Linear Regression

The input variable is denoted by x,
also called as the input feature

The output variable is denoted as y, also called the target variable

Training Set

Living Area of the House in sq. ft. (X)	Price of the House in dollars (Y)
1000	50000
400	25000
350	24000
2000	90000

Training Example

$$(x^{(2)}, y^{(2)}) = (400, 2500)$$

A Training Set is a list of Training Examples

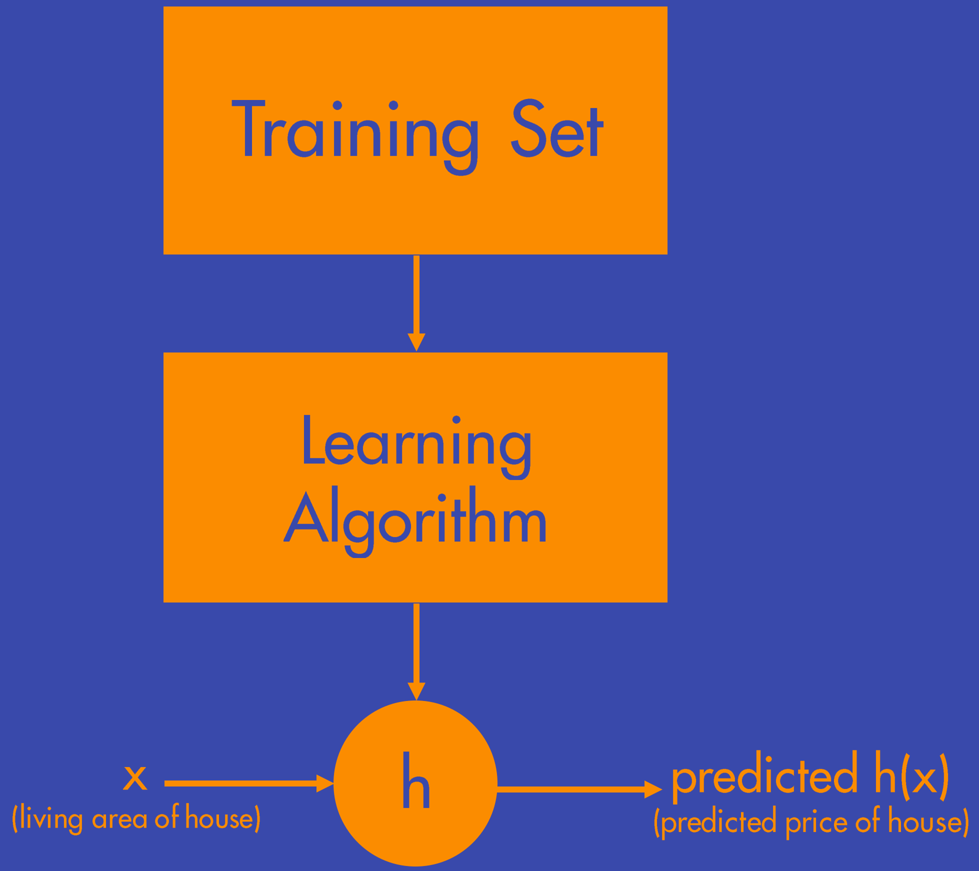
Hypothesis

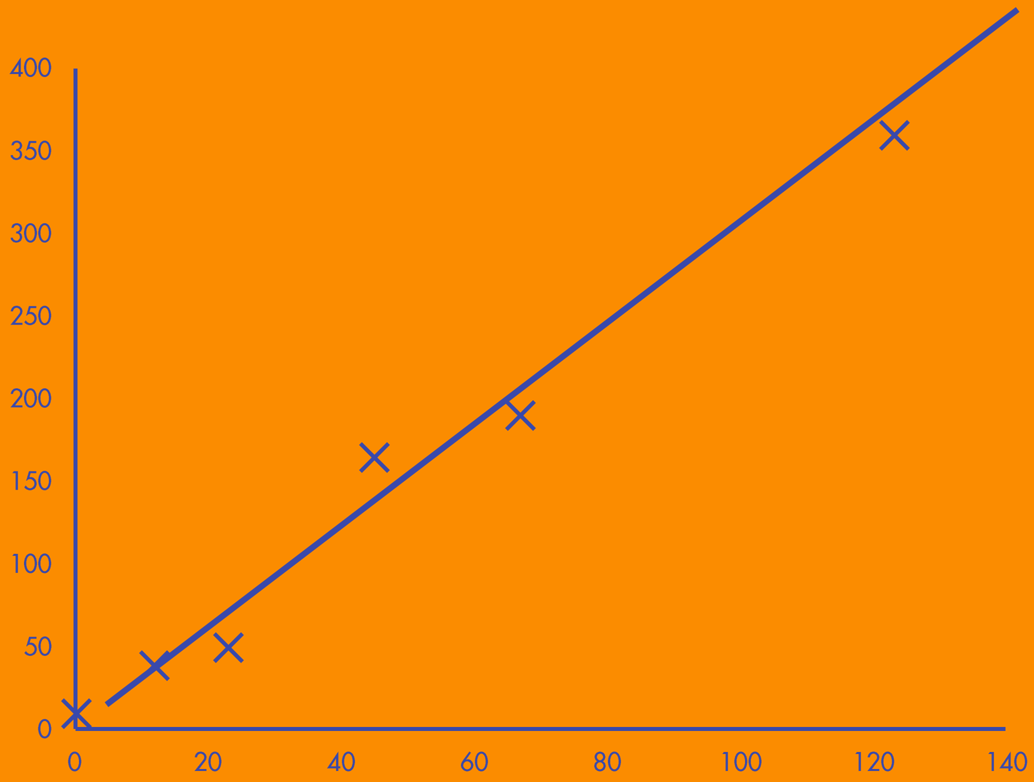
h is a function that maps the feature x into a predicted output $h(x)$.

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

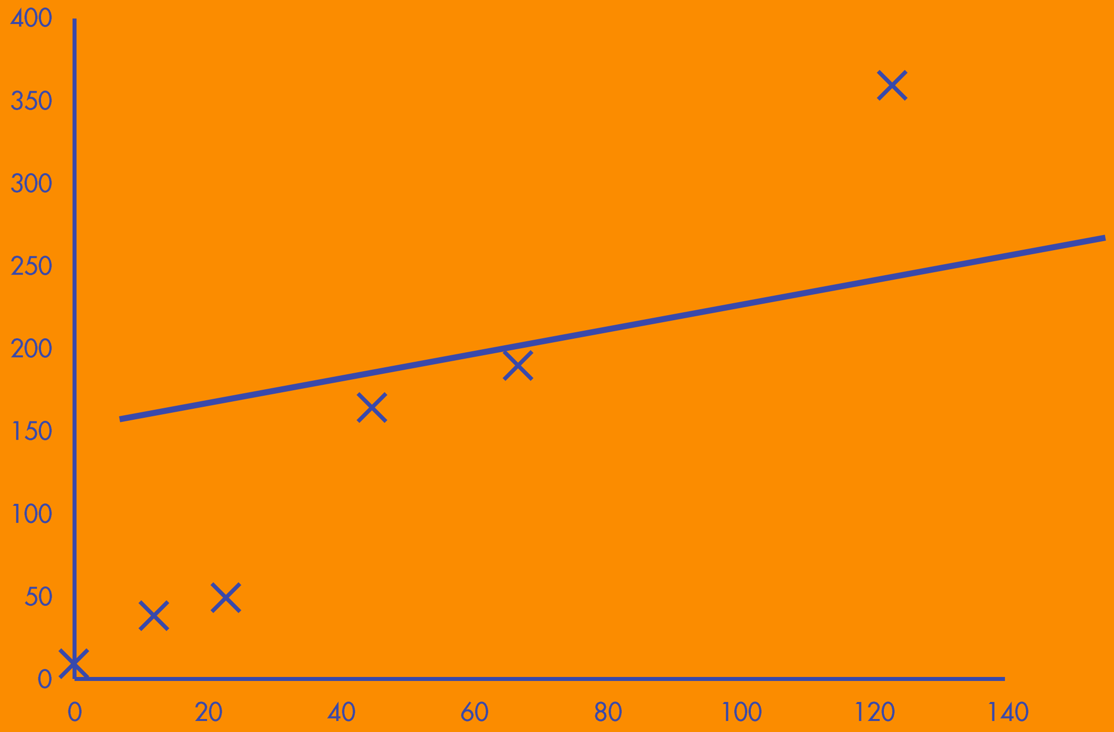
The algorithm is trying to learn the best h for the training set.

