MACHINE LEARNING (ML-5)

Dr. NEERAJ GUPTA, Department of CEA, GLA University, Mathura

AGENDA

- Linear regression with one variable

QUIZ

X	Y
2	4
3	9
5	25
9	81
7	49
11	121
10.5	WHAT?

QUIZ

X	Y
2	4
3	9
5	25
9	81
7	49
11	121
10.5	110.25

HOW DO YOU FIND THAT?

You find the relation between X and Y

HOW DO YOU FIND THAT?

You find the relation between X and Y

$$Y = X.X = X^2$$

$$Y = f(X)$$

HOW DO YOU FIND THAT?

You find the relation between X and Y

$$Y = X.X = X^2$$

$$Y=f(X)$$

Which one is dependent variable ?

HOW DO YOU FIND THAT?

You find the relation between X and Y

$$Y = X.X = X^2$$

$$Y = f(X)$$

Which one is dependent variable ? ANSWER = Y

HOW DO YOU FIND THAT?

You find the relation between X and Y

$$Y = X.X = X^2$$

$$Y = f(X)$$

Which one is dependent variable ? ANSWER = Y

So What is X?

HOW DO YOU FIND THAT?

You find the relation between X and Y

$$Y = X.X = X^2$$

$$Y = f(X)$$

Which one is dependent variable ? ANSWER = Y

SO What is X? X= Independent variable

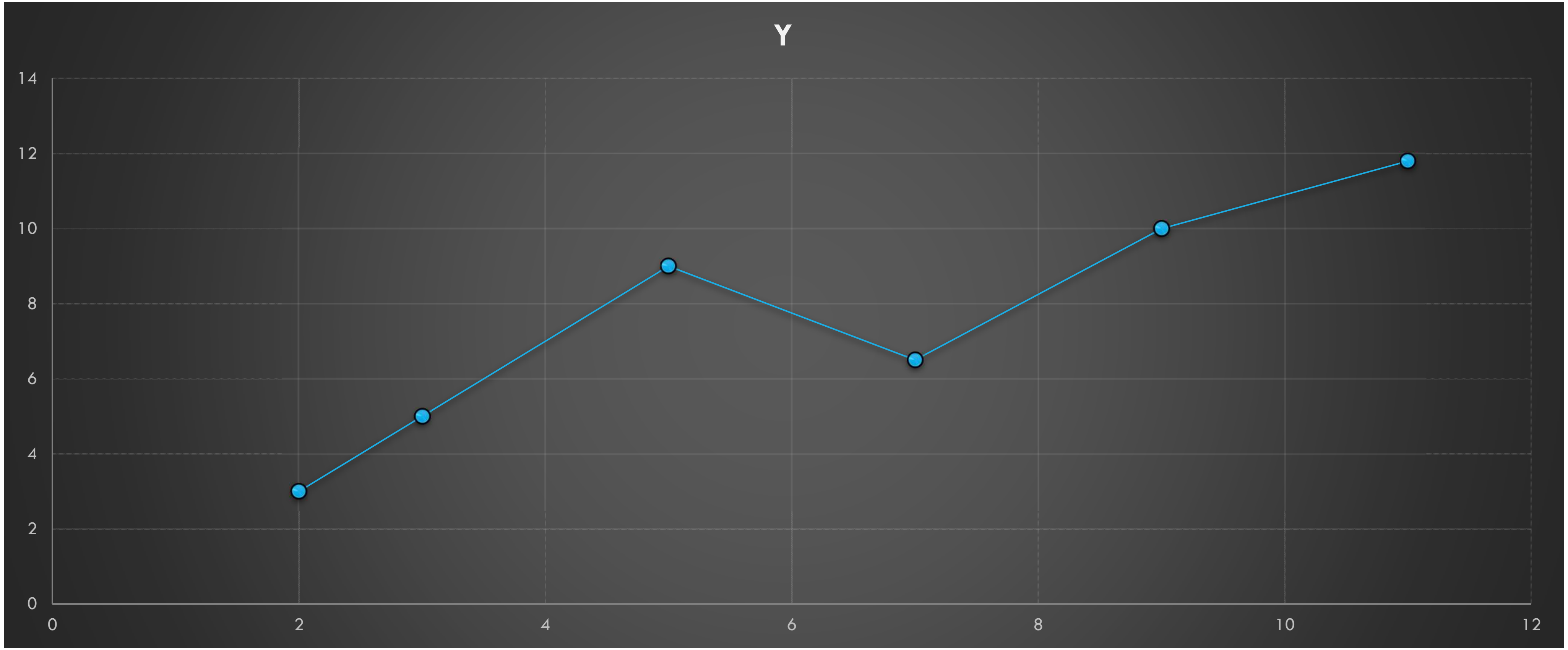
QUIZ

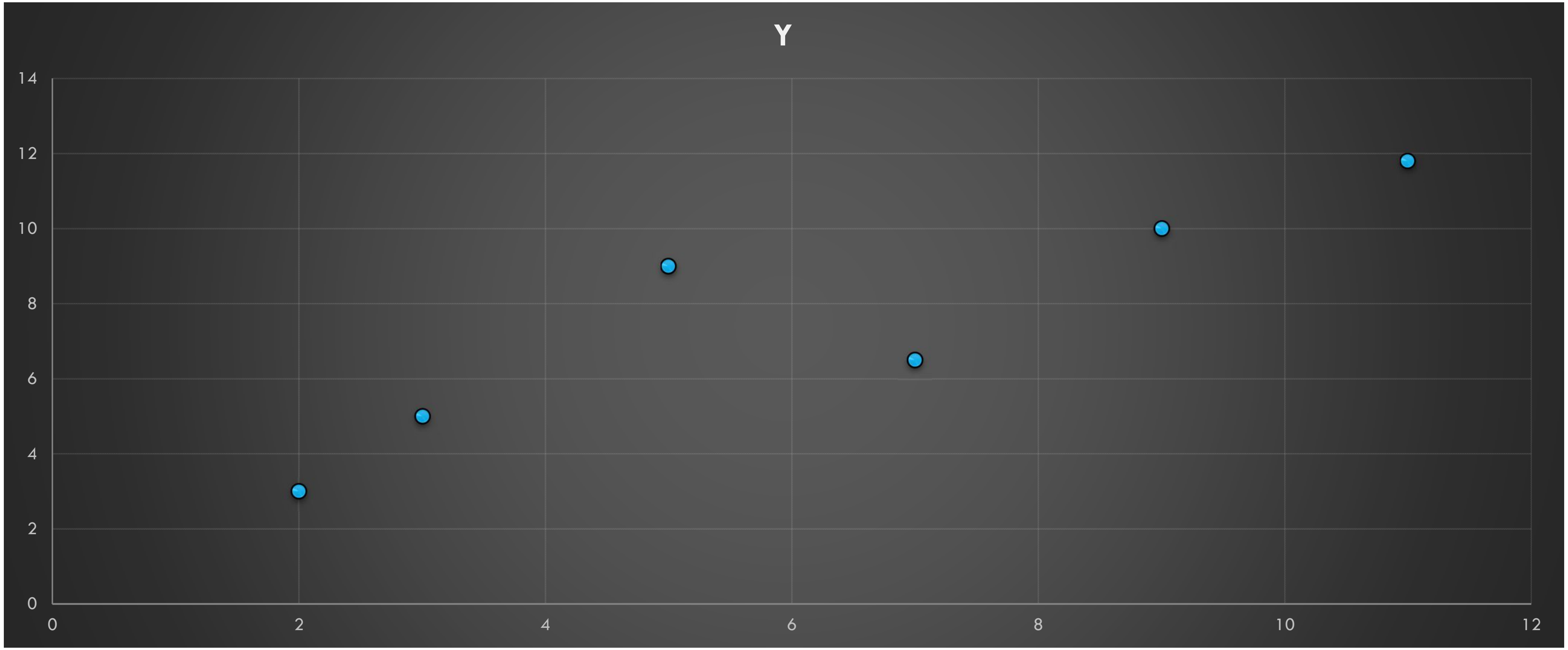
X	Y
2	3
3	5
5	9
9	10
7	6.5
11	11.8
10.5	WHAT?

QUIZ

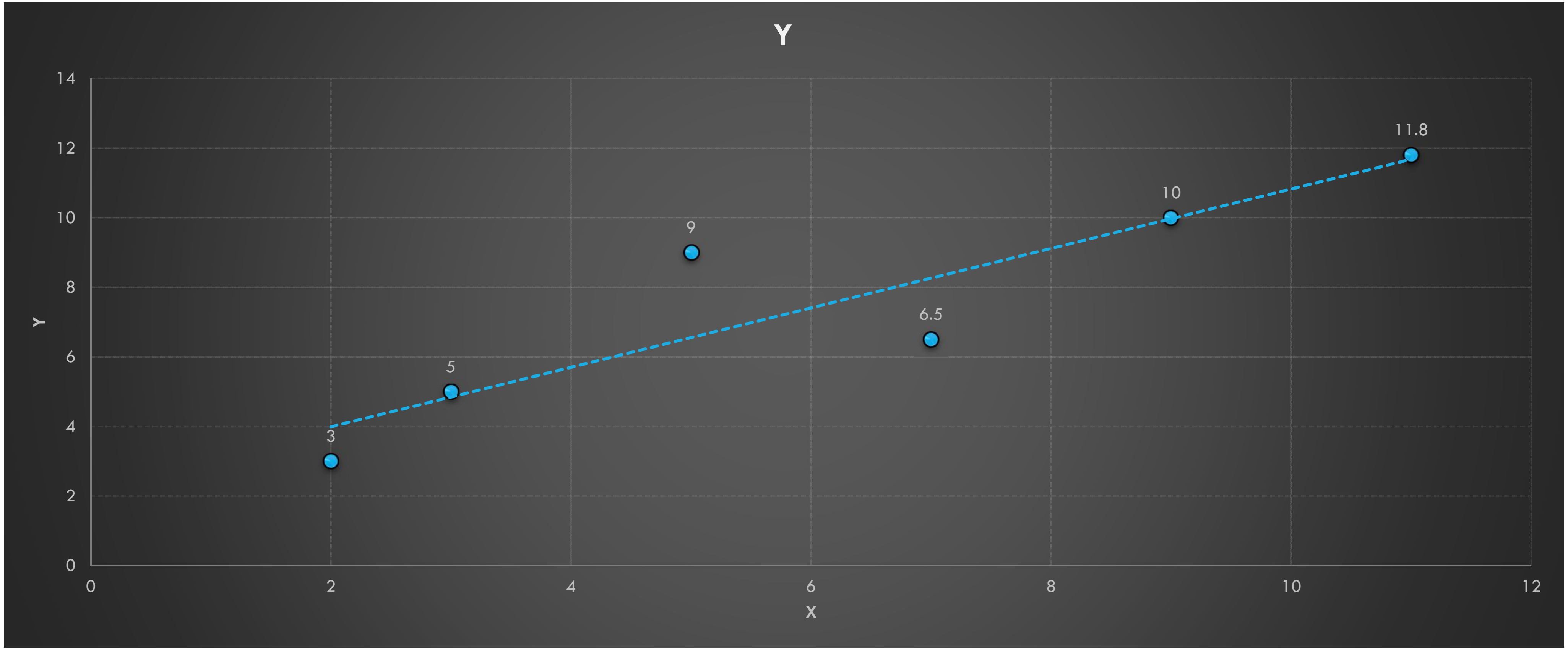
X	Y
2	3
3	5
5	9
9	10
7	6.5
11	11.8
10.5	Is it difficult to find out the relation ?

GRAPH IS SOLUTION ?




Y
14
12
10
8
6
4
2
0
0
2
4
6
8
10
12

APPROXIMATION



FIND THE EQUATION OF LINE ?

Two Points are given (3, 5) and (9,10)

Find equation of line ?

First right answer = 1Choclate (with in 90 secs)

What will be slope (m) and y intercept (c)?

$Y = m.X + c$

FIND THE EQUATION OF LINE ?

Two Points are given (3, 5) and (9,10)

Find equation of line ?

First right answer = 1Choclate

What will be slope (m) and y intercept (c)?

$Y = m.X + c$

$Y = 0.83X+2.5$

DEFINITION

Finding the relation between dependent variable and Independent variable is called Linear Regression.

OR

Finding the best fit line between dependent variable and Independent variable is called Linear Regression.

DEFINITION

Finding the relation between dependent variable and Independent variable is called

Linear Regression.

Now X=10.5 , X=2 , 13, 7

Y= what?

Put the values in equation Y=0.83X+2.5

What are you doing here? (USES OF LINEAR REGRESSION)

DEFINITION

Finding the relation between dependent variable and Independent variable is called

Linear Regression.

Now X=10.5 , X=12 , 13, 2.5

Y= what?

What are you doing here?

(USES OF LINEAR REGRESSION)

FORECASTING

PREDICTION

THE ERROR (RESIDUALS)

ERROR =GIVEN DATA(ACTUAL DATA) – PREDICTED DATA

e=Y(actual) –Y(predicted)

THE ERROR (RESIDUALS)

ERROR =GIVEN DATA(ACTUAL DATA) – PREDICTED DATA

e=Y(actual) –Y(predicted)

"Question"

"How can we find the best fit line?"

THE ERROR (RESIDUALS)

ERROR =GIVEN DATA(ACTUAL DATA) – PREDICTED DATA

e=Y(actual) –Y(predicted)

“Question”

“How can we find the best fit line?”

“Answer”

If Y(actual) =Y(predicted) or e=0

Or

Minimise the error

HOW TO FIND BEST FIT LINE

Derivation of linear regression equations : (FOR SINGLE VARIABLE)

given a set of n points (X_i, Y_i) on a scatterplot,

find the best-fit line, Y_i' =a+ bX_i

such that the sum of squared errors in Y, $\sum(Y_i - Y_i')^2$ is minimized.

SSE method or least Square method:

(Sum of Squared Error method)

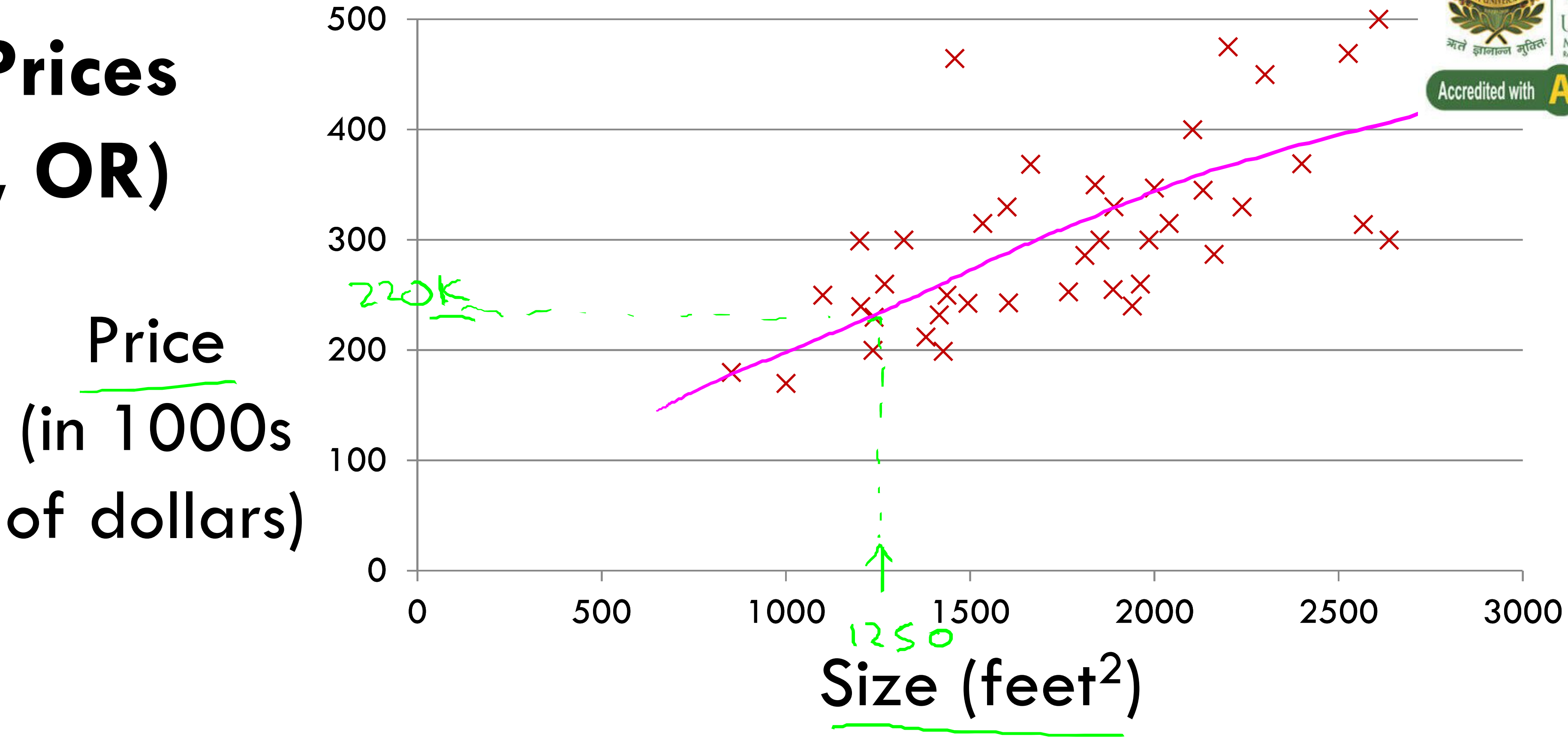
Find the a (y intercept), b (slope of line).

Linear regression with one variable

MODEL REPRESENTATION*

SRC : * Andrew NG

Housing Prices (Portland, OR)



<u>Supervised Learning</u>

Given the "right answer" for each example in the data.

<u>Regression Problem</u>

Predict real-valued output

Classification: Discrete-valued output

Training set of housing prices (Portland, OR)

Size in feet² (x)	Price ($) in 1000's (y)
2104	460
1416	232
1534	315
852	178
…	…

$m = 47$

Notation:

- **m** = Number of training examples
- **x**'s = "input" variable / features
- **y**'s = "output" variable / "target" variable

(x, y) - one training example

$(x^{(i)}, y^{(i)})$ - i^{th} training example

$x^{(1)} = 2104$

$x^{(2)} = 1416$

$y^{(1)} = 460$

Training Set

↓

Learning Algorithm

↓

Size of house → h → Estimated price

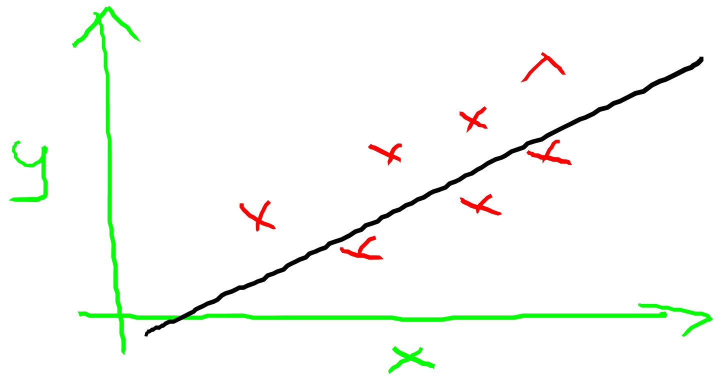
x hypothesis (estimated value of y)

h maps from x's to y's.

How do we represent h ?

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

Shorthand: $h(x)$



$h(x) = \theta_0 + \theta_1 x$

Linear regression with one variable. (x)

Univariate linear regression.

