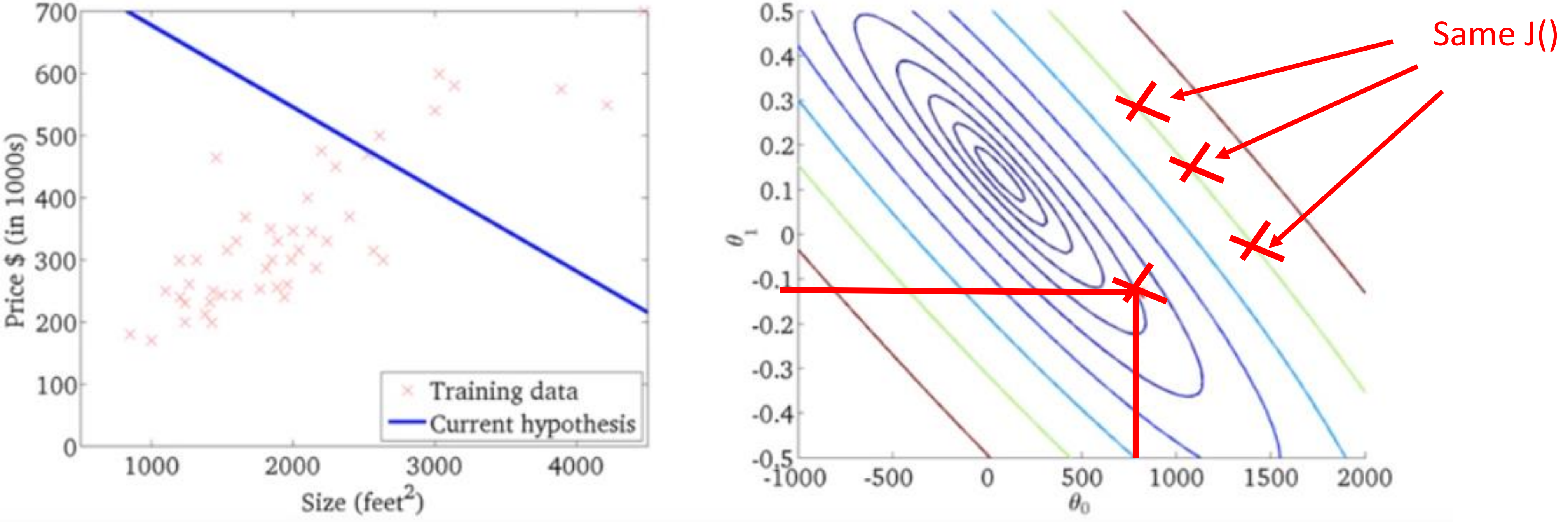
Contour plots/figures

$h_\theta(x)$

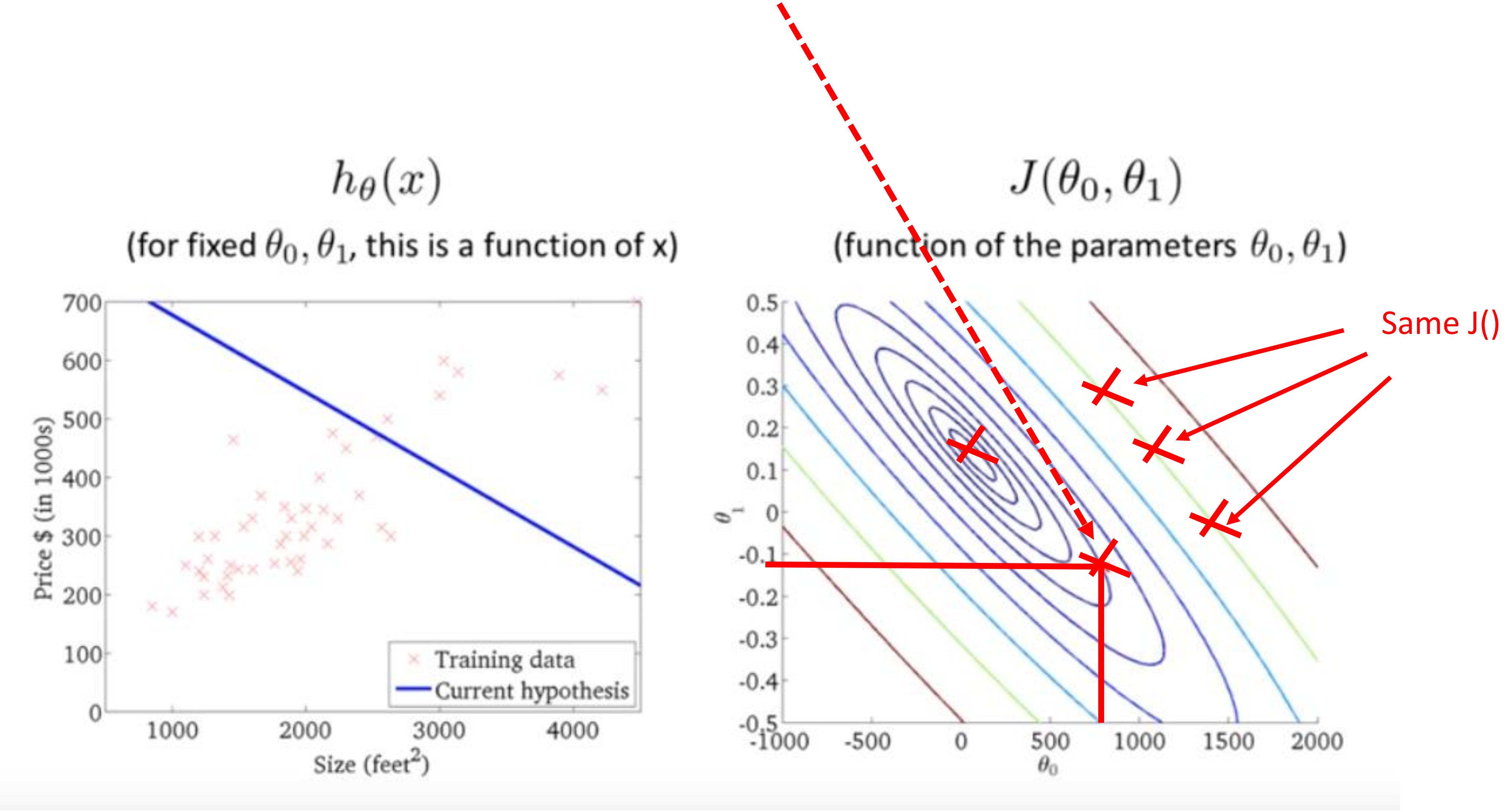
(for fixed θ_0, θ_1, this is a function of x)