Model Representation




400
350
300
250
200
150
100
50
0
0 20 40 60 80 100 120 140

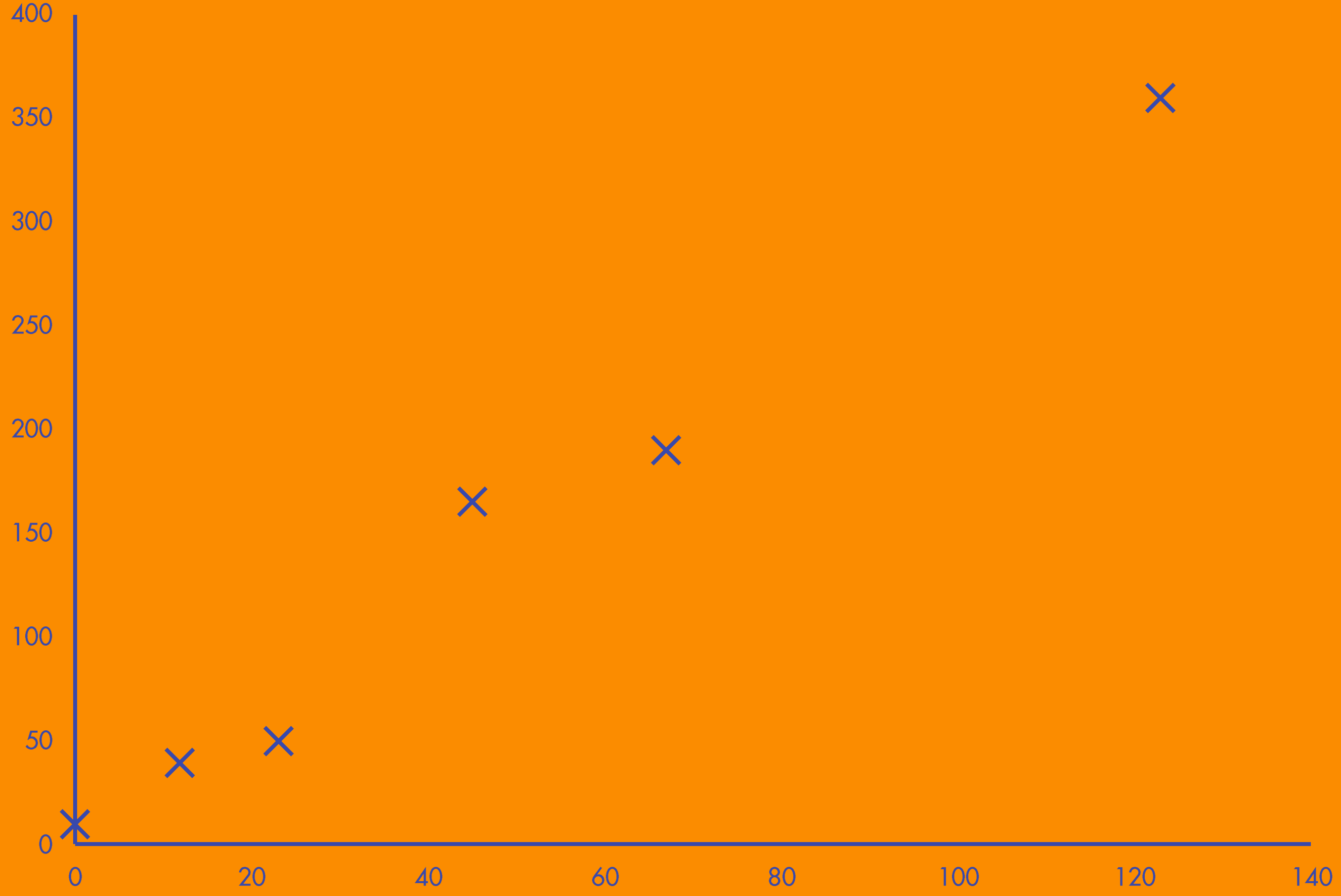
h_1


400
350
300
250
200
150
100
50