└ one variable

Linear regression with one variable

COST FUNCTION*

SRC : * Andrew NG

Training Set

Size in feet^2 (x)	Price (\$) in 1000's (y)
2104	460
1416	232
1534	315
852	178
…	…

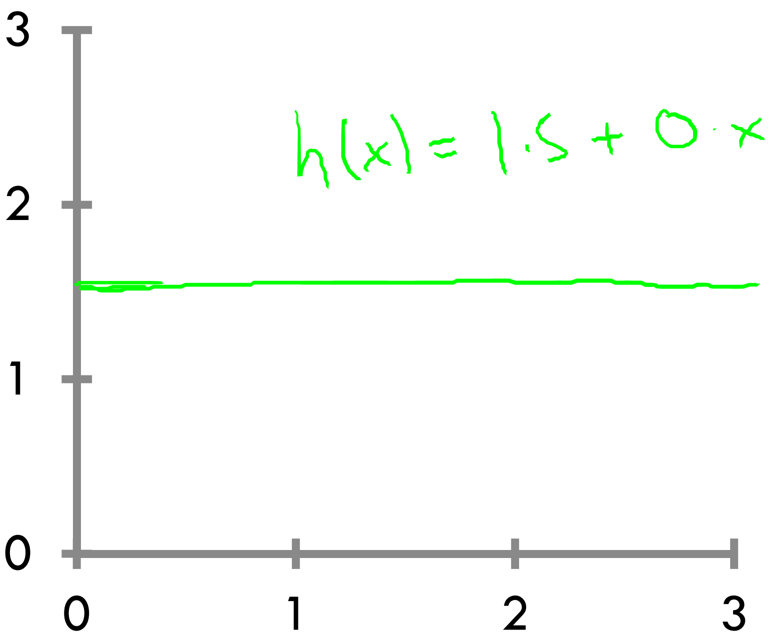
$m = 47$

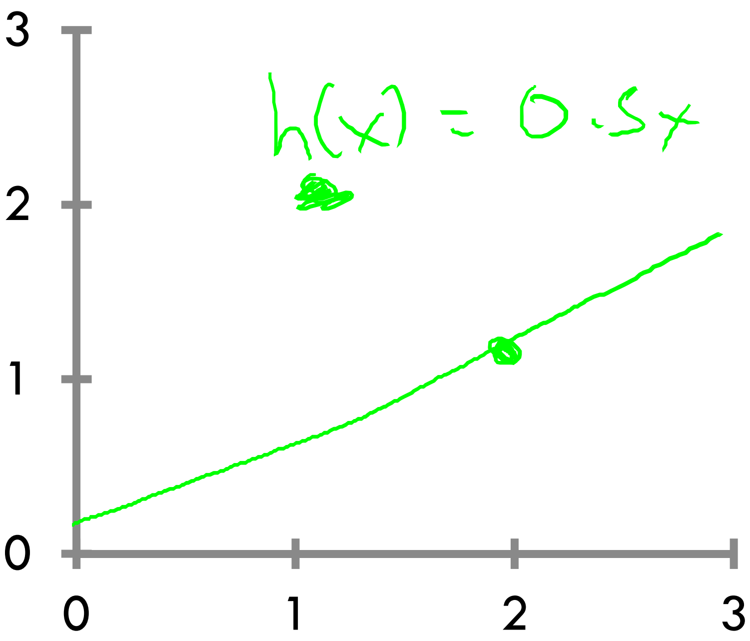
Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

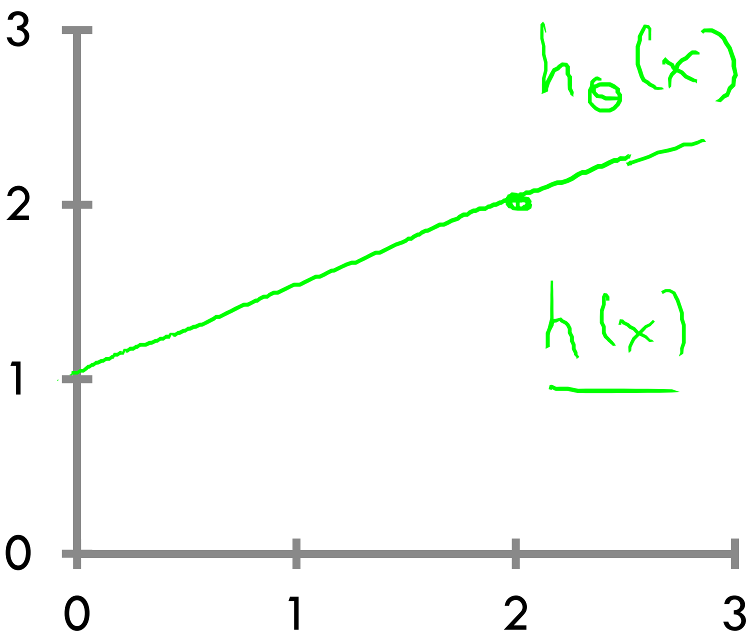
θ_i's: Parameters

How to choose θ_i's ?

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



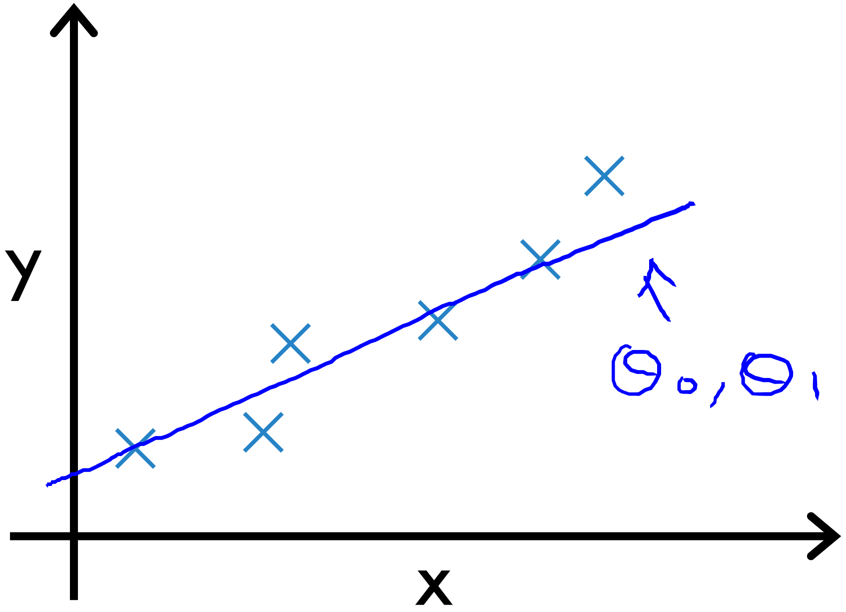




→ $\theta_0 = 1.5$
→ $\theta_1 = 0$

→ $\theta_0 = 0$
→ $\theta_1 = 0.5$

→ $\theta_0 = 1$
→ $\theta_1 = 0.5$



θ_0, θ_1

$(x^{(i)}, y^{(i)})$

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

x, y

training examples

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

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

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

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

Squared error function

Linear regression with one variable

COST FUNCTION INTUITION I *

SRC : * Andrew NG

Simplified

Hypothesis:

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

Parameters:

$$\theta_0, \theta_1$$

Cost Function:

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

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

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

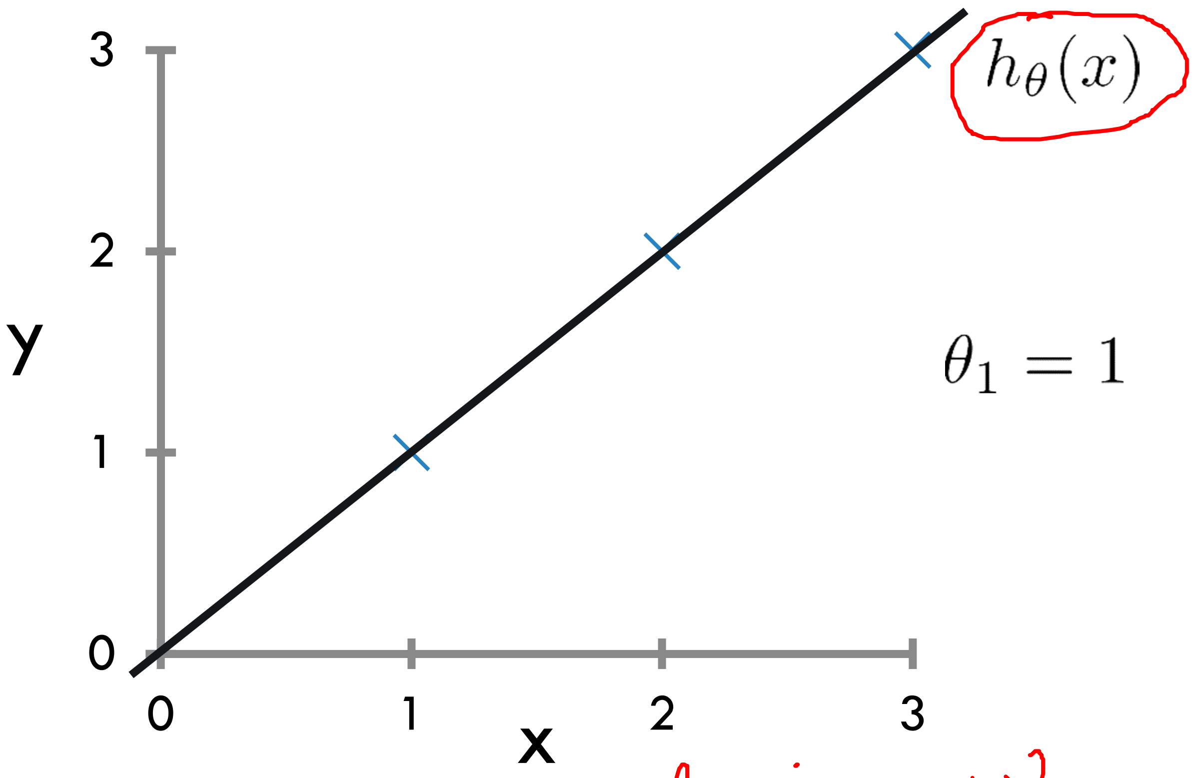
$$\theta_1$$

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

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

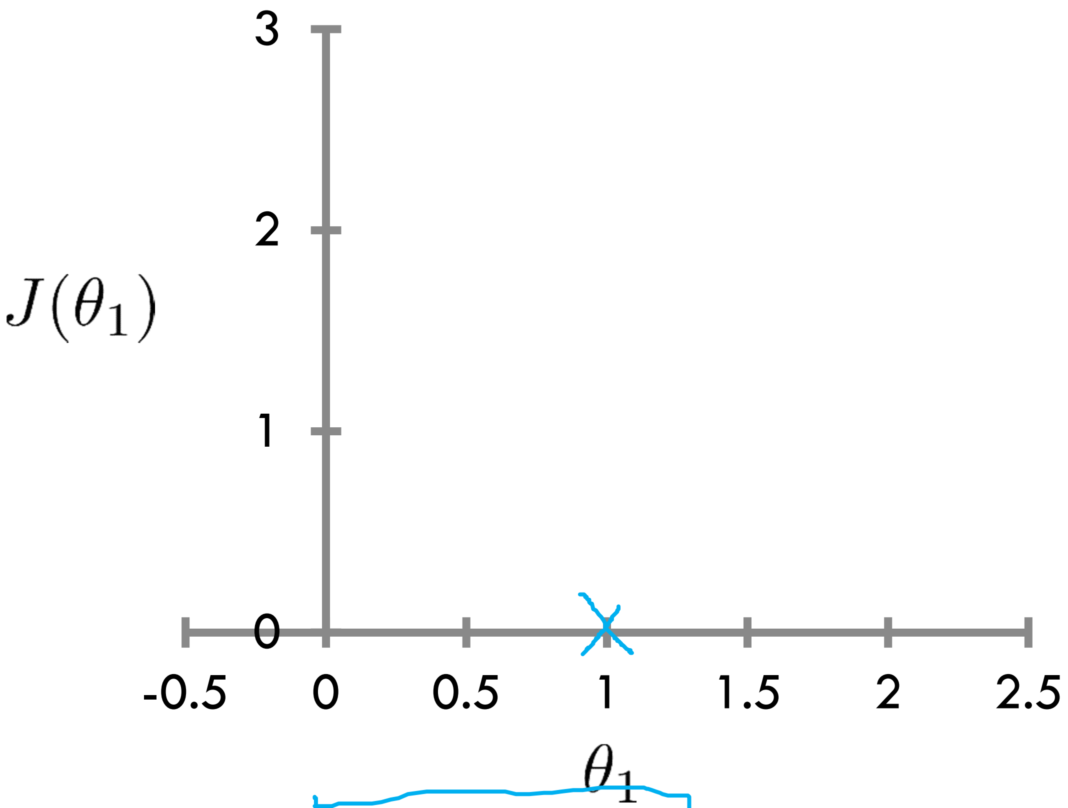
$h_\theta(x)$

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



$J(\theta_1)$

(function of the parameter θ_1)



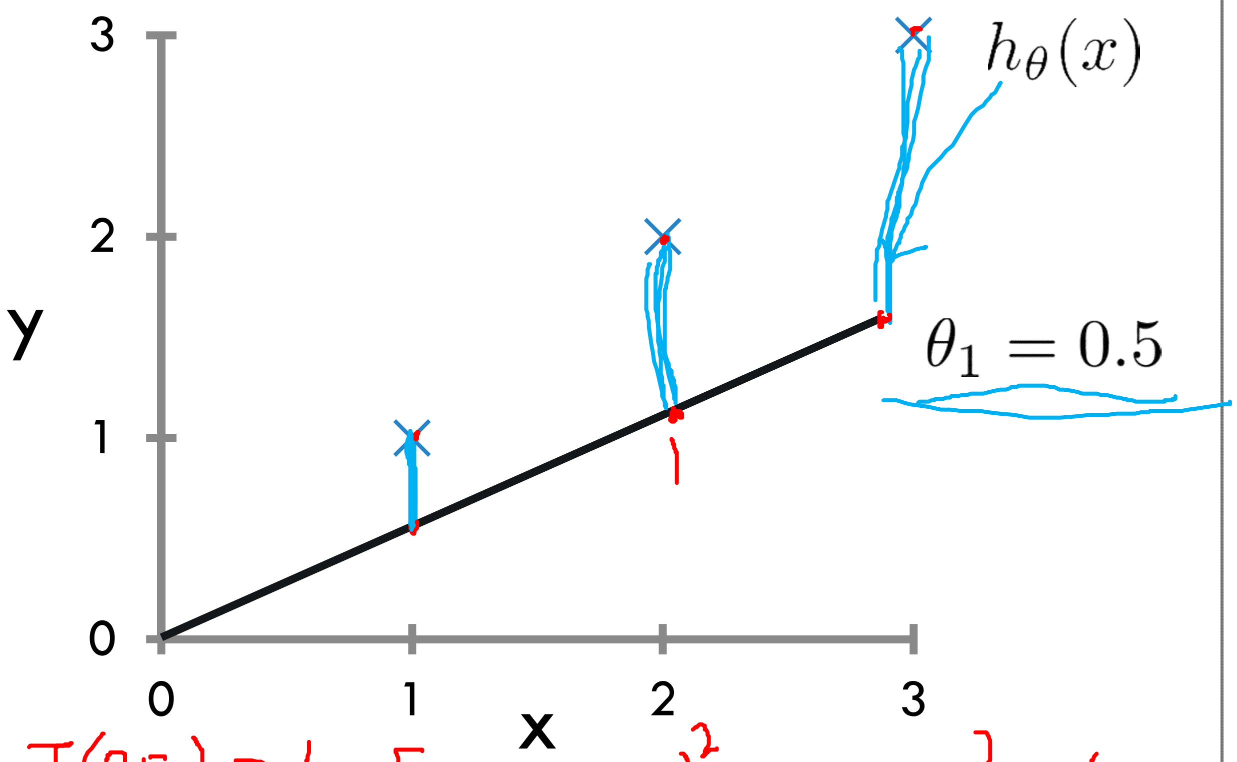
$$J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^i) - y^i\right)^2$$

$$= \frac{1}{2\times 3}\left(0^2 + 0^2 + 0^2\right) = 0$$

$$\boxed{J(1) = 0}$$

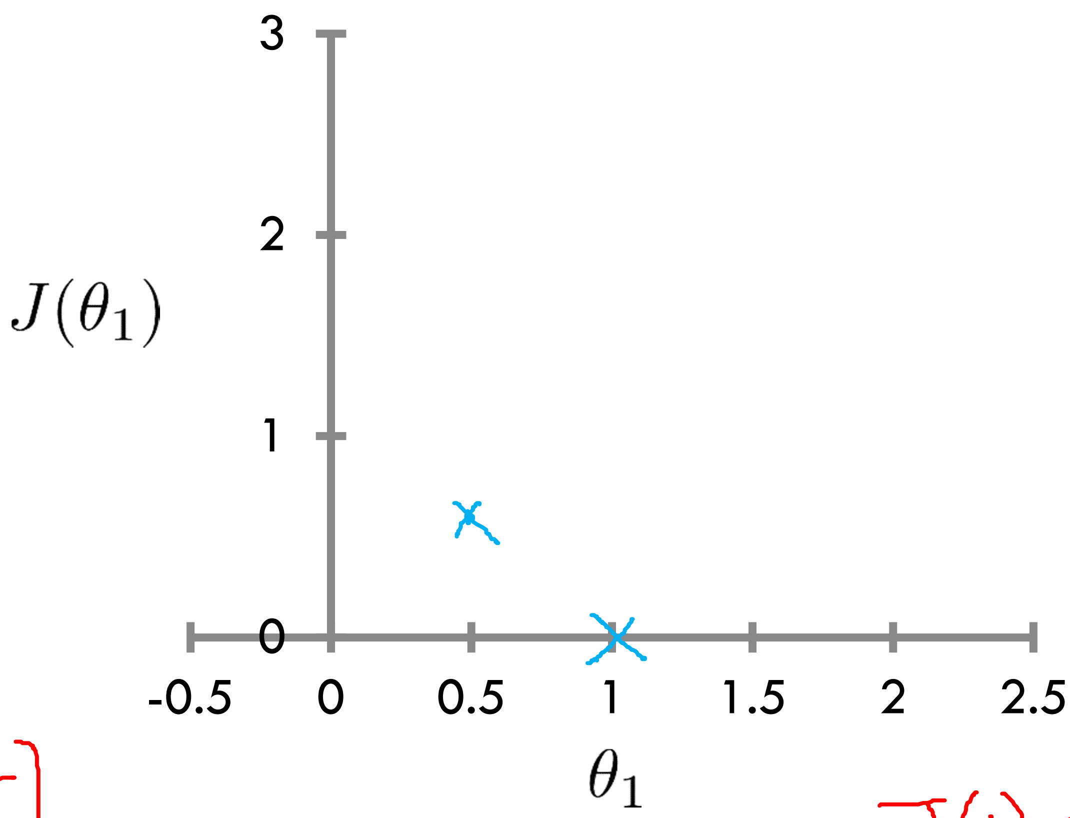
$h_\theta(x)$

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



$J(\theta_1)$

(function of the parameter θ_1)



$J(0.5) = \frac{1}{2\times 3}\left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right]$

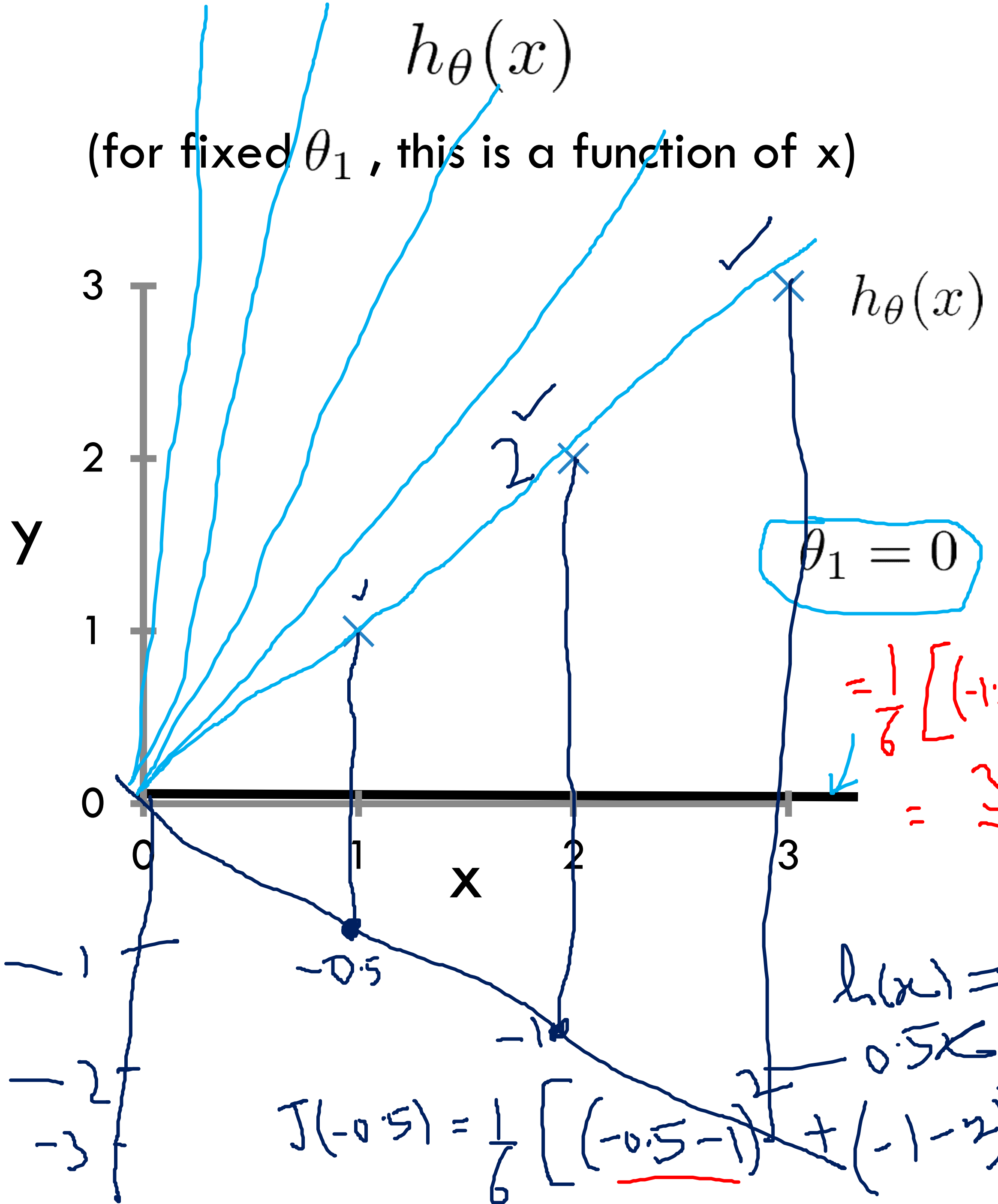
$= \frac{1}{6}\left((0.5)^2 + (-1)^2 + (-1.5)^2\right) = \frac{3.5}{6} \approx 0.58$