$J(\theta_0, \theta_1)$

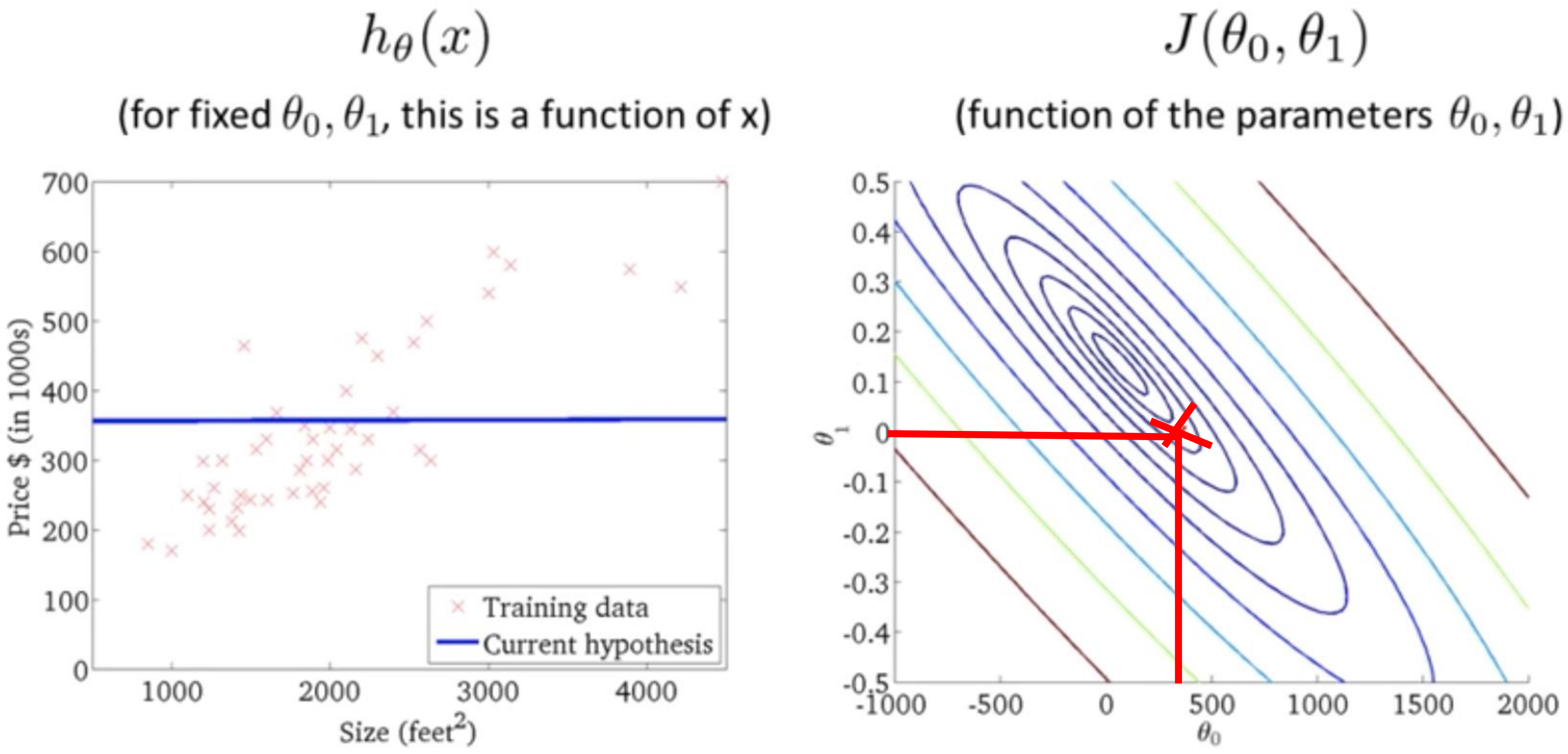
(function of the parameters θ_0, θ_1)



The cost is high under this parameter



Another example with different parameter

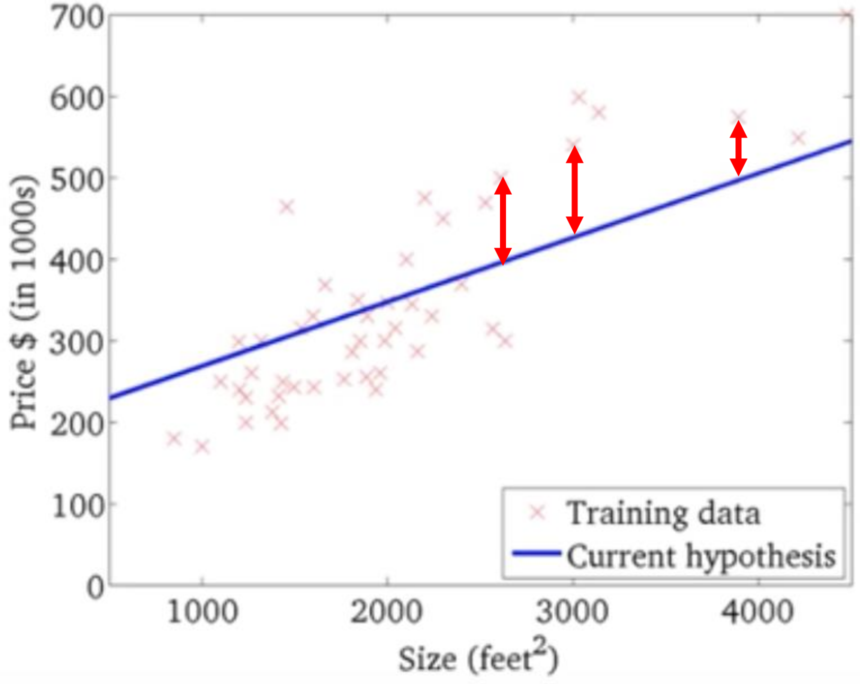


Is this model better than previous one?