0
0 20 40 60 80 100 120 140

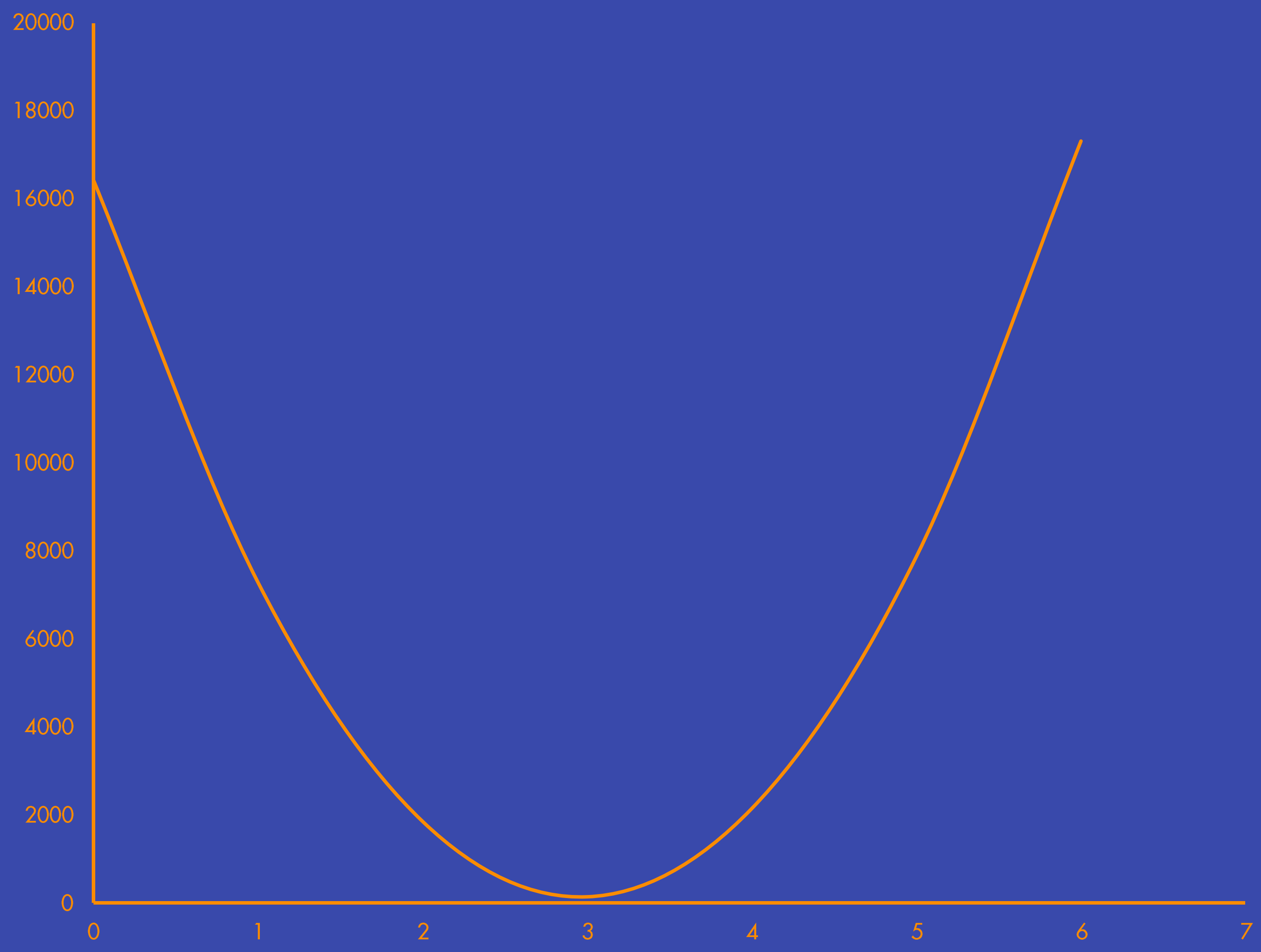
h_2

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

The smaller the value of $J(\theta_0, \theta_1)$, the better the fit of the hypothesis to the training set



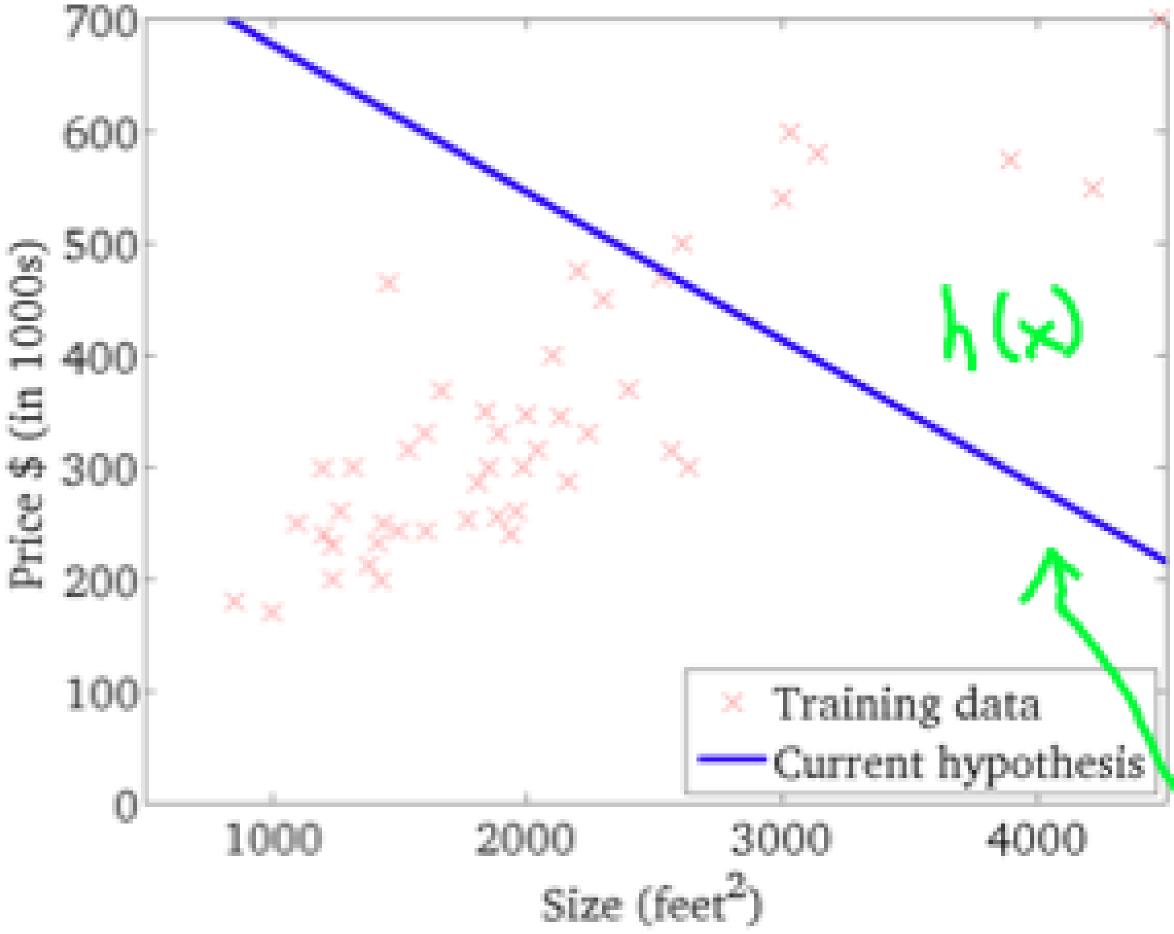
x	y
12	39
23	50
45	165
0	10
123	360
67	190