$J(1) = 0$

$J(0.5) = 0.58$

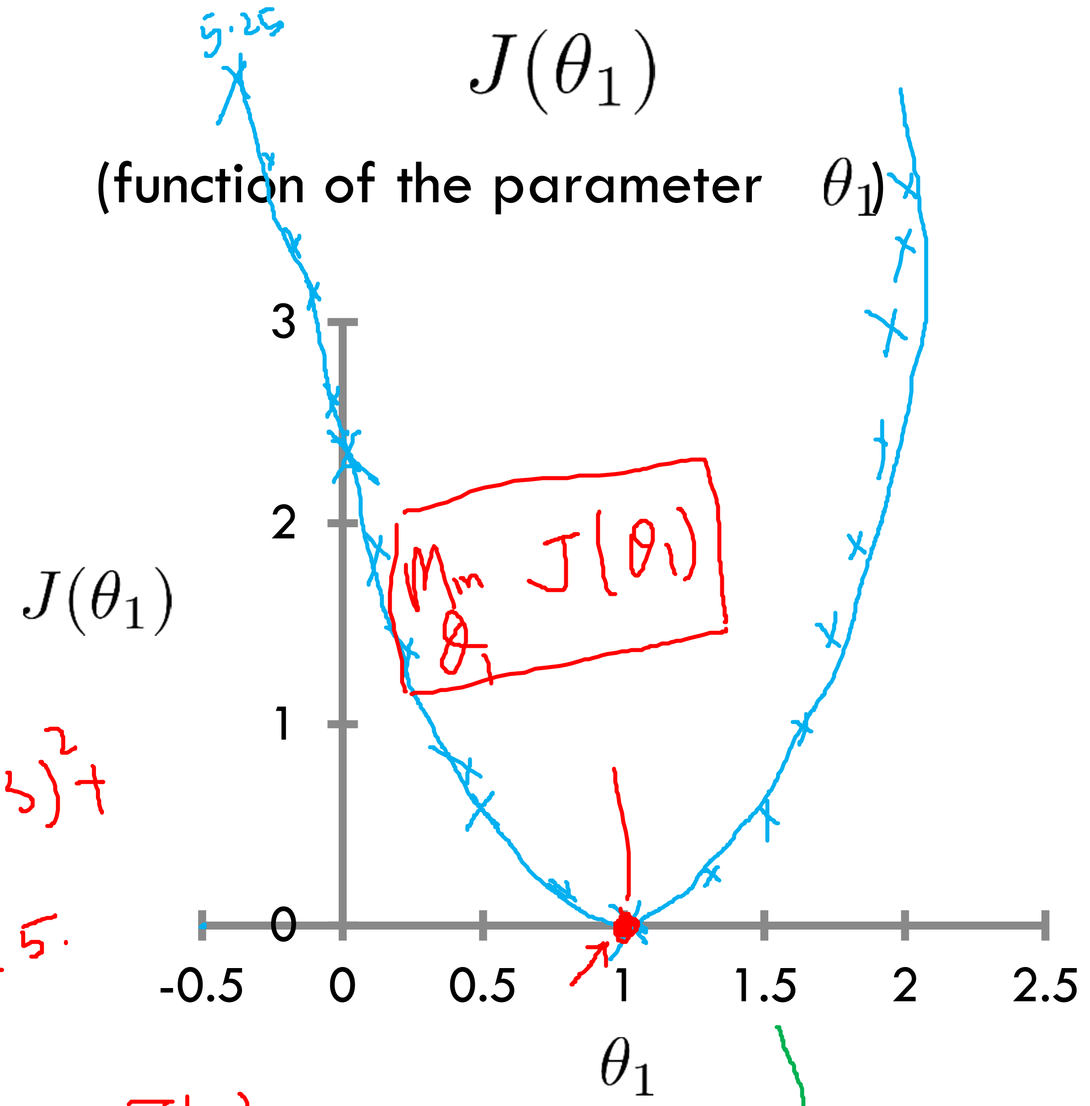
$h_\theta(x)$

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



$J(\theta_1)$

(function of the parameter θ_1)



$= \frac{1}{6}\left[(-1.5)^2 + (-3)^2 +\right.$

$= \frac{31.5}{6} = 5.$

$h(x) = -0.5x$

$J(-0.5) = \frac{1}{6}\left[(-0.5-1)^2 + (-1-2)^2 + (-1.5-3)^2\right]$

$J(1) = 0$

$J(0.5) = 0.58$

$J(0) = 2.33$

$\theta_1 = -0.5$

$J(\theta_1) = ?$ $h(x) = -0.5x$

Linear regression with one variable

COST FUNCTION INTUITION II *

SRC : * Andrew NG

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

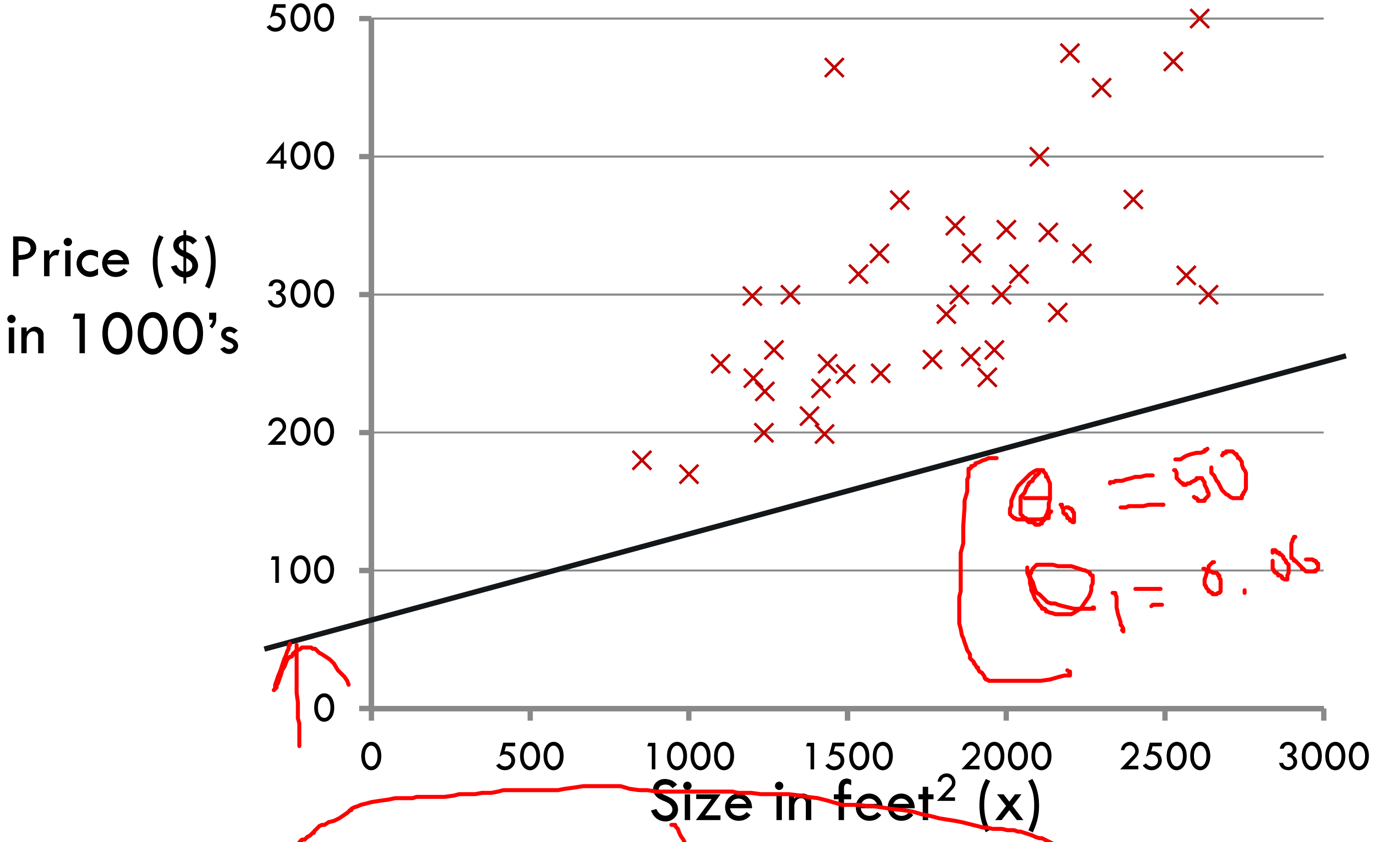
Parameters: θ_0, θ_1

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

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

$h_\theta(x)$

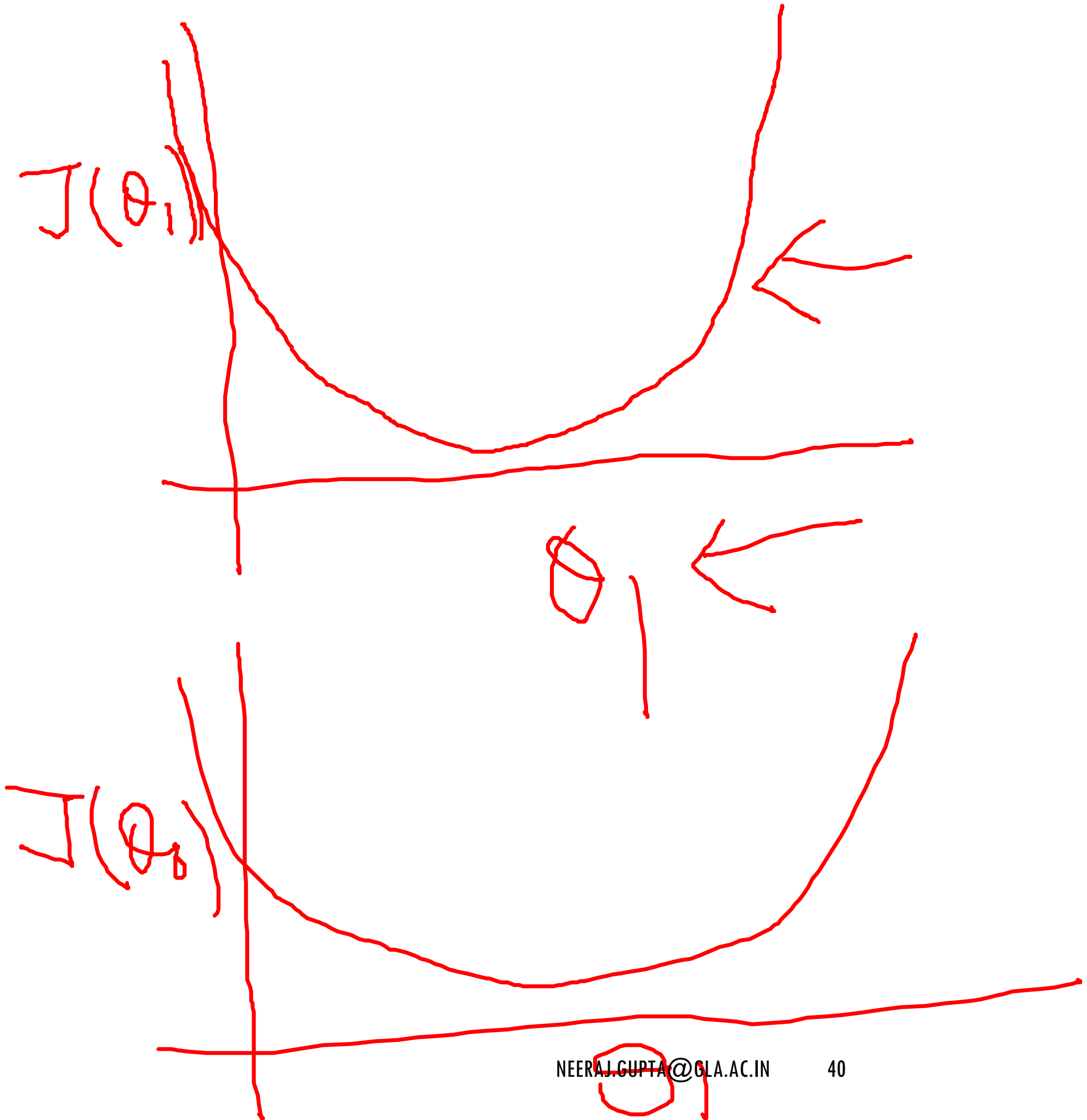
(for fixed θ_0, θ_1 , this is a function of x)



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

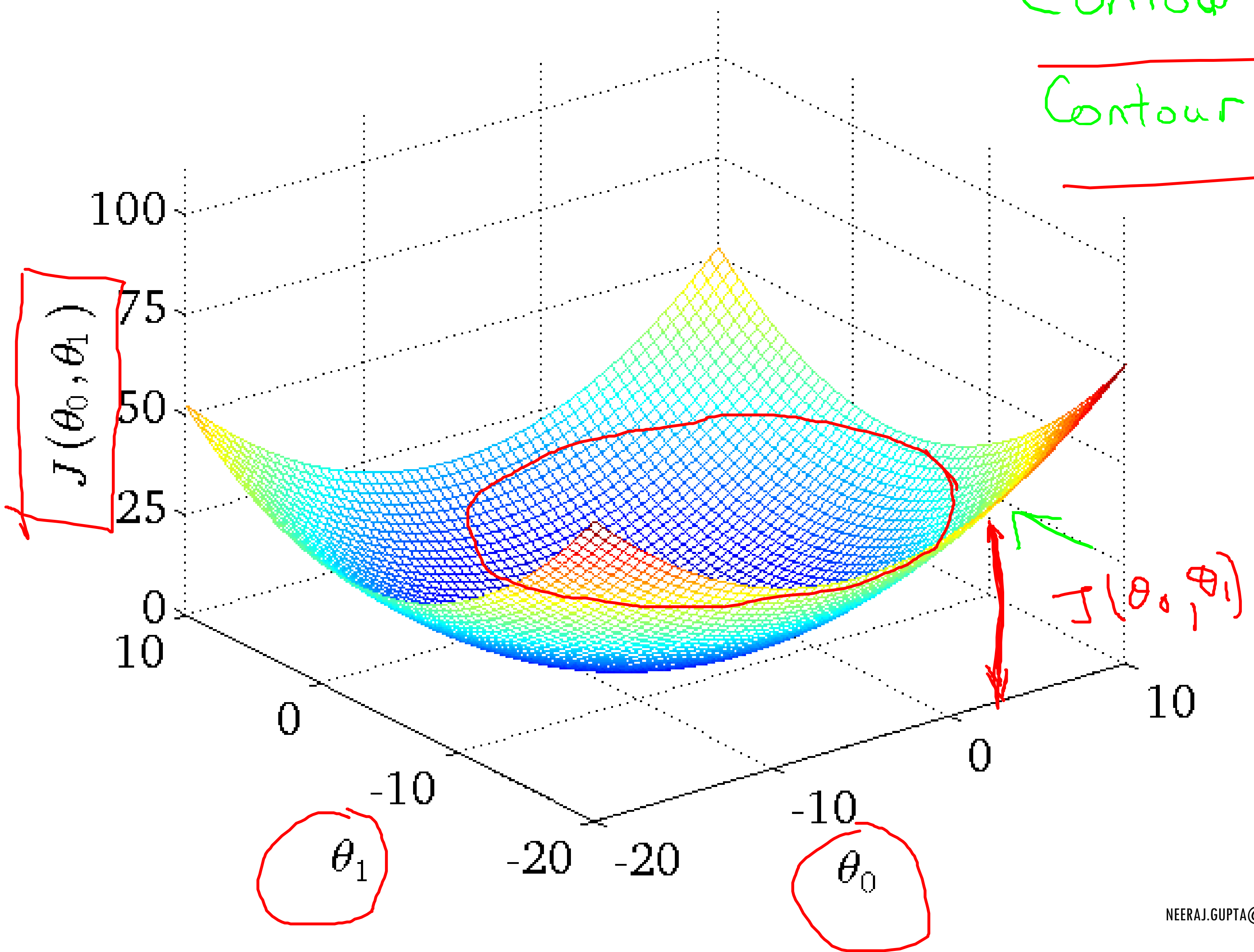
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