Ultimate example

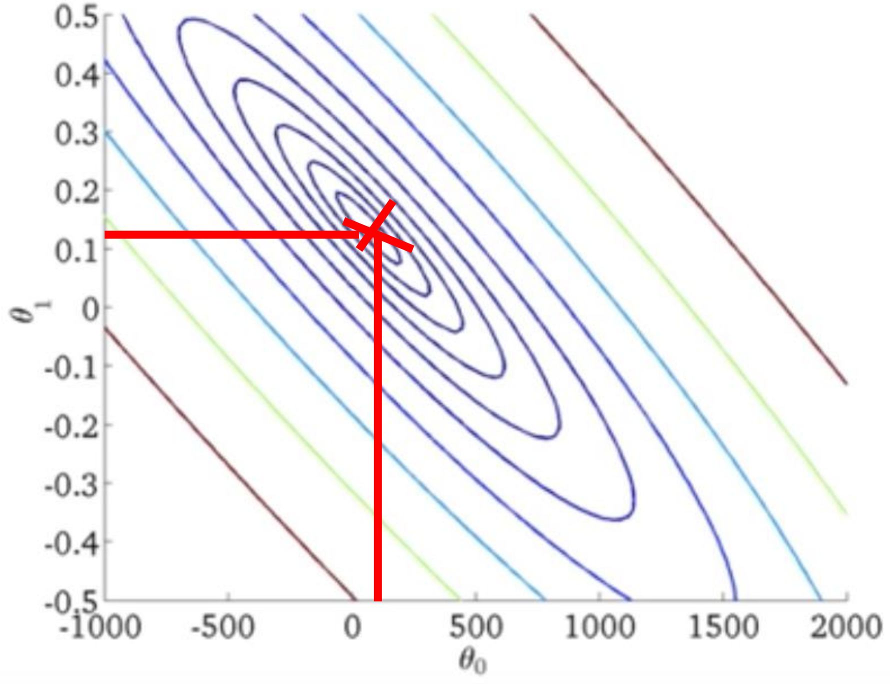
$h_\theta(x)$

(for fixed θ_0, θ_1, this is a function of x)



$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



Is this model better than previous one?

So, the problem really becomes, how can we automatically find parameters that minimize cost function J?

Gradient Descent!