θ_1	$J(\theta_1)$
0	16420.5
1	7271.3
2	1841.5
3	131
4	2139.8
5	7868
6	17315

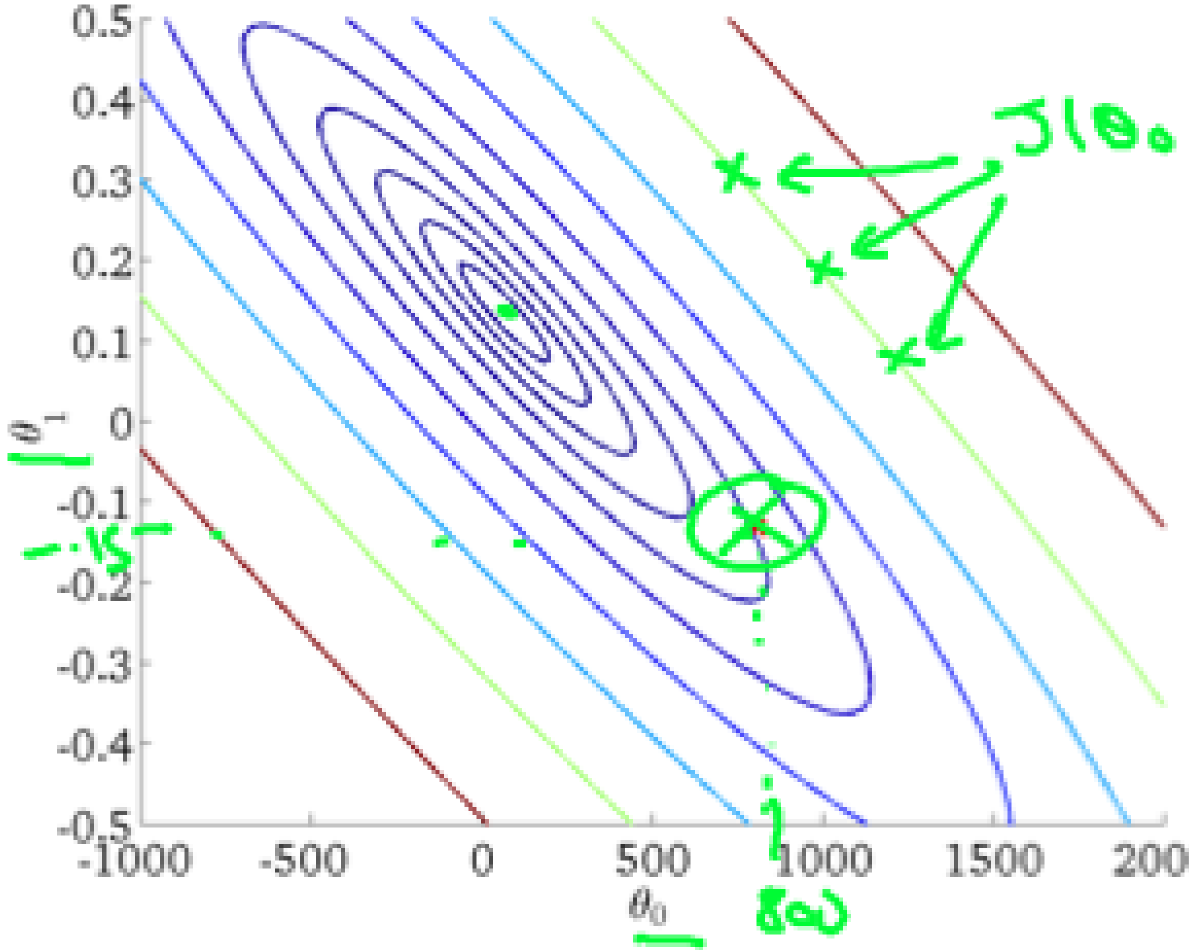
$h_\theta(x)$