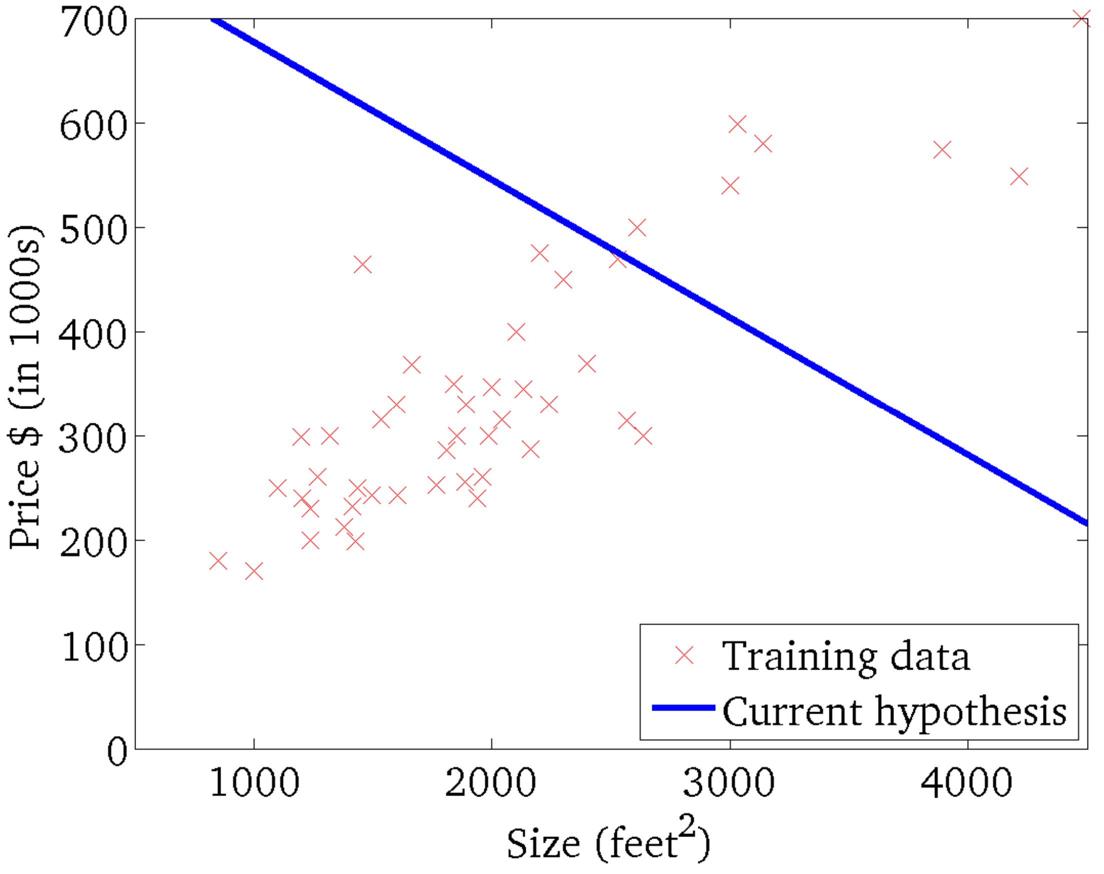
Contour plots

Contour figures -



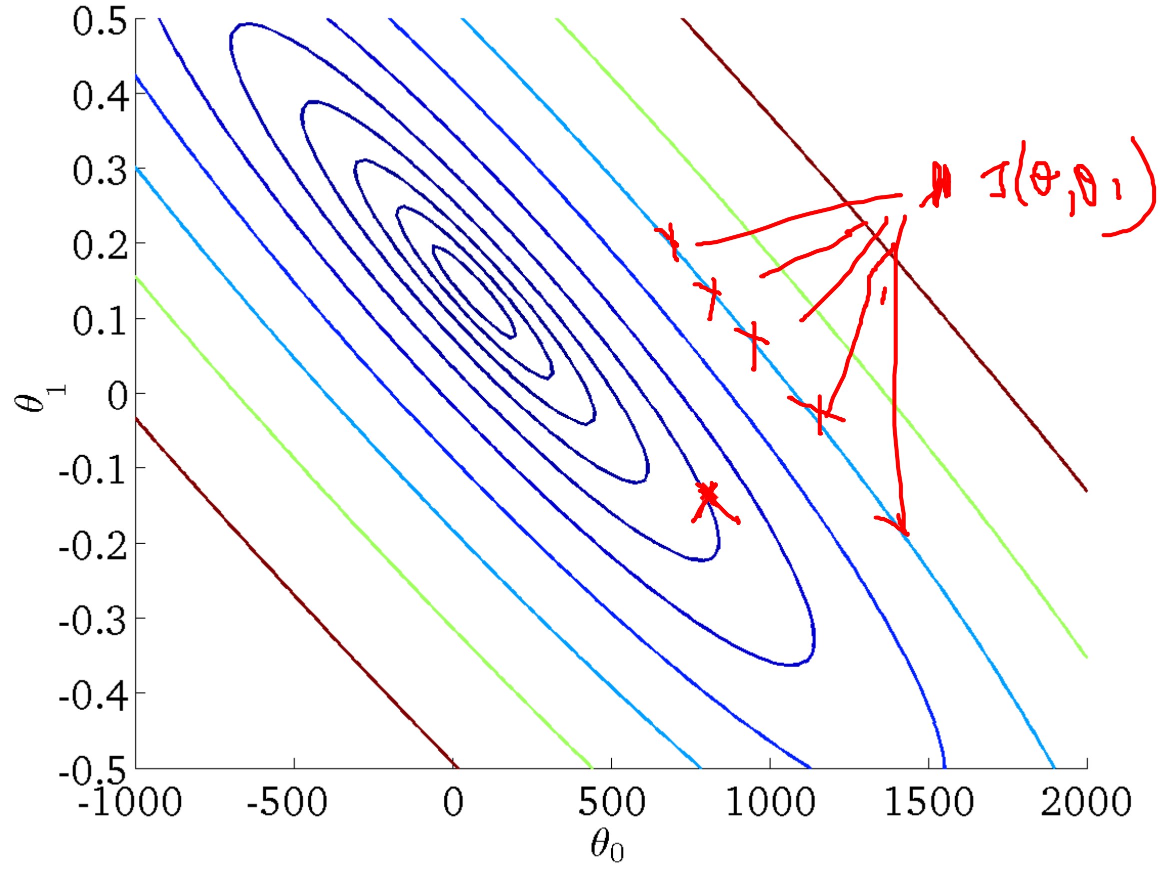
$h_\theta(x)$

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



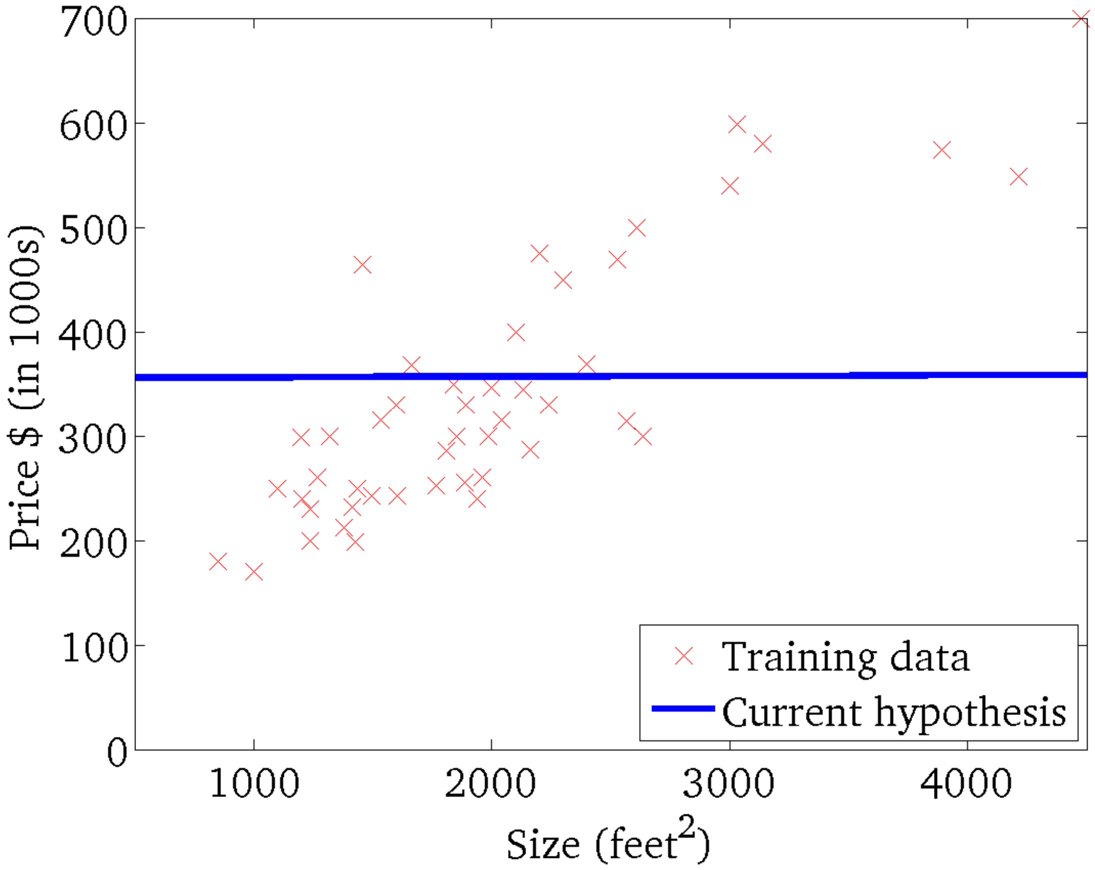
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



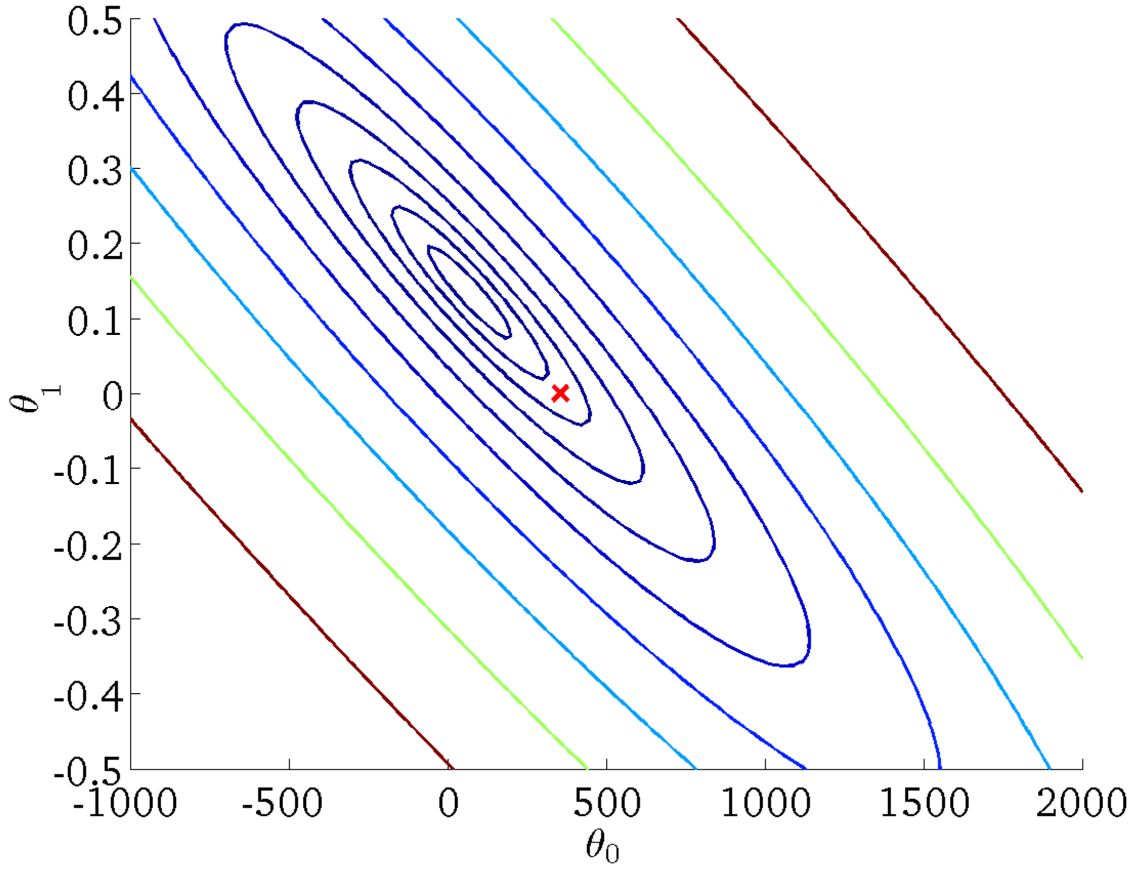
$h_\theta(x)$

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



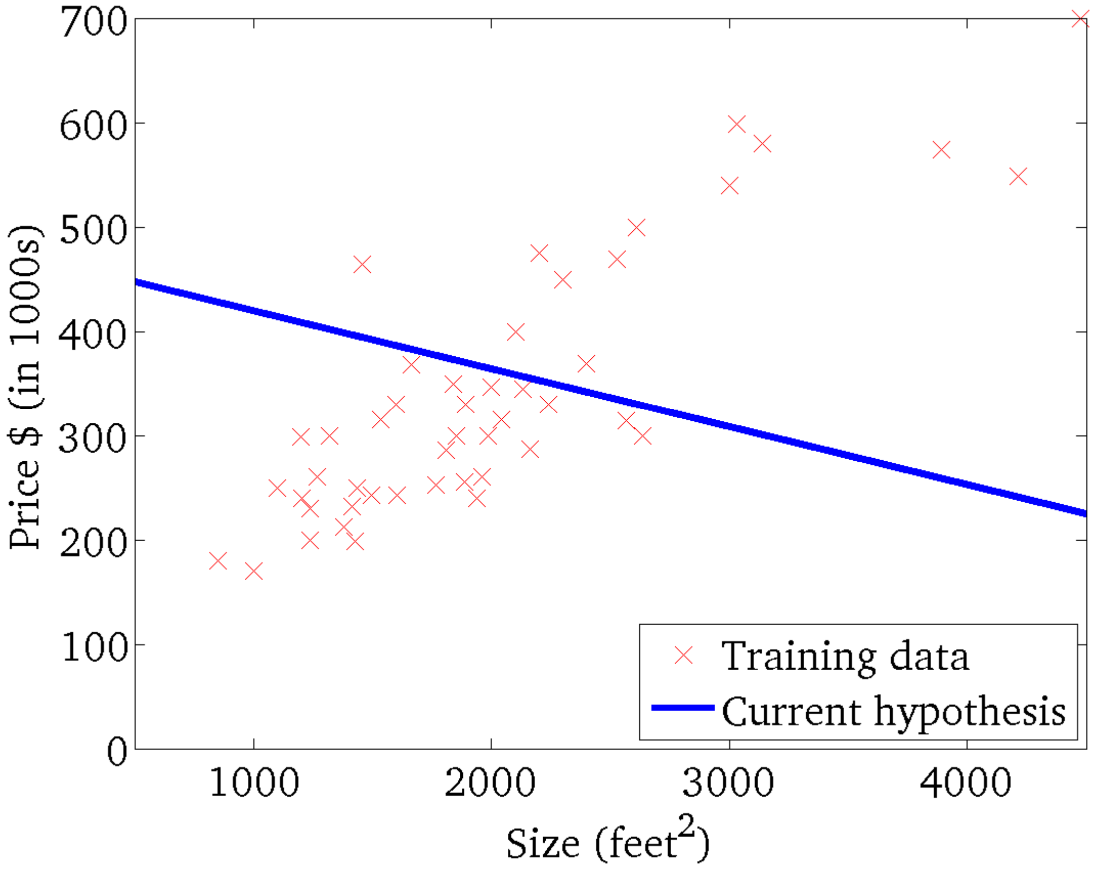
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



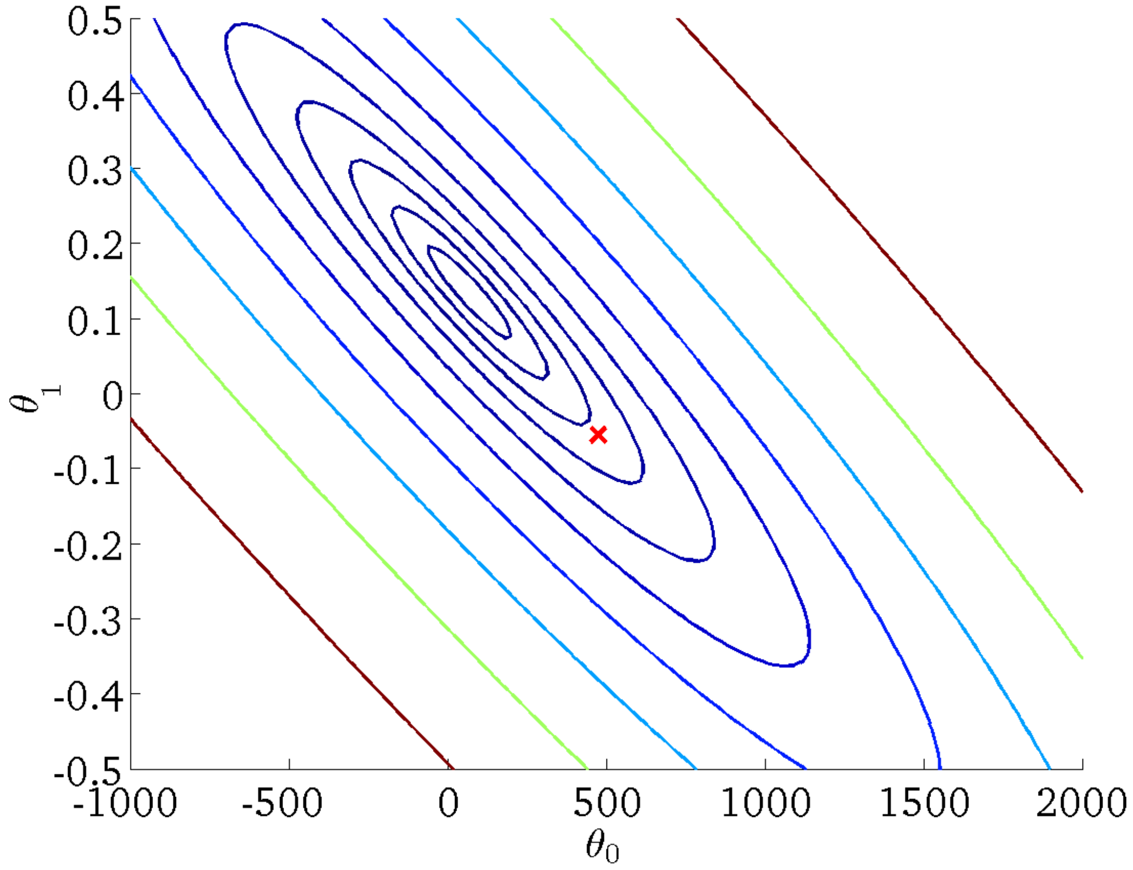
$h_\theta(x)$

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



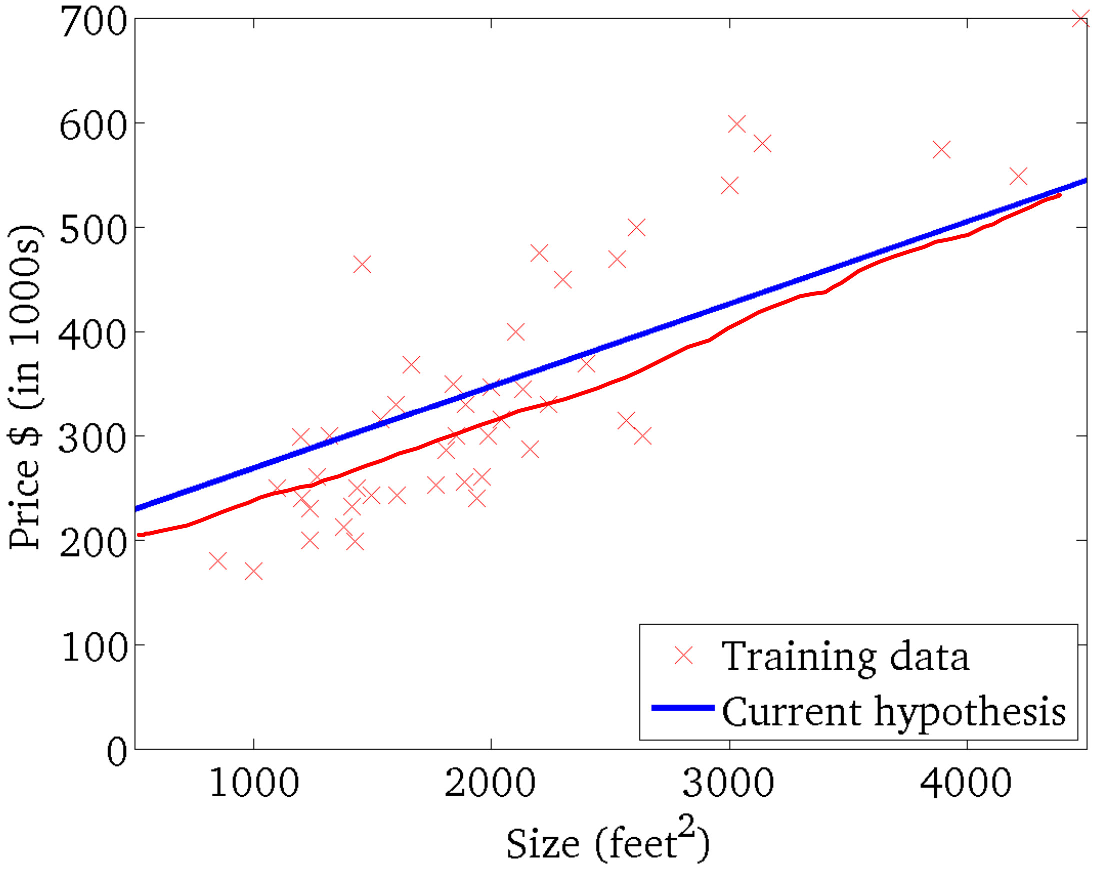
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



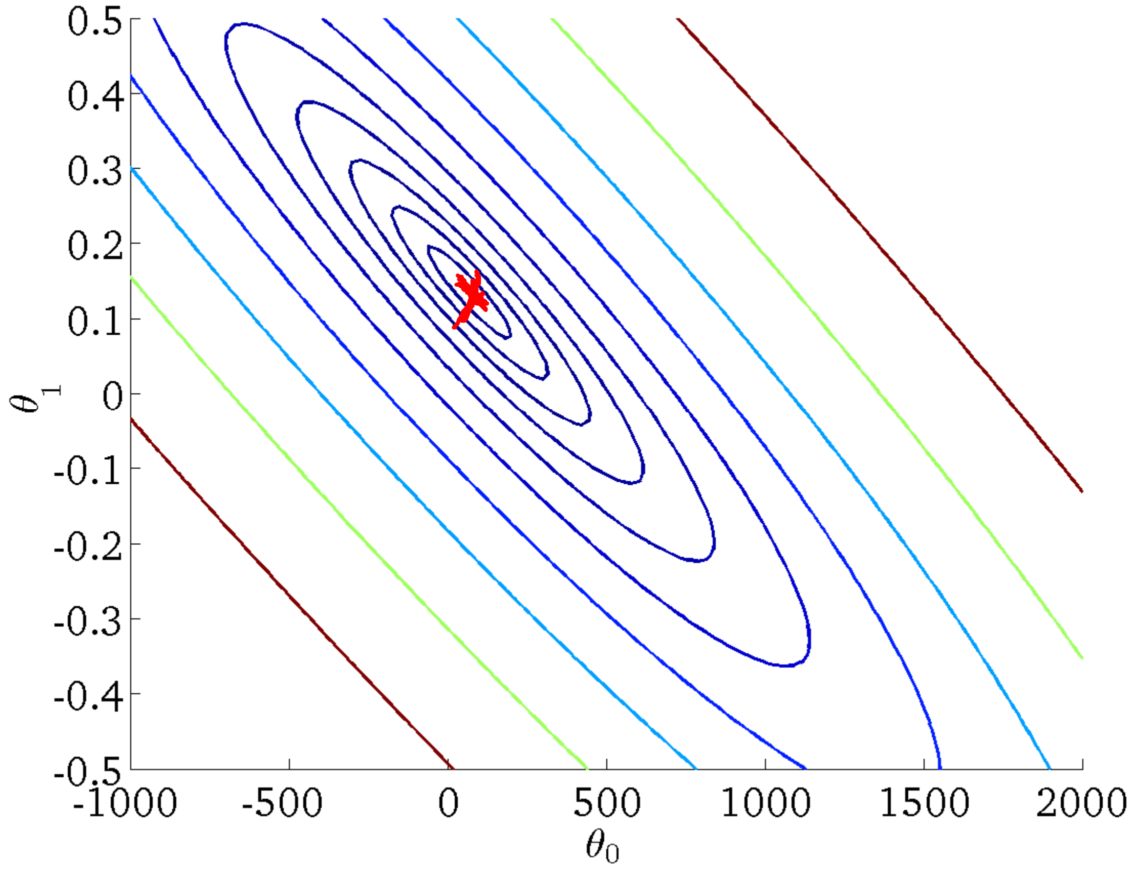
$h_\theta(x)$

(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)