Have some function $J(\theta_0, \theta_1)$

Can be even more theta

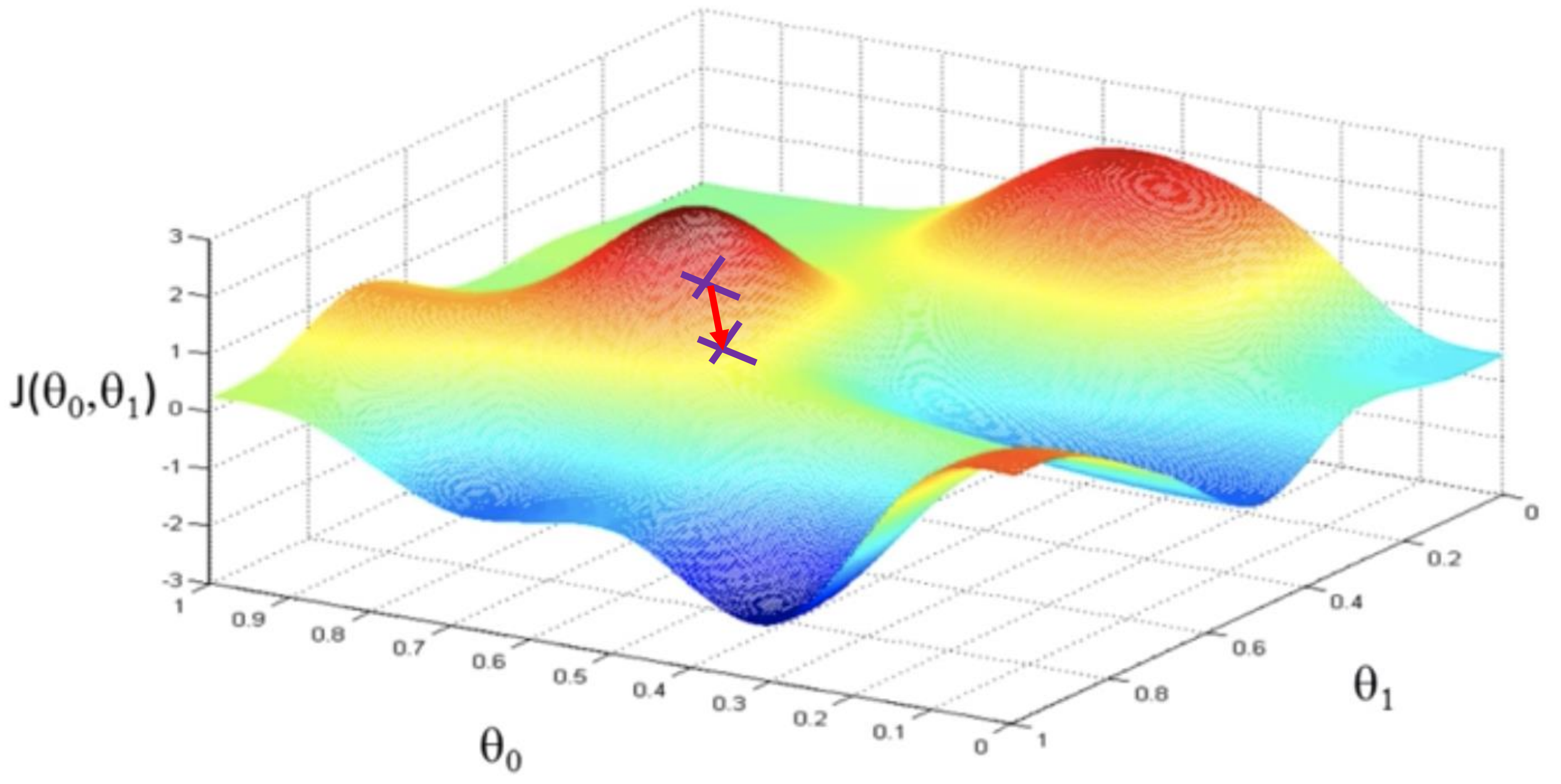