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



$J(\theta_0, \theta_1)$

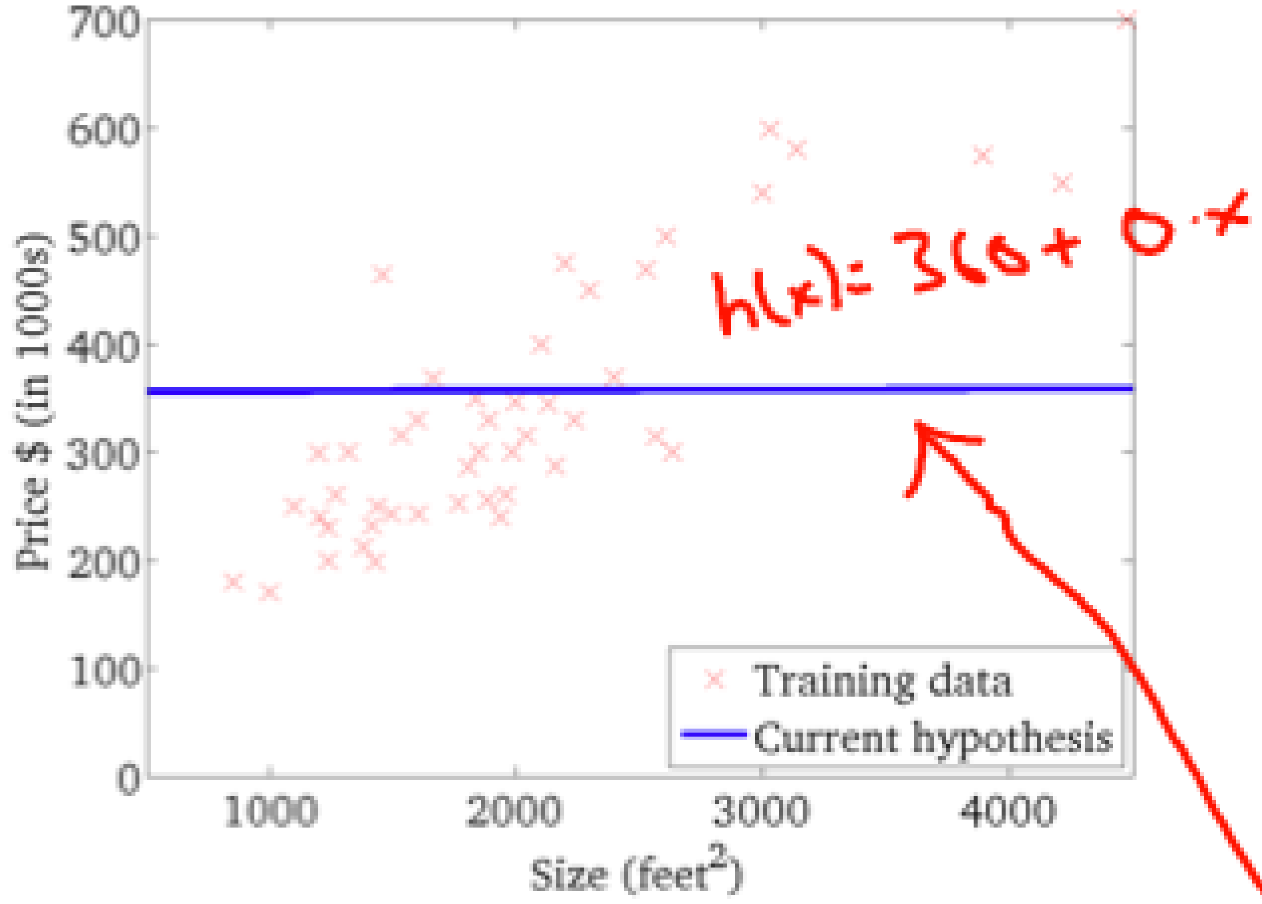
(function of the parameters θ_0, θ_1)