Linear regression with one variable

GRADIENT DESCENT *

SRC : * Andrew NG

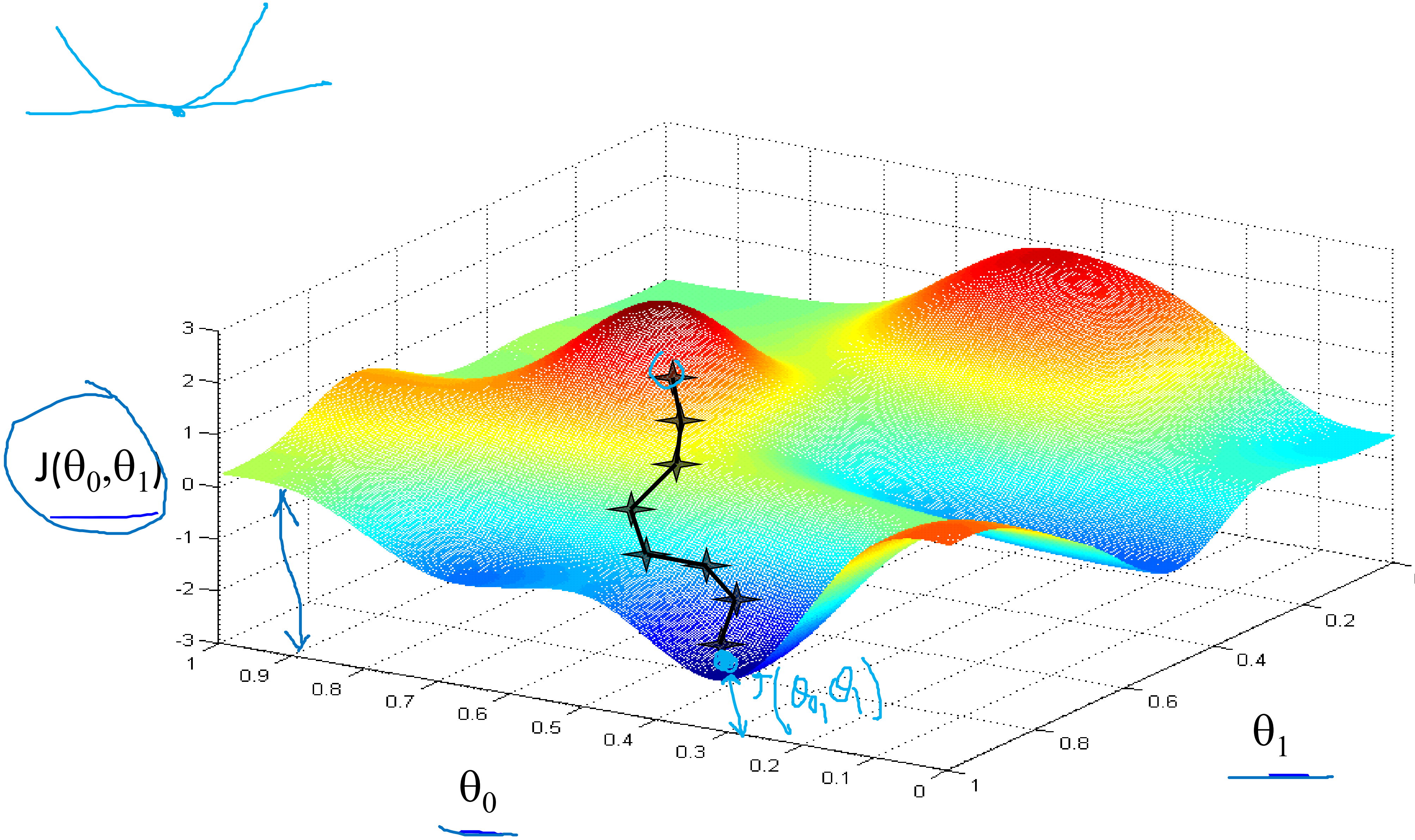
Have some function $J(\theta_0, \theta_1)$ $J(\theta_0, \theta_1, \ldots \theta_n)$

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

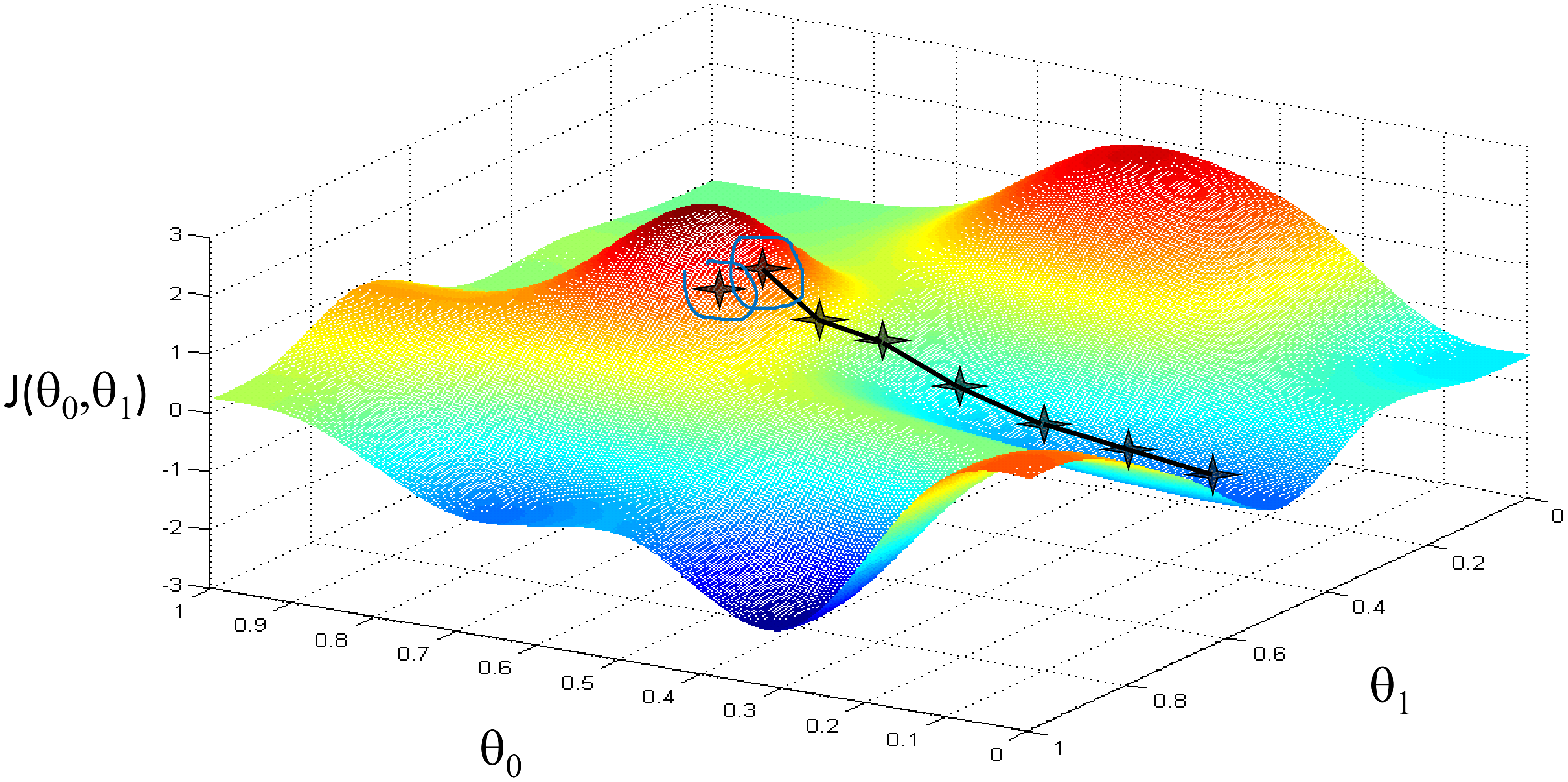
Outline:

- Start with some θ_0, θ_1 (Say $\theta_0 = 0$, $\theta_1 = 0$)
- Keep changing θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$ min

  until we hopefully end up at a minimum



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
0.2
0.4
0.6
0.8
1
$J(\theta_0,\theta_1)$
θ_0
θ_1
$J(\theta_0, \theta_1)$



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

Gradient descent algorithm

Learning Rate

repeat until convergence {

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

}

assignment

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:

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

Linear regression with one variable

GRADIENT DESCENT INTUITION*

SRC : * Andrew NG

Gradient descent algorithm

repeat until convergence {

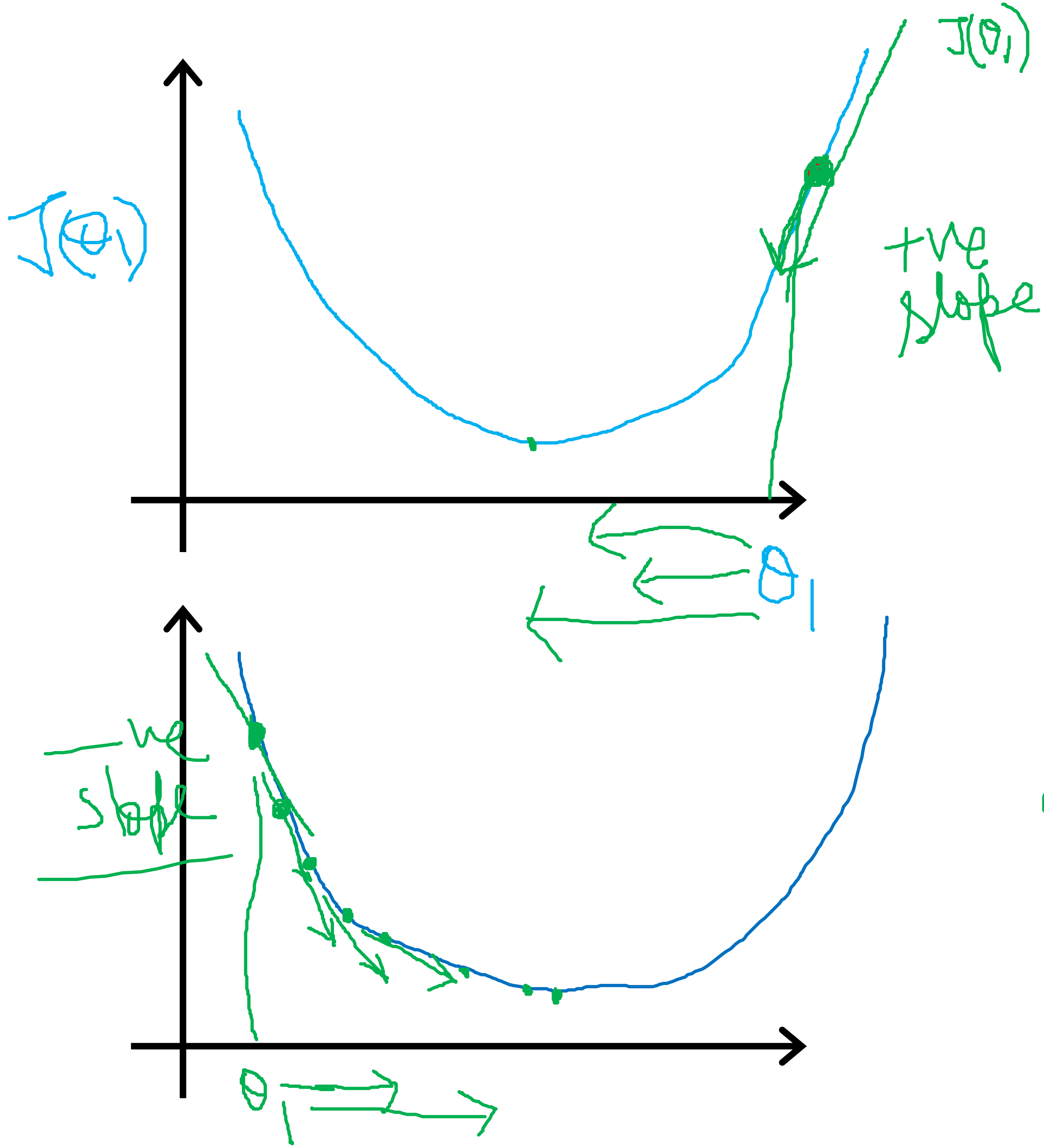
$$\theta_j := \theta_j - \alpha \boxed{\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)} \quad \text{(simultaneously update } j = 0 \text{ and } j = 1\text{)}$$

}

derivative

Learning Rate

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



$$\theta_1 = \theta_1 - \underset{0.01}{\alpha} \boxed{\frac{d}{d\theta_1} J(\theta_1)} \; > 0 \; +ve$$

$$\theta_1 := \theta_1 - \alpha\,(\text{positive no})$$

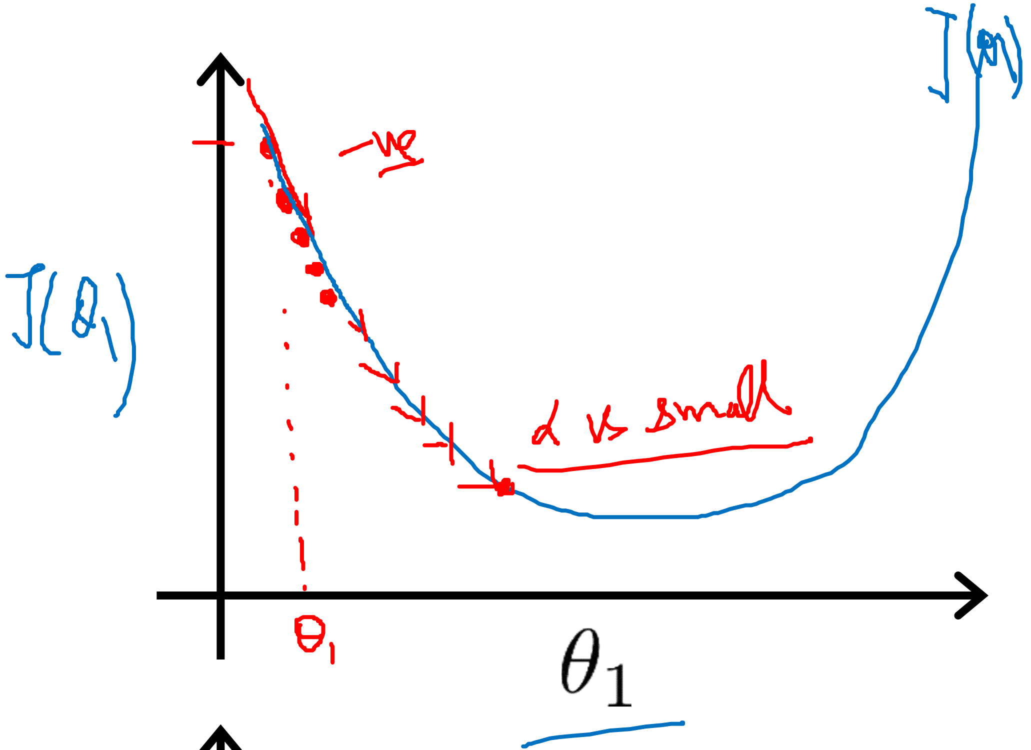
$$= \text{small}$$

$$\theta_1 := \theta_1 - \alpha\,(\text{negative no})$$

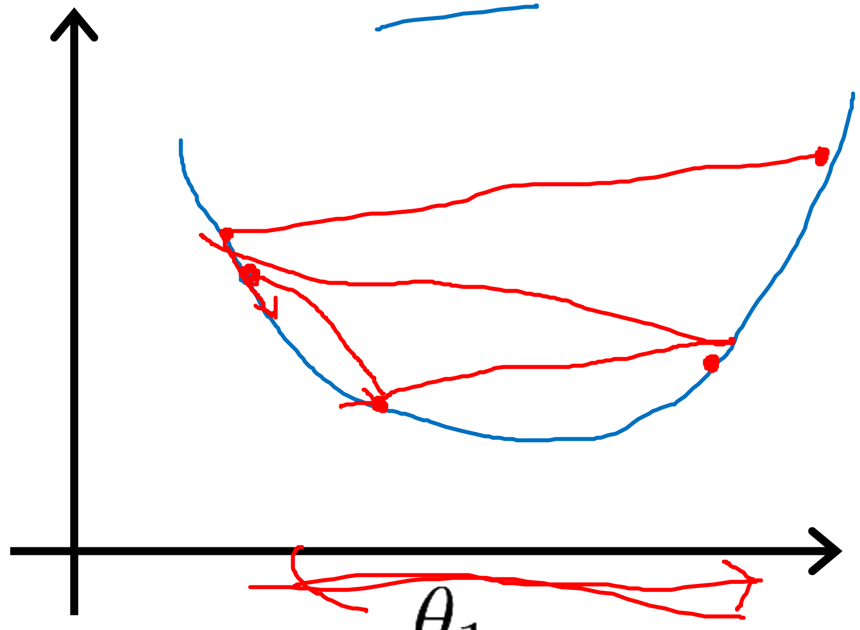
$$=$$

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

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.



Linear regression with one variable

GRADIENT DESCENT FOR LINEAR REGRESSION*

SRC : * Andrew NG

Gradient descent algorithm

repeat until convergence {

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

$$(\text{for } j = 1 \text{ and } j = 0)$$

}

Linear Regression Model

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

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

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

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

$$\cancel{x^2} \quad 2x$$

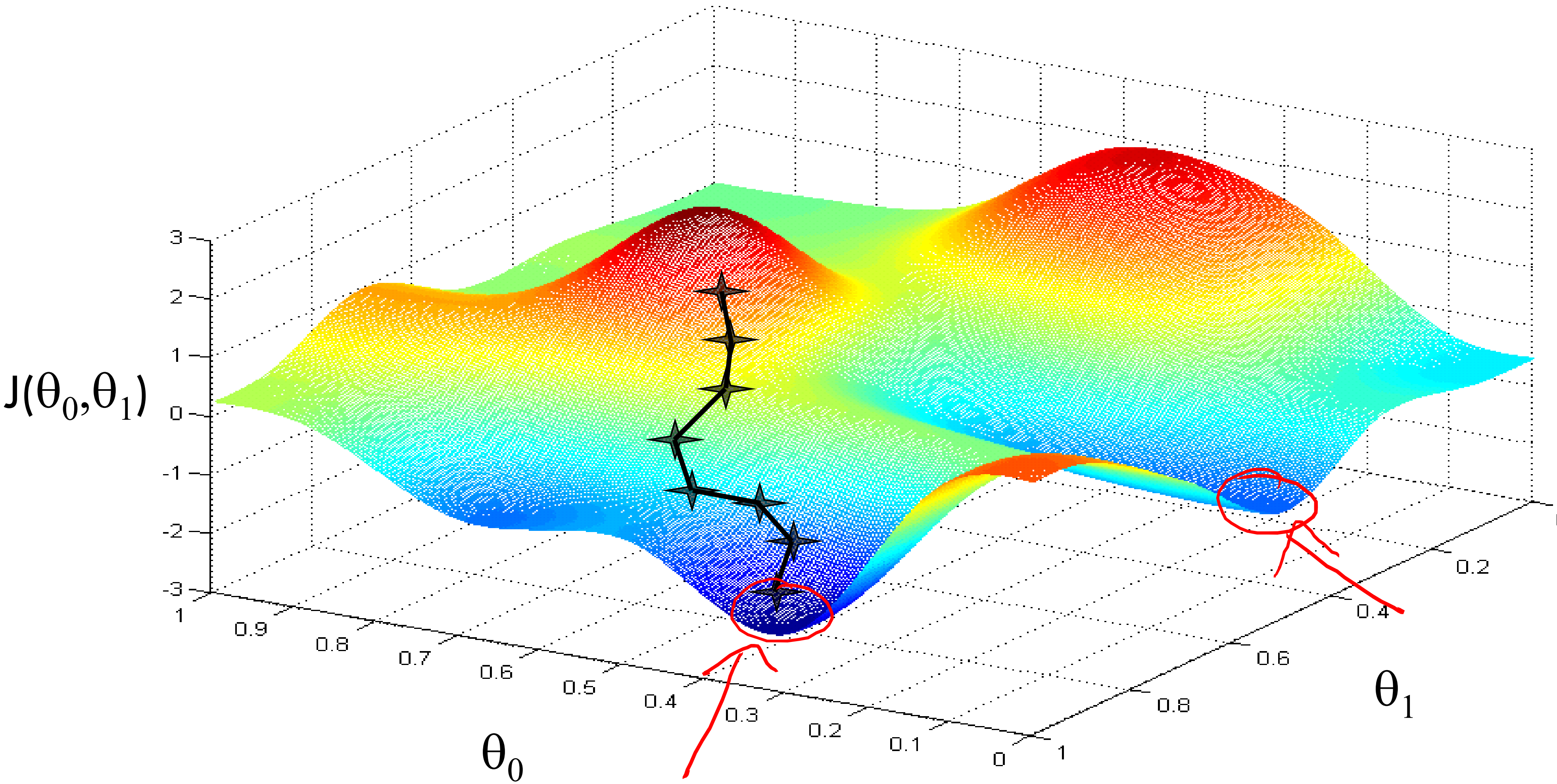
$$j = 0 : \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)$$