Want $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

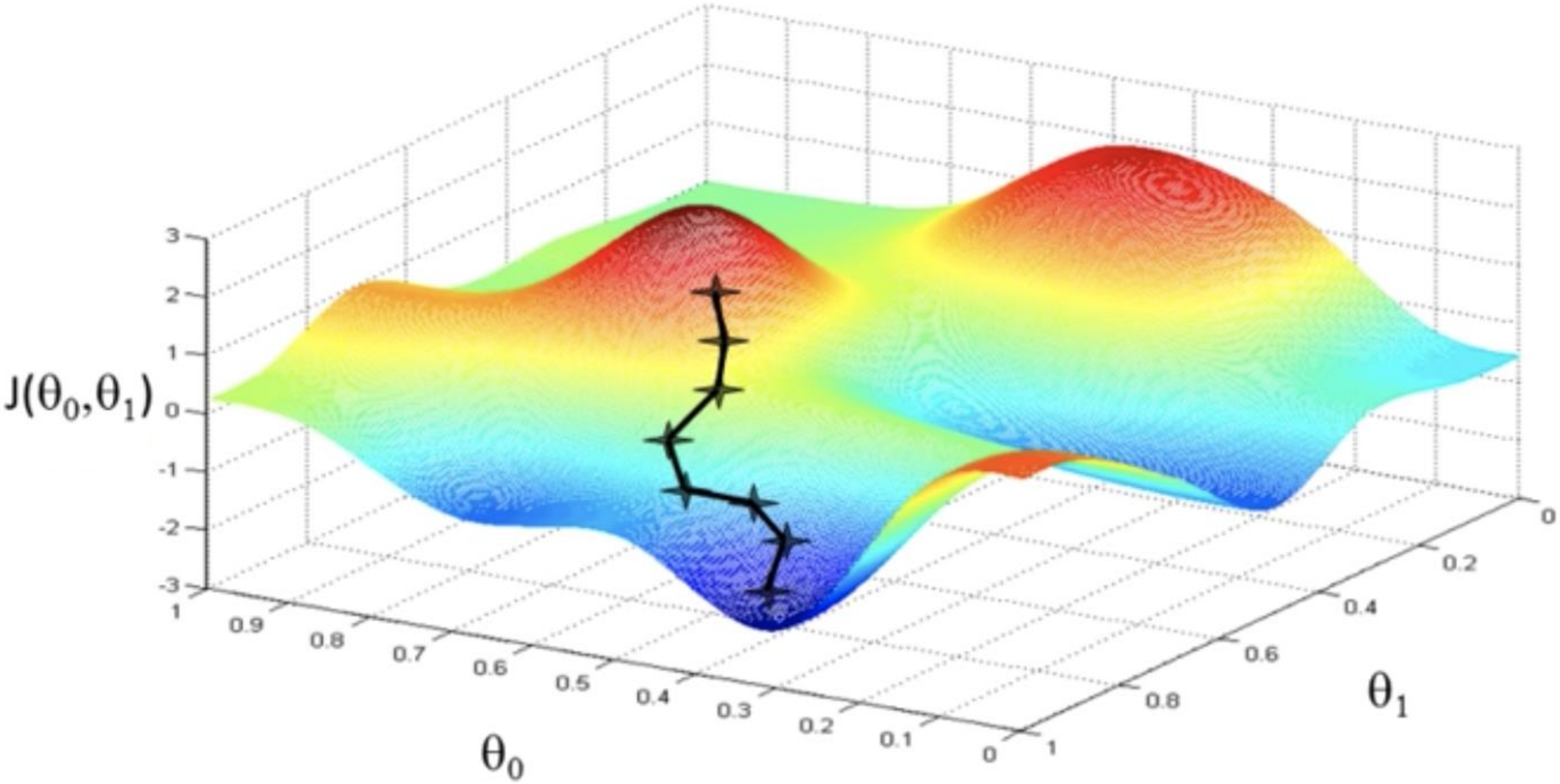
Outline:

- Start with some θ_0, θ_1 e.g., both are zeros
- Keep changing θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$

  until we hopefully end up at a minimum

Which direction I should move if I want to go down (as rapidly as possible) ?





$J(\theta_0,\theta_1)$
3
2
1
0
-1
-2
-3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
θ_0
0
0.2
0.4
0.6
0.8
1
θ_1

Gradient descent algorithm

repeat until convergence {

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

}

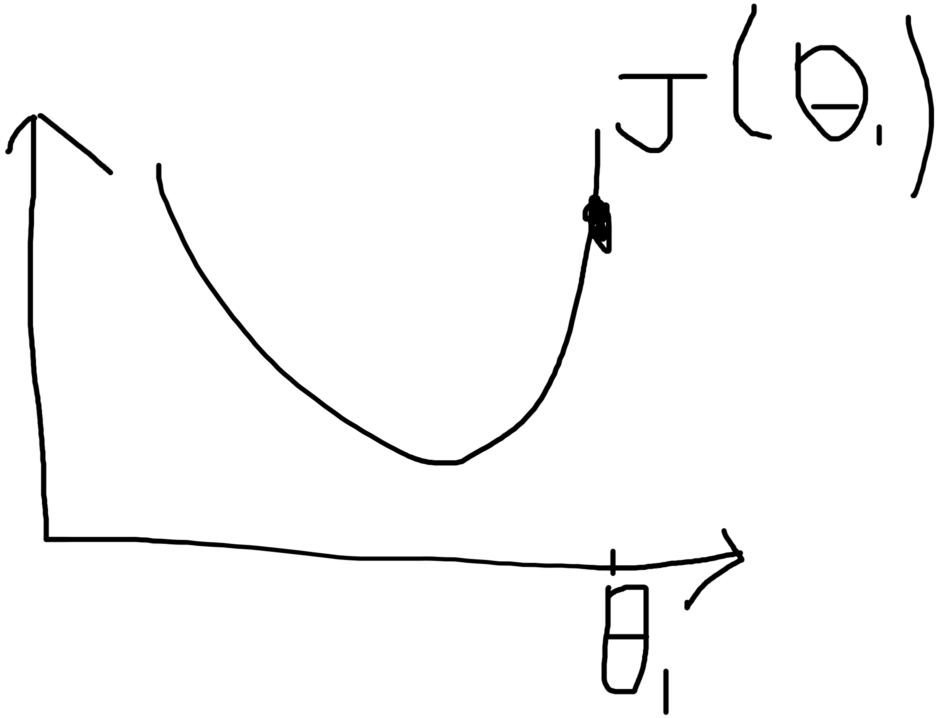
learning rate

Correct: Simultaneous update

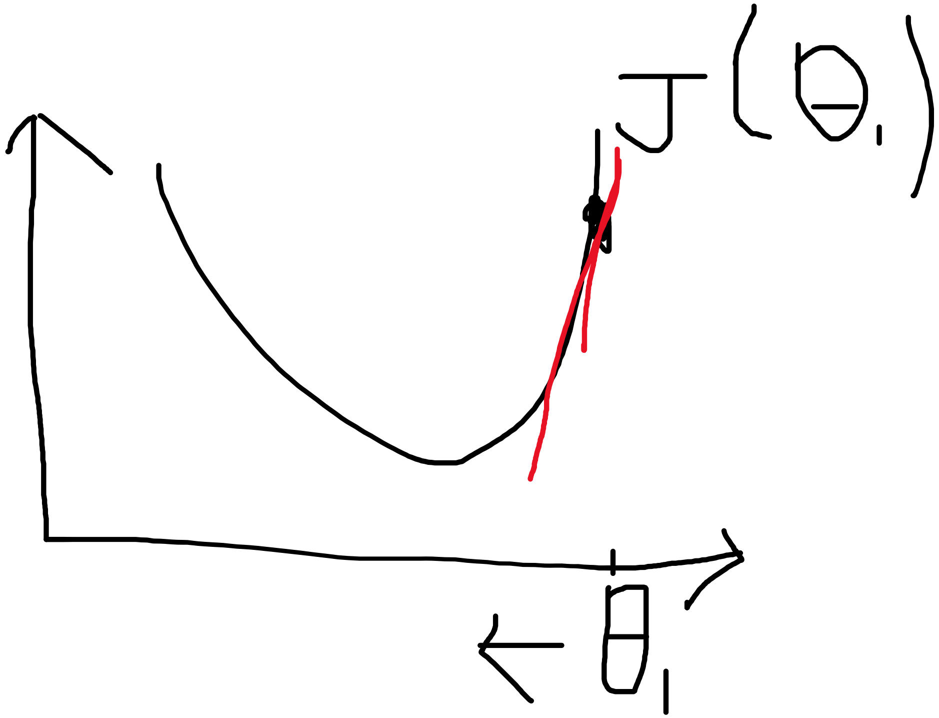