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

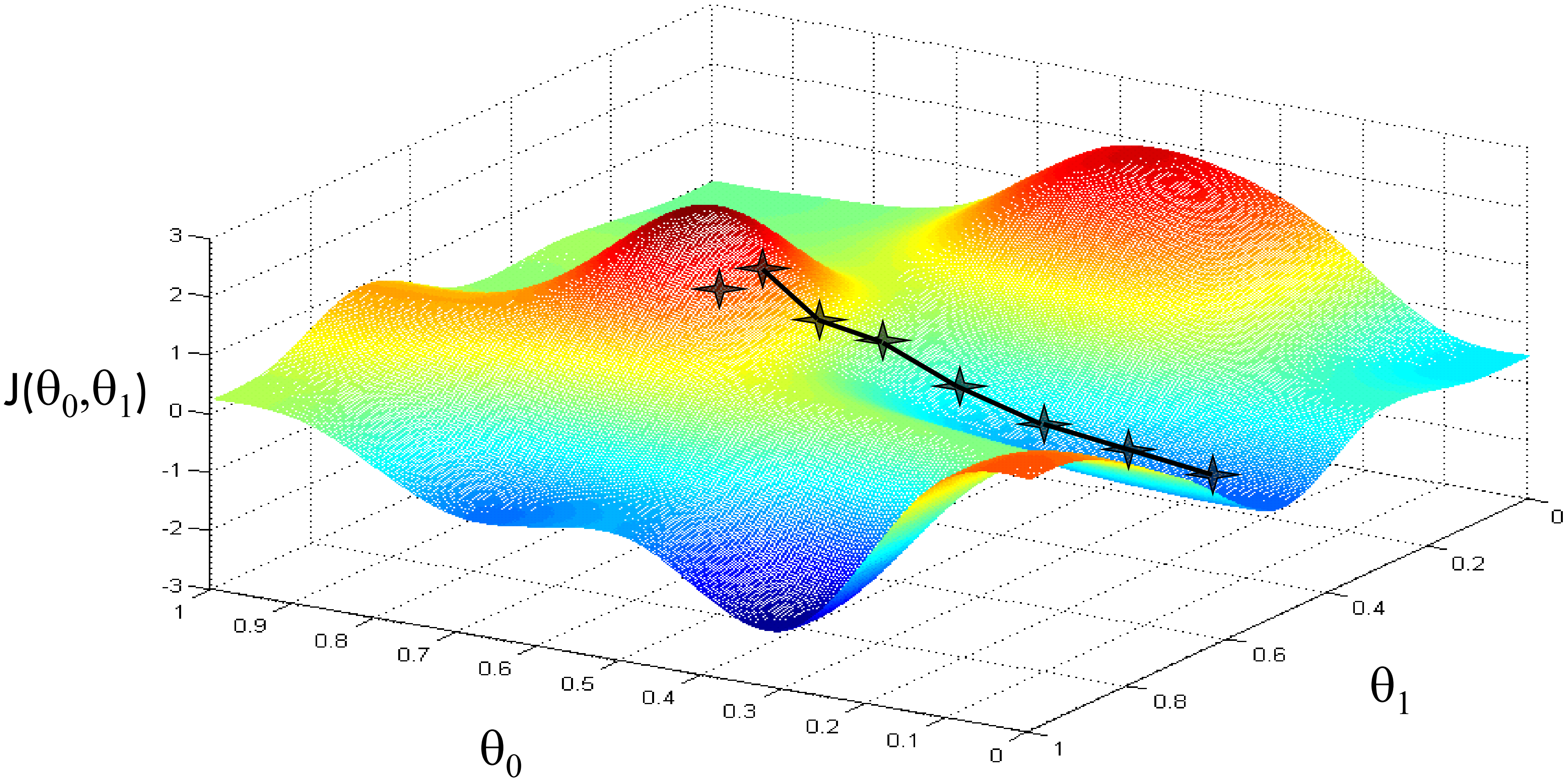
Gradient descent algorithm

$$
\begin{aligned}
&\text{repeat until convergence } \{ \\
&\quad \left.\begin{aligned}
\theta_0 &:= \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) \\
\theta_1 &:= \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) \cdot x^{(i)}
\end{aligned}\right\} \\
&\}
\end{aligned}
$$

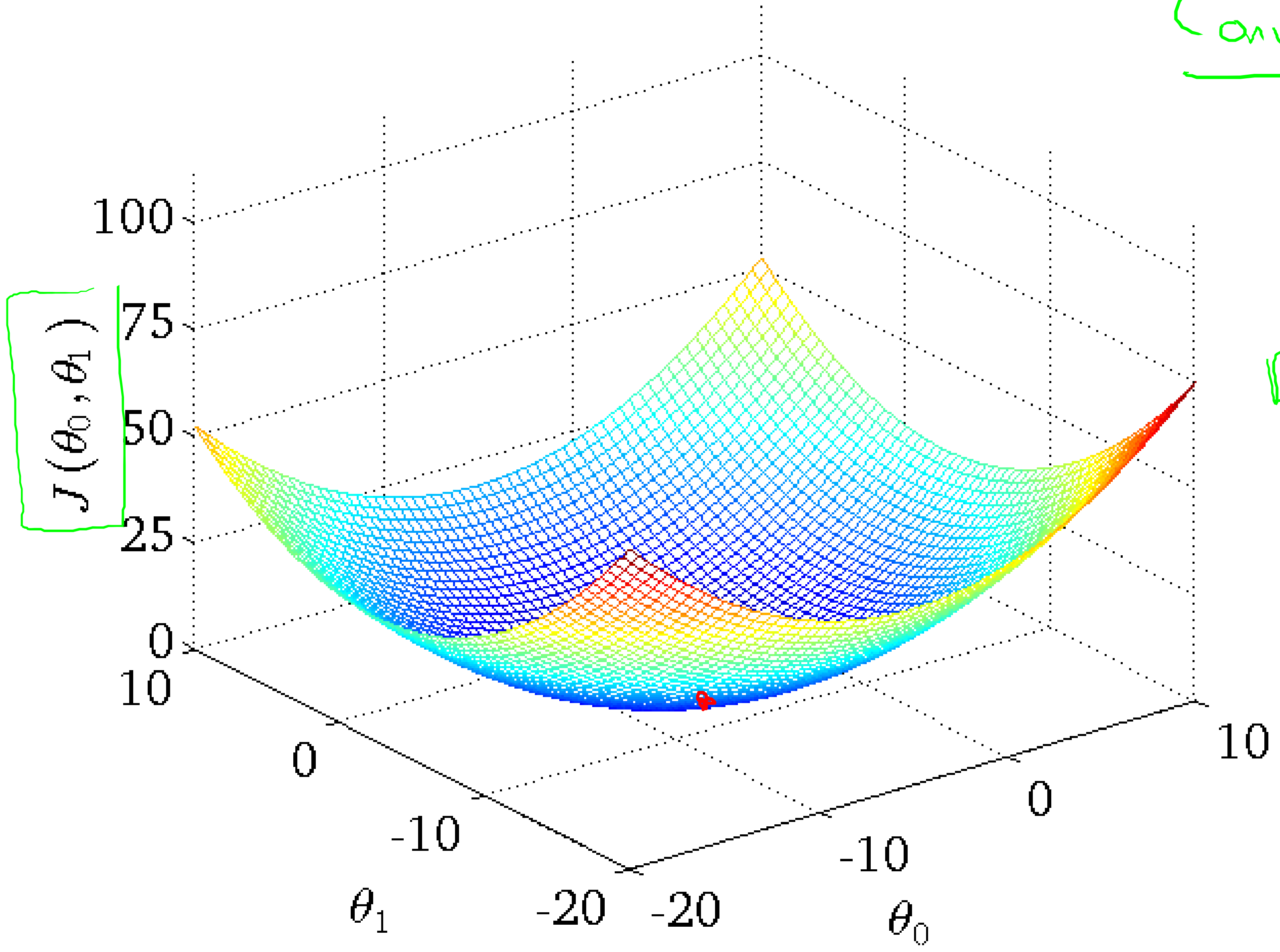
update θ_0 and θ_1 simultaneously



$J(\theta_0,\theta_1)$
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


3
2
1
0
-1
-2
-3
J(θ0,θ1)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
θ0
0
0.2
0.4
0.6
0.8
1
θ1

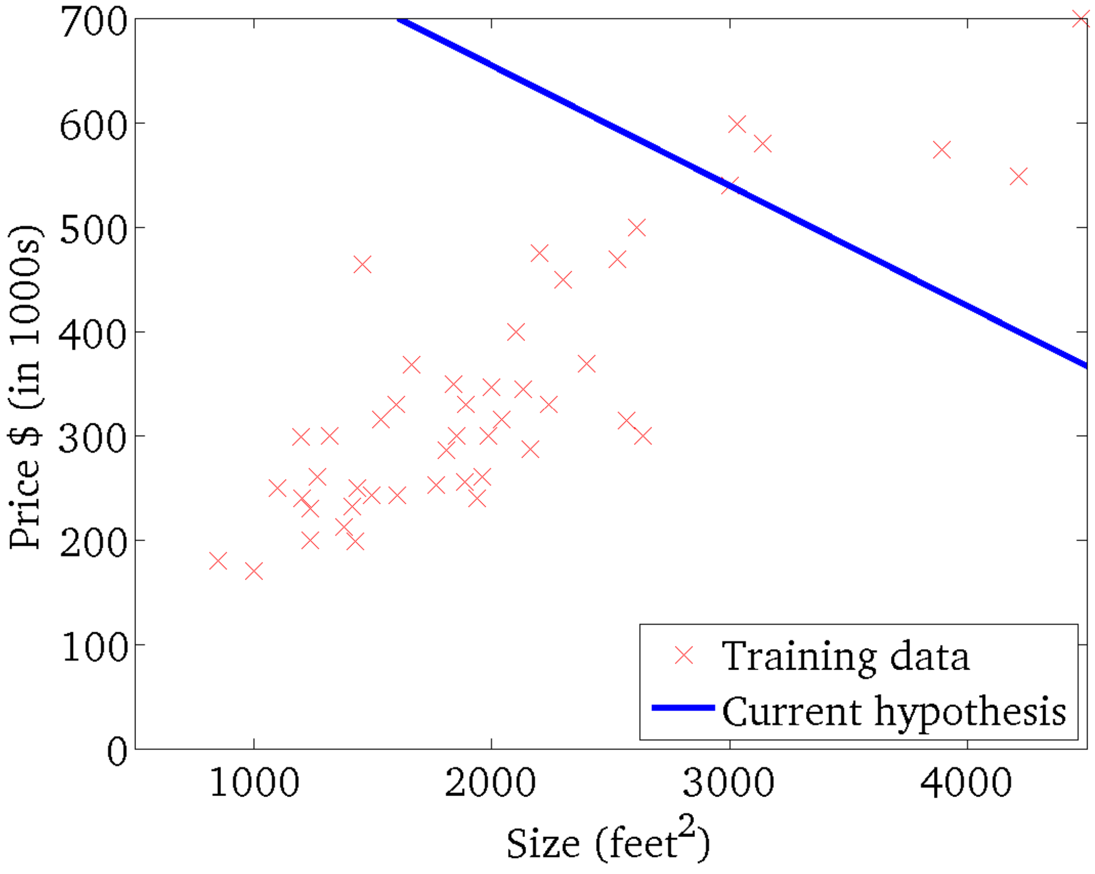
"Convex function"



Bowl-shaped

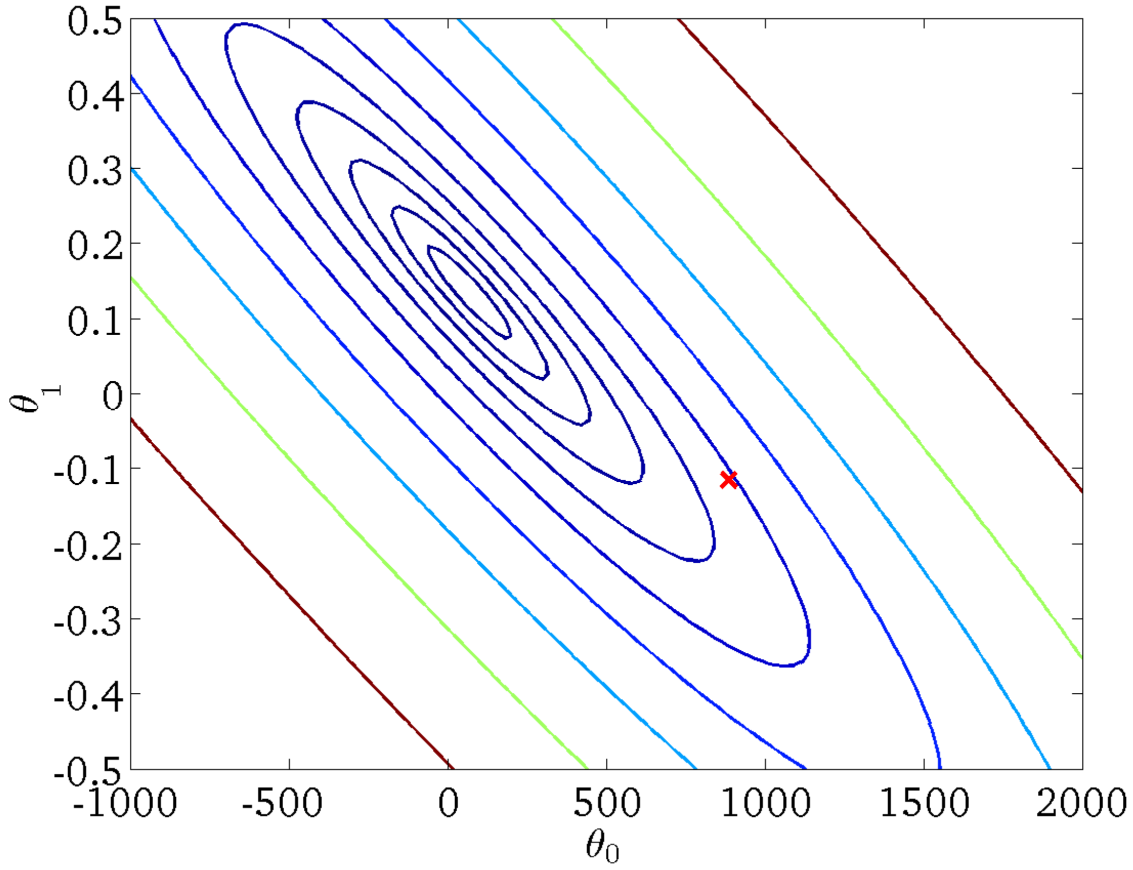
$h_\theta(x)$

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



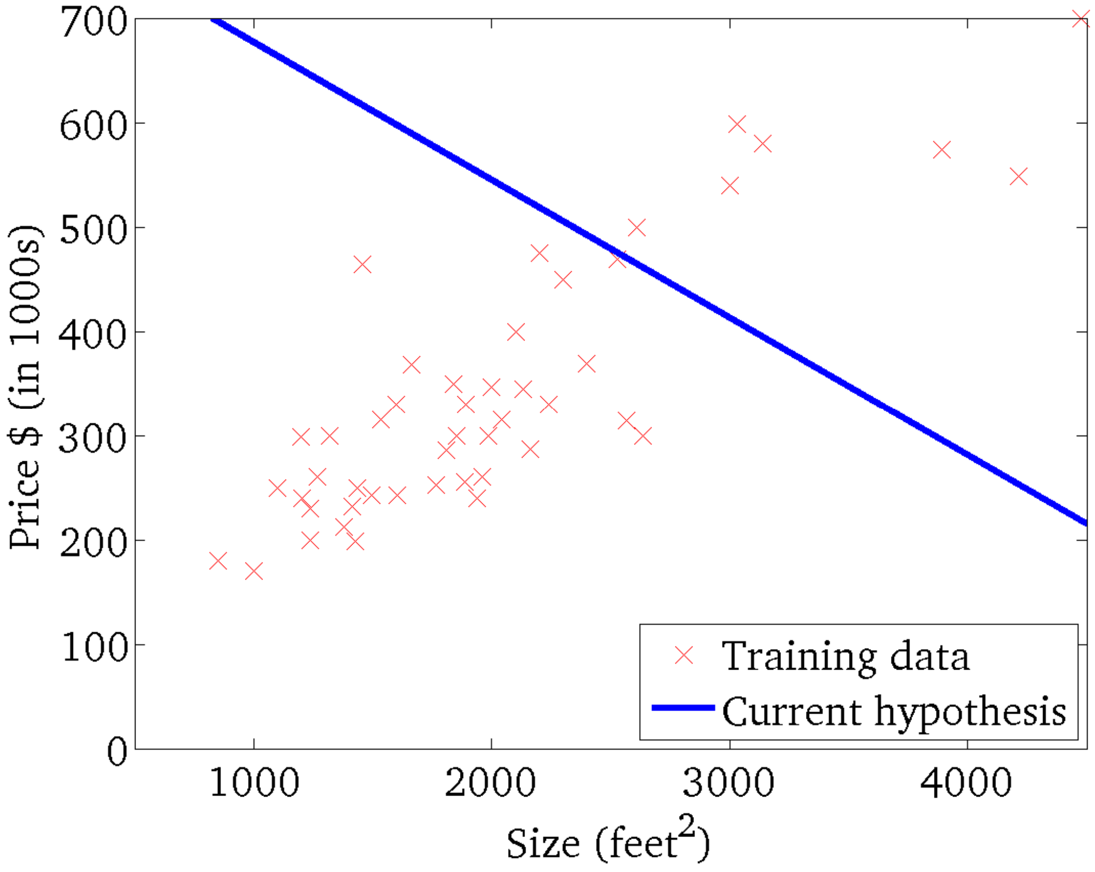
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



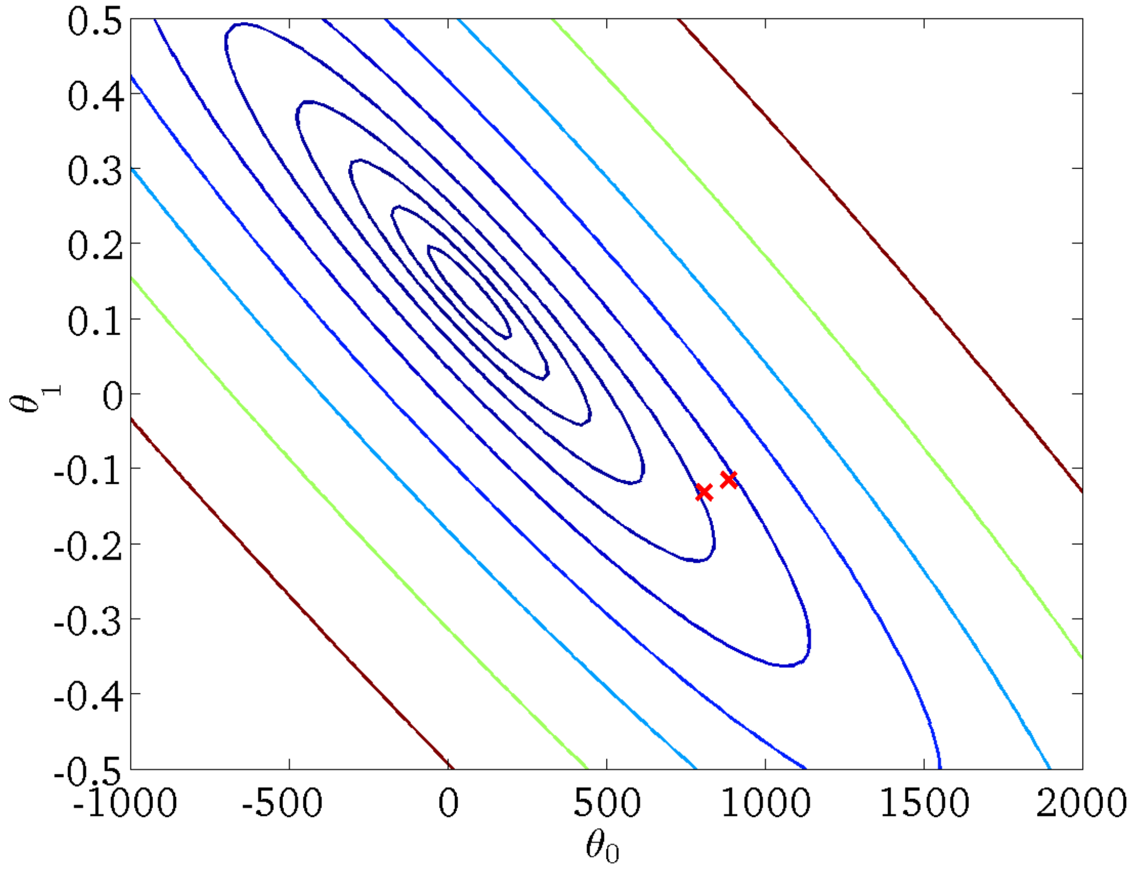
$$h_\theta(x)$$

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



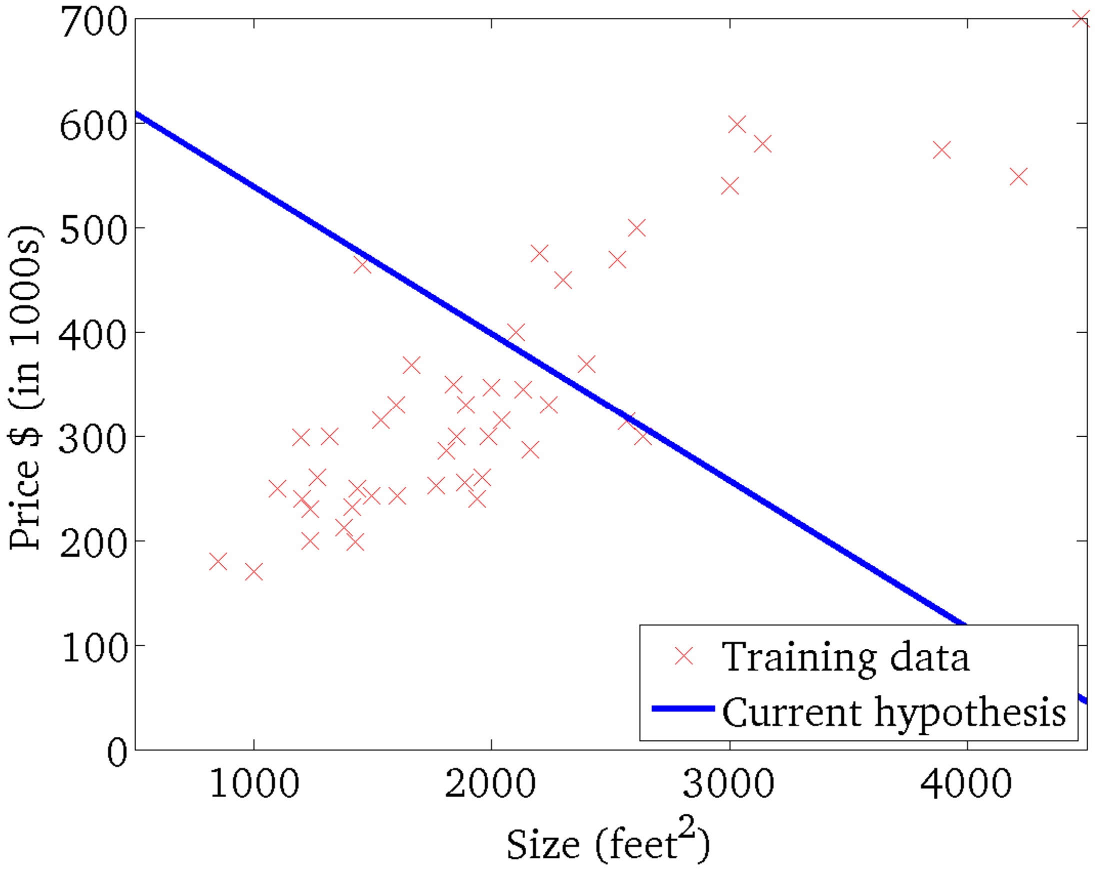
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



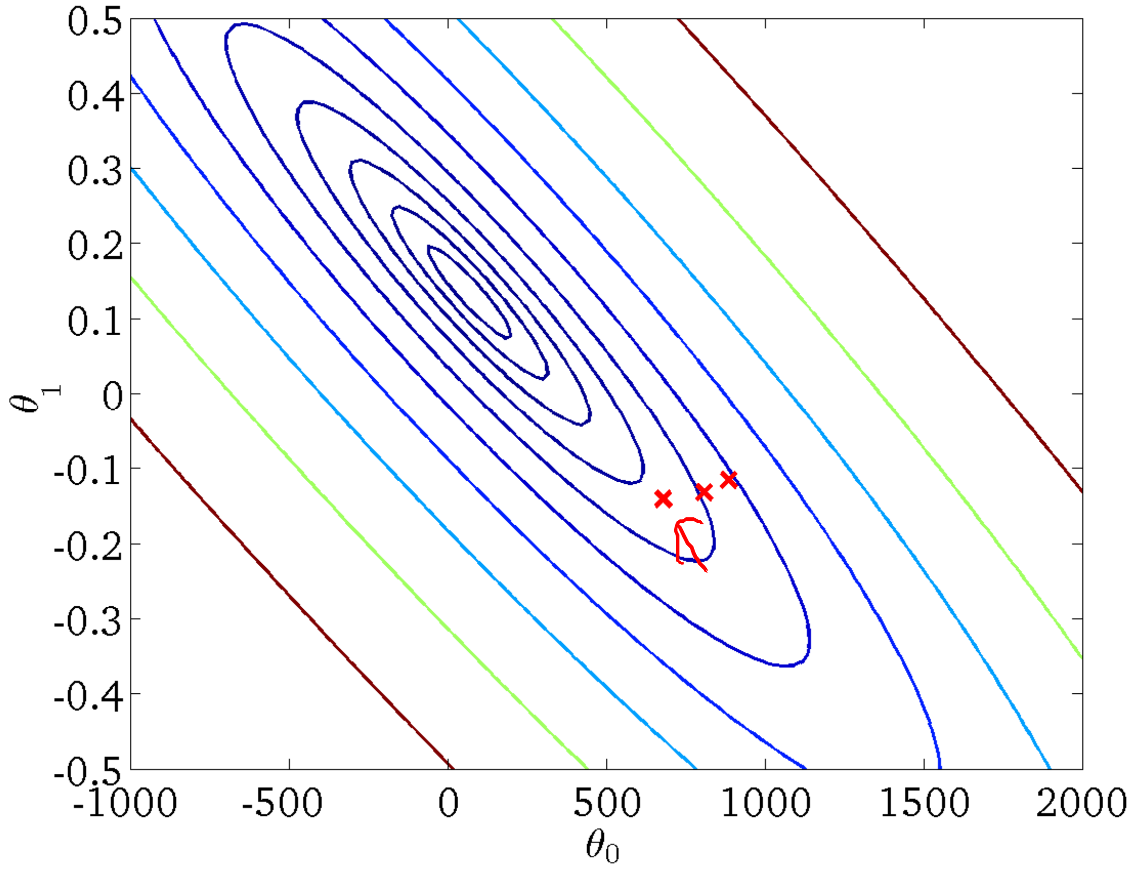
$h_\theta(x)$

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



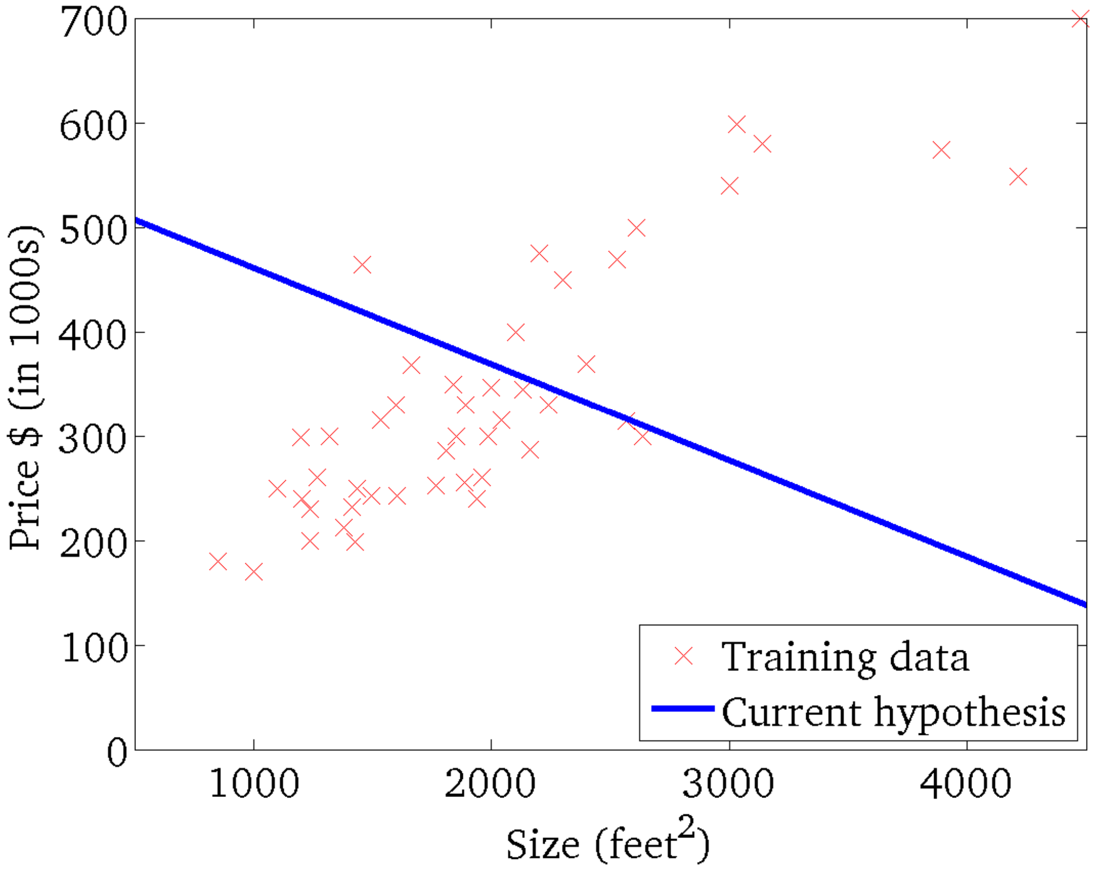
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



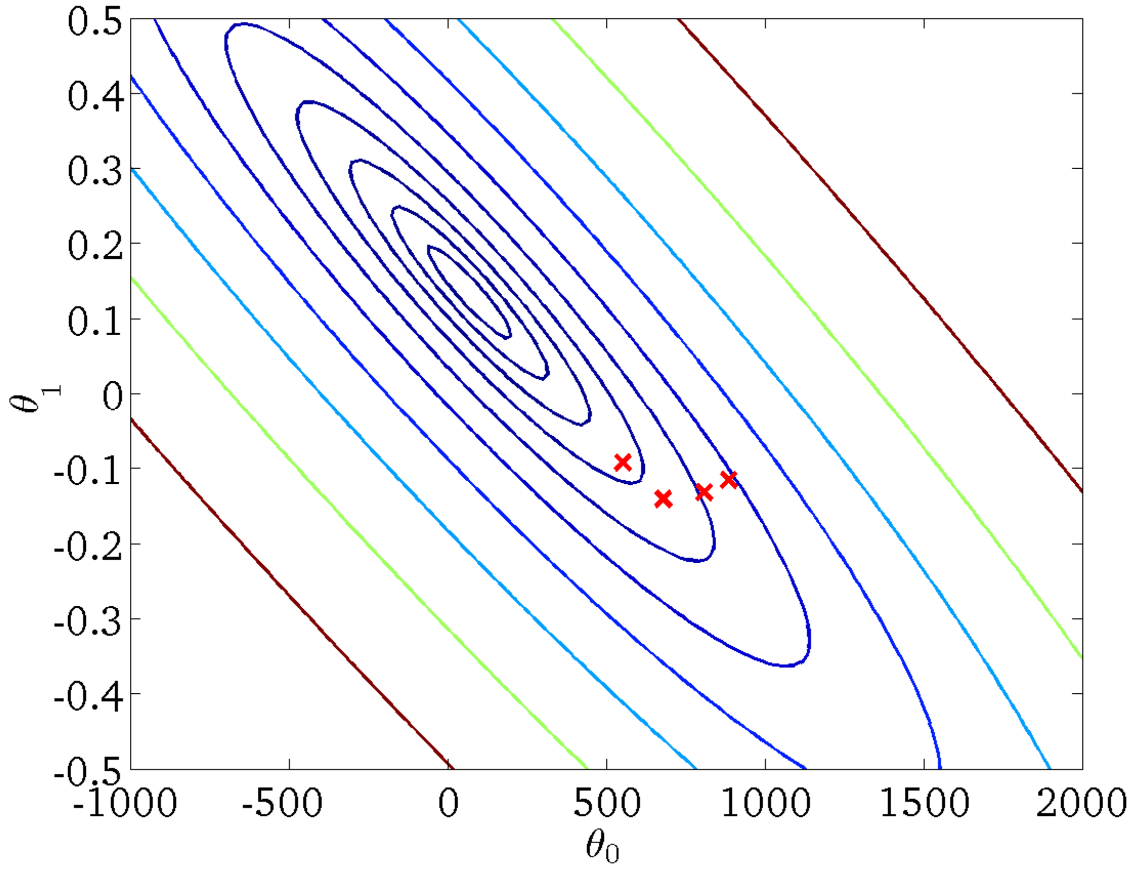
$$h_\theta(x)$$

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



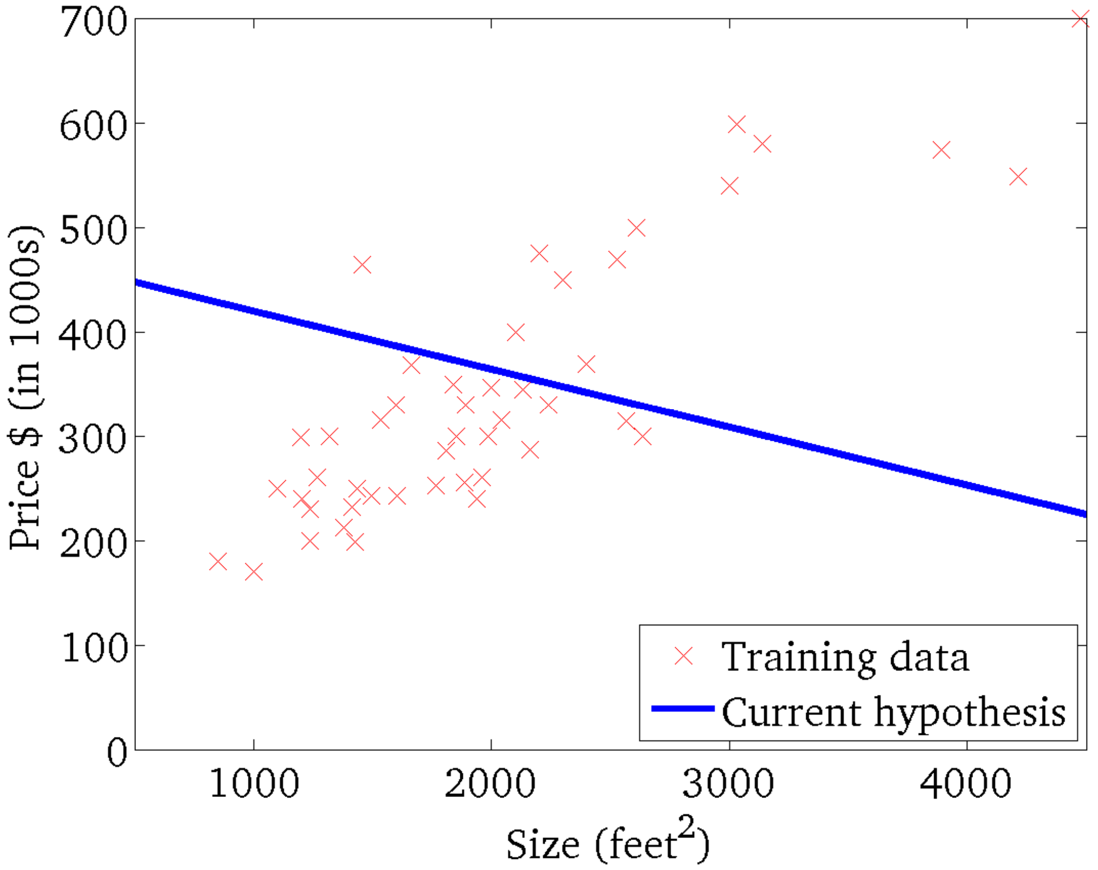
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



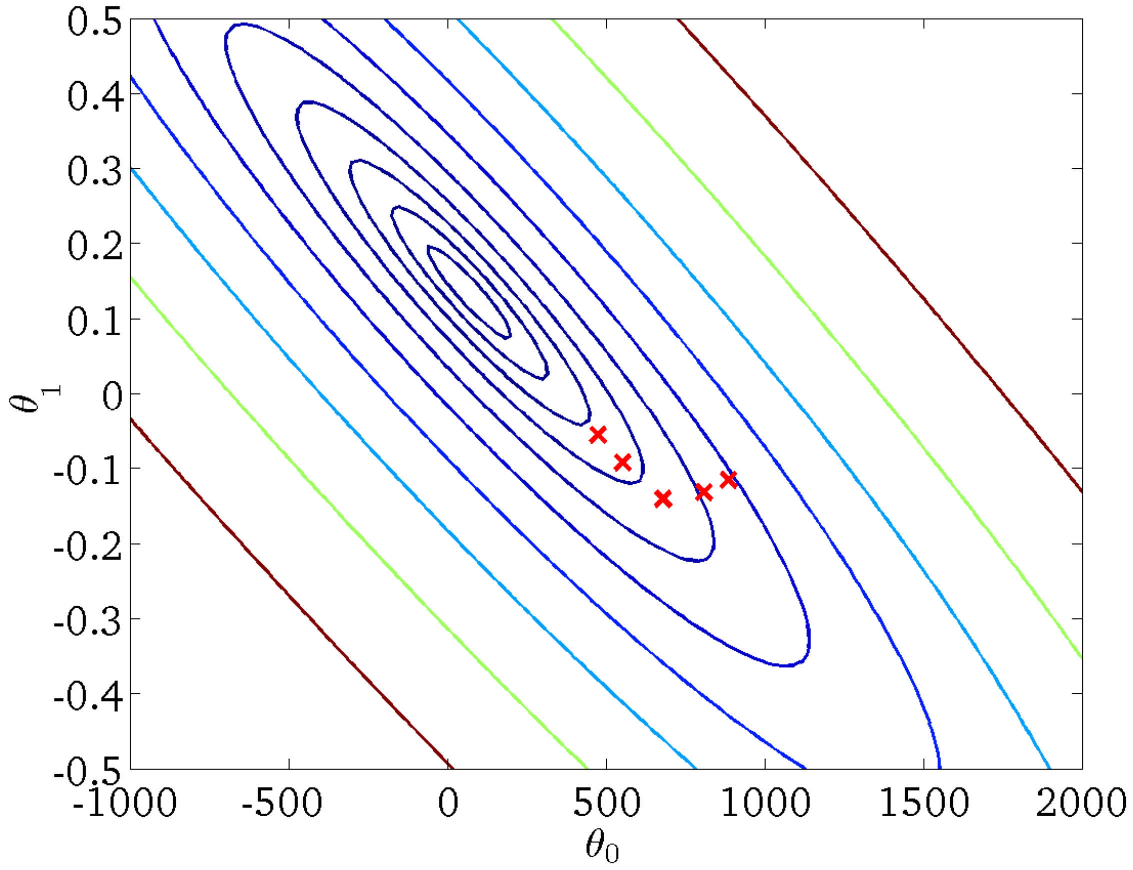
$h_\theta(x)$

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



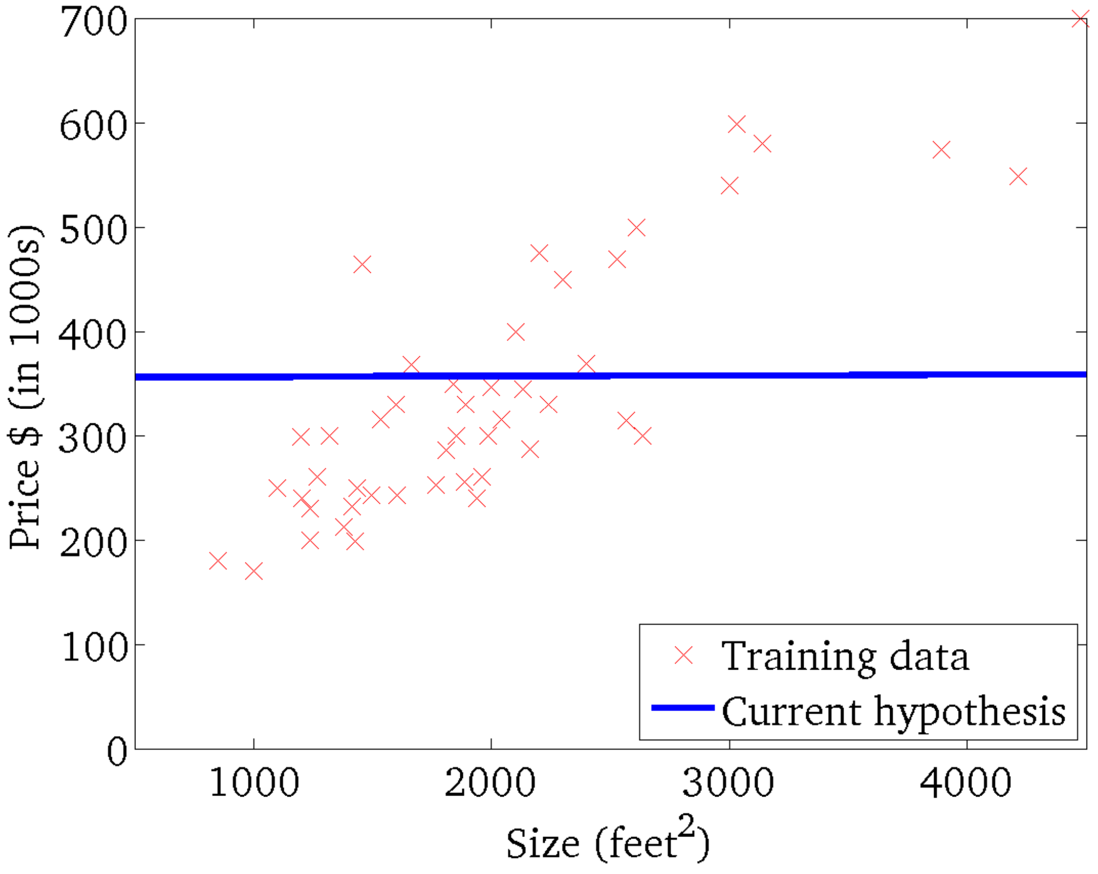
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