$$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$
$$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$
$$\theta_0 := \text{temp0}$$
$$\theta_1 := \text{temp1}$$

Incorrect:

$$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$
$$\theta_0 := \text{temp0}$$
$$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$
$$\theta_1 := \text{temp1}$$

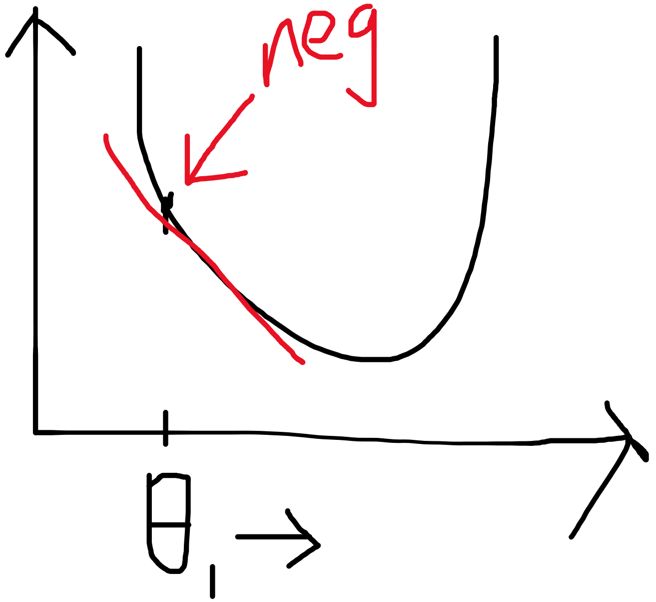


$$\theta_1 := \theta_1 - \alpha \frac{d}{d\theta_1} J(\theta_1)$$



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

$$\theta_1 = \theta_1 - \alpha (\text{pos})$$



$$\theta_1 = \theta_1 - \alpha(\text{neg})$$