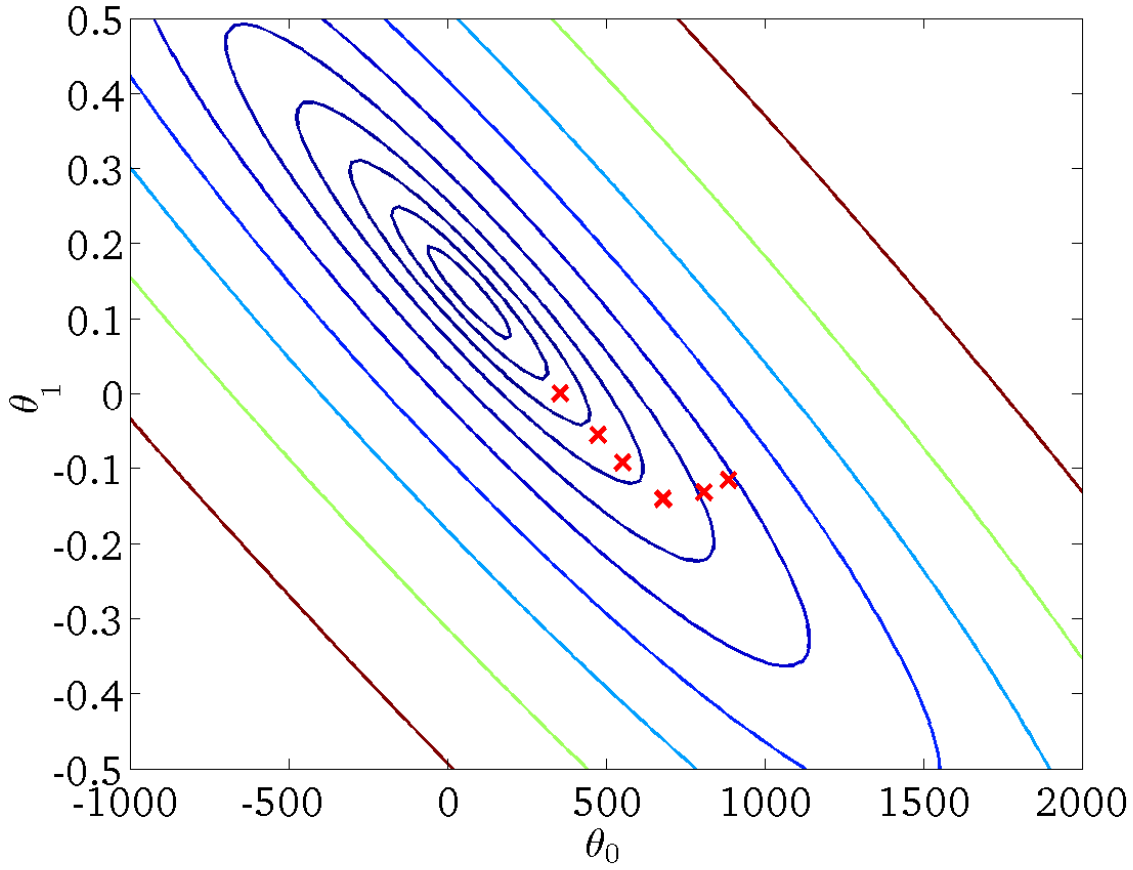
$$h_\theta(x)$$

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



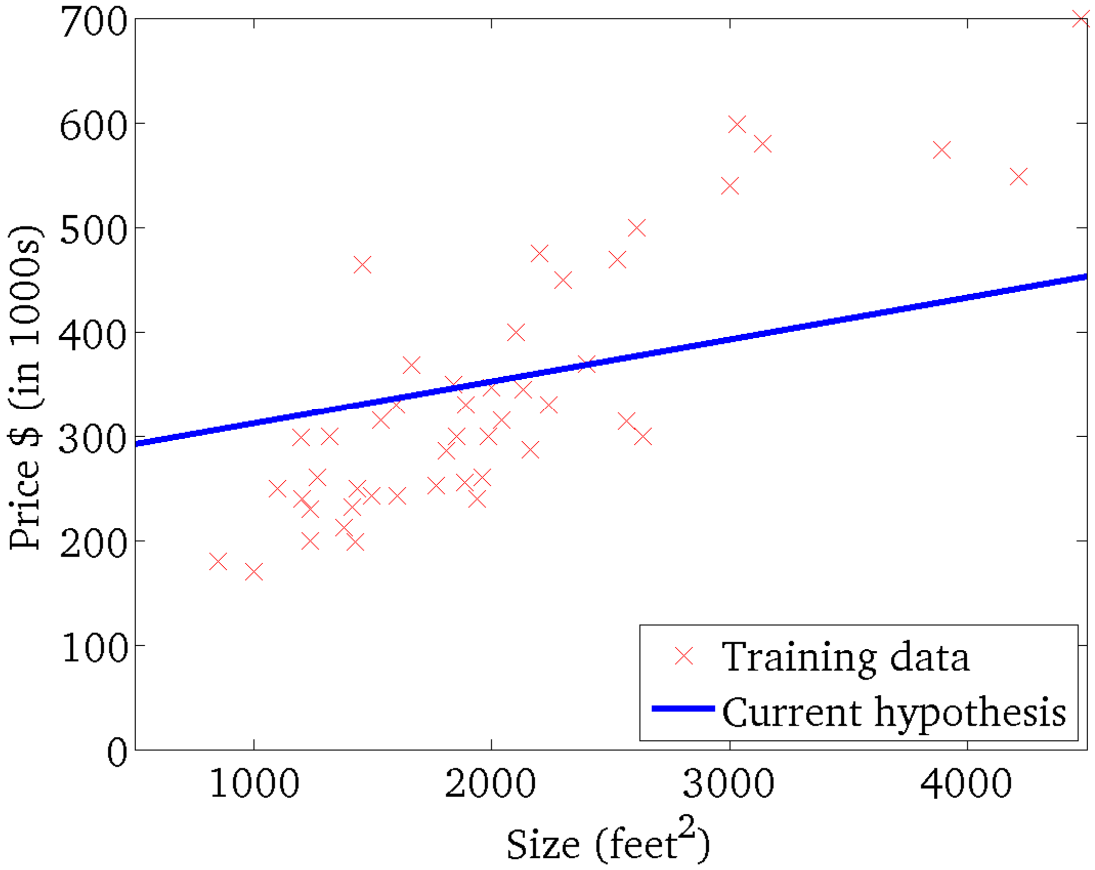
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



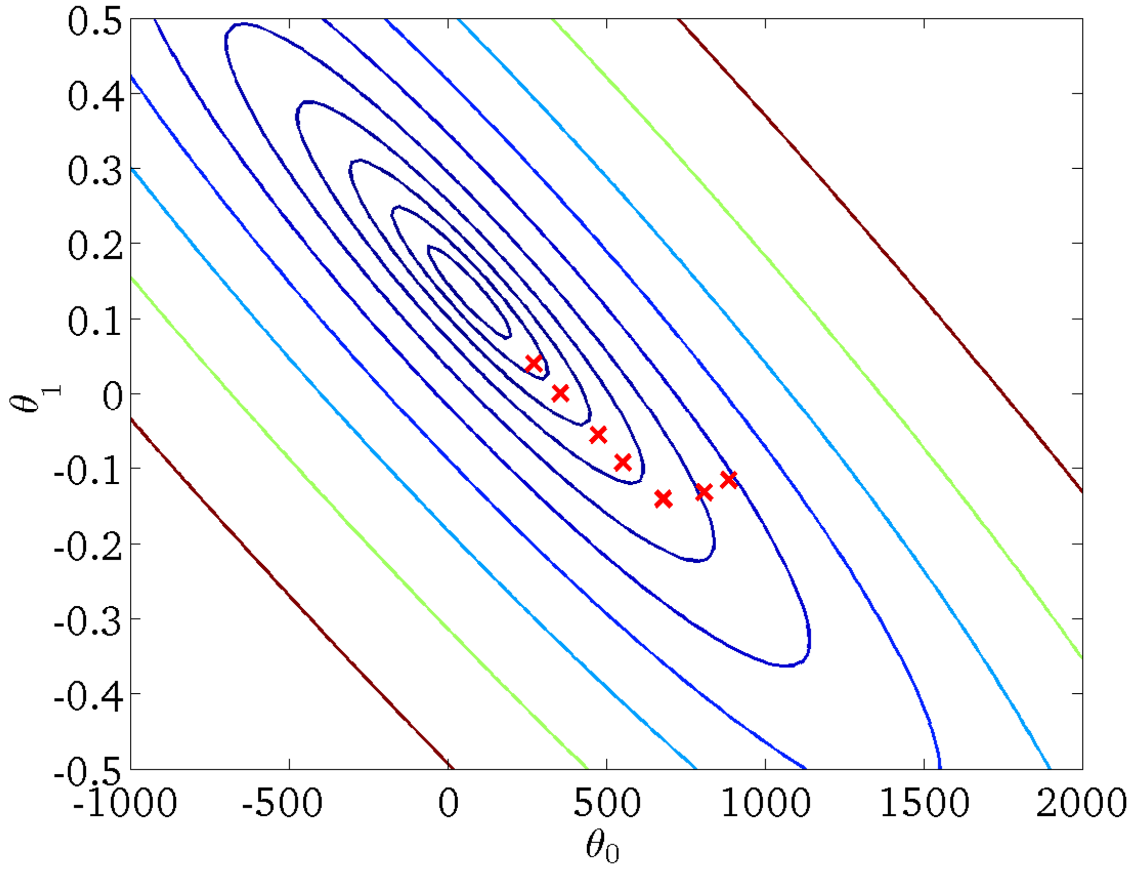
$h_\theta(x)$

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



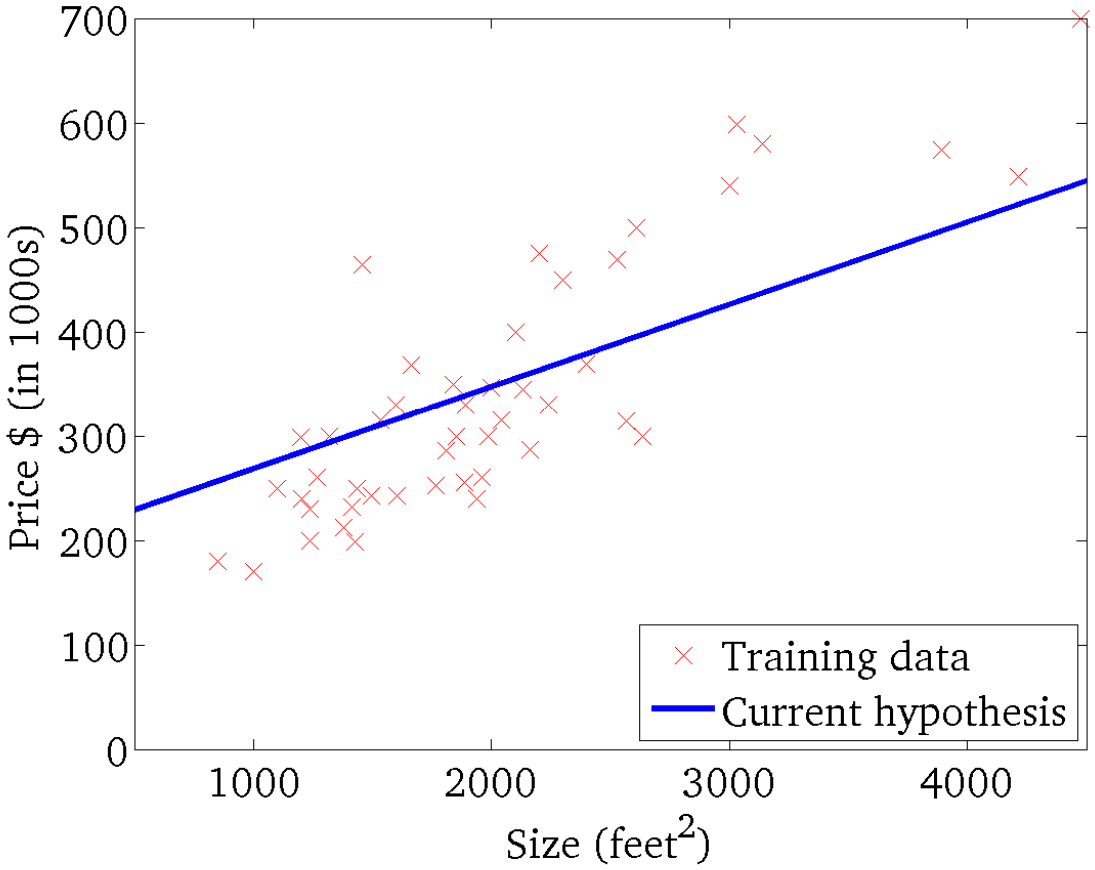
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



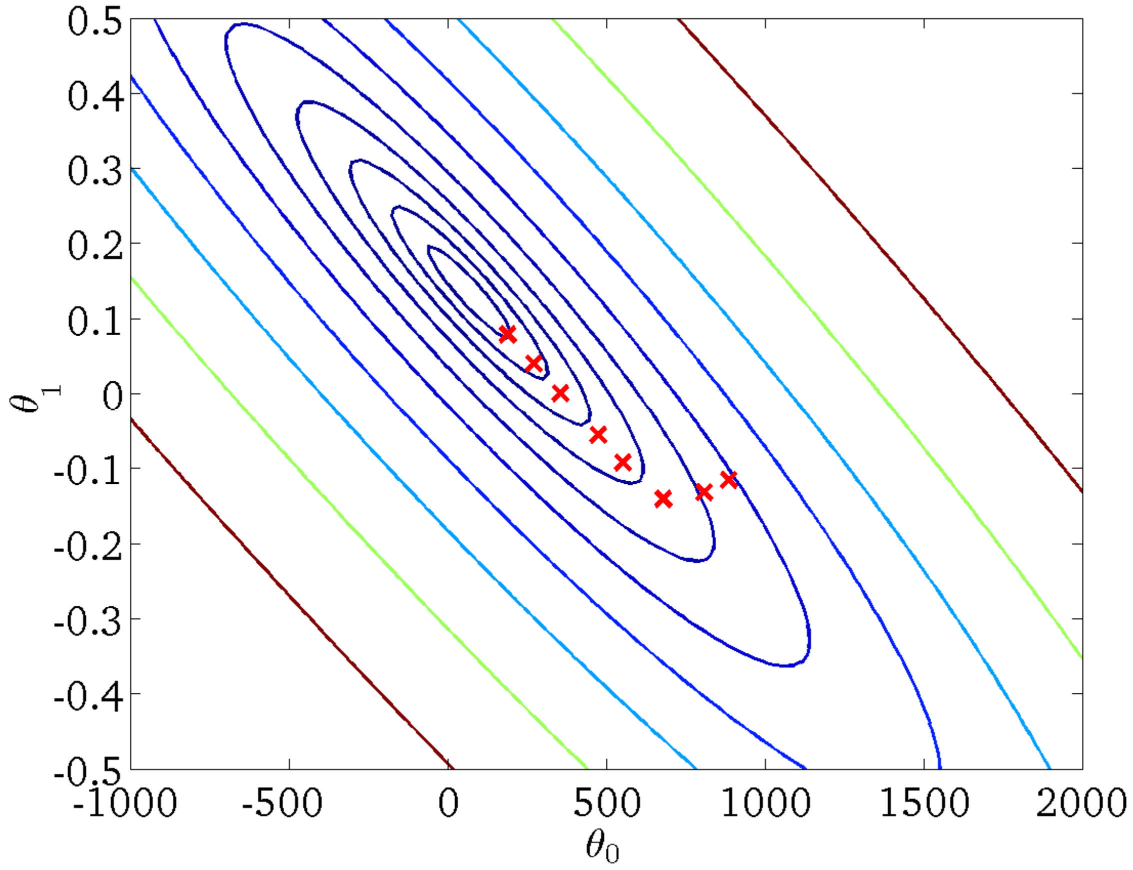
$h_\theta(x)$

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



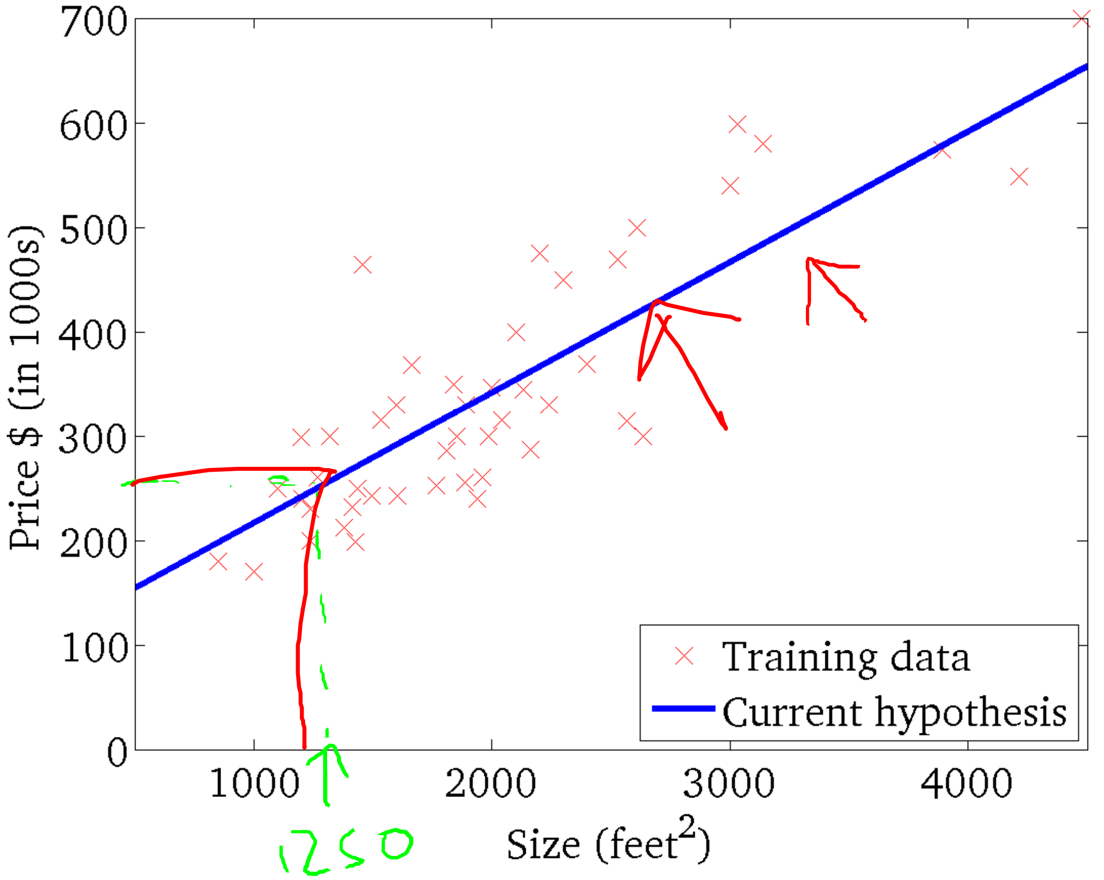
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



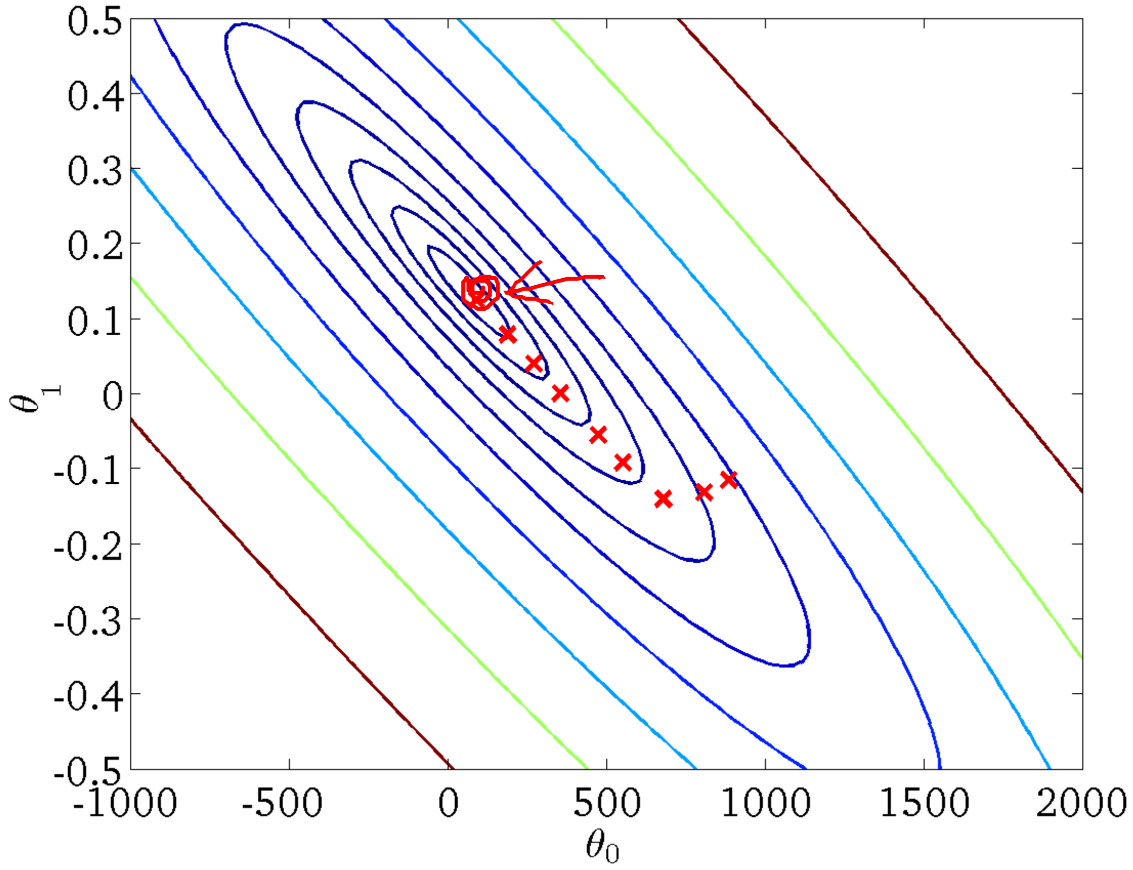
$h_\theta(x)$

(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)



“Batch” Gradient Descent

“Batch”: Each step of gradient descent uses all the training examples.

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

50

THANKS


Keep Learning
Keep Growing


Dr. Neeraj Gupta
Assistant Professor, Dept. of CEA
neeraj.gupta@gla.ac.in