θ_0, θ_1

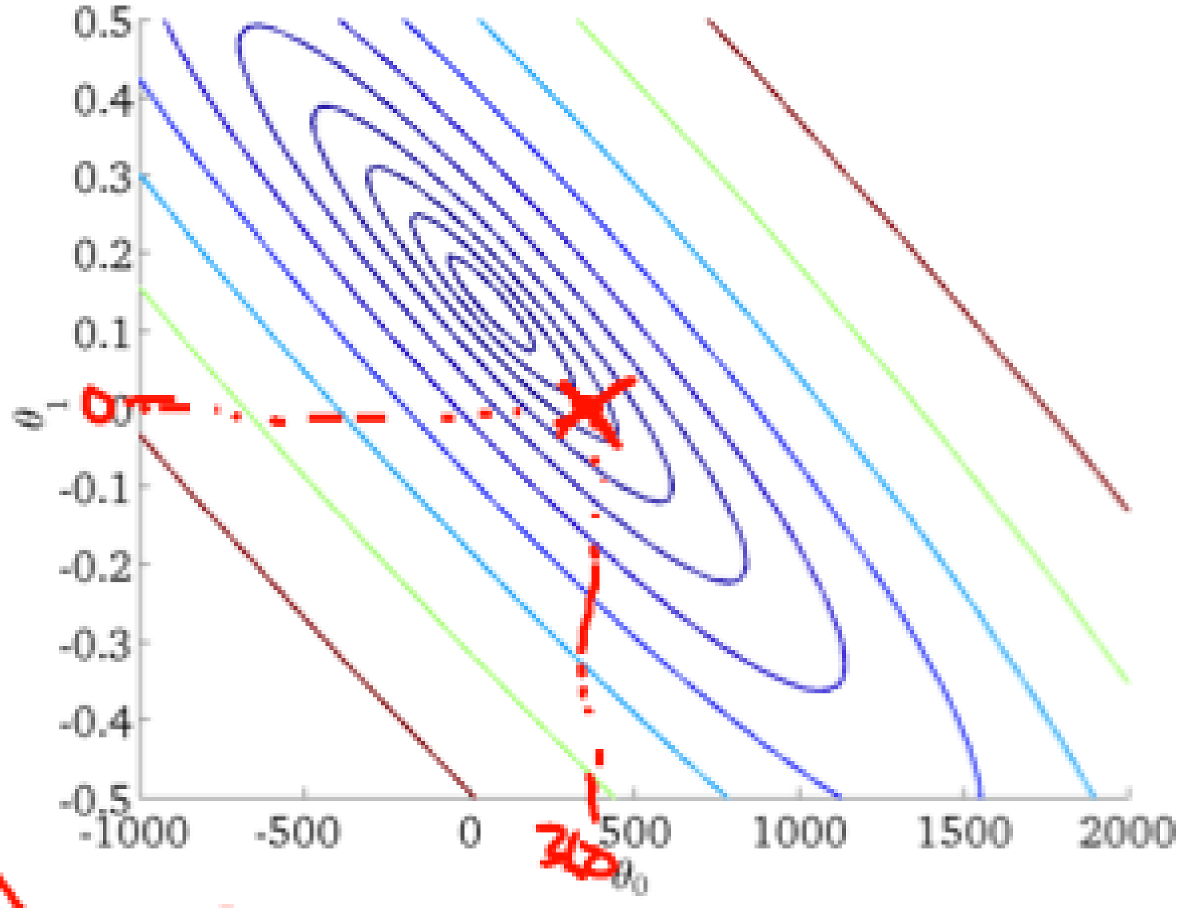
$h_\theta(x)$

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



$J(\theta_0, \theta_1)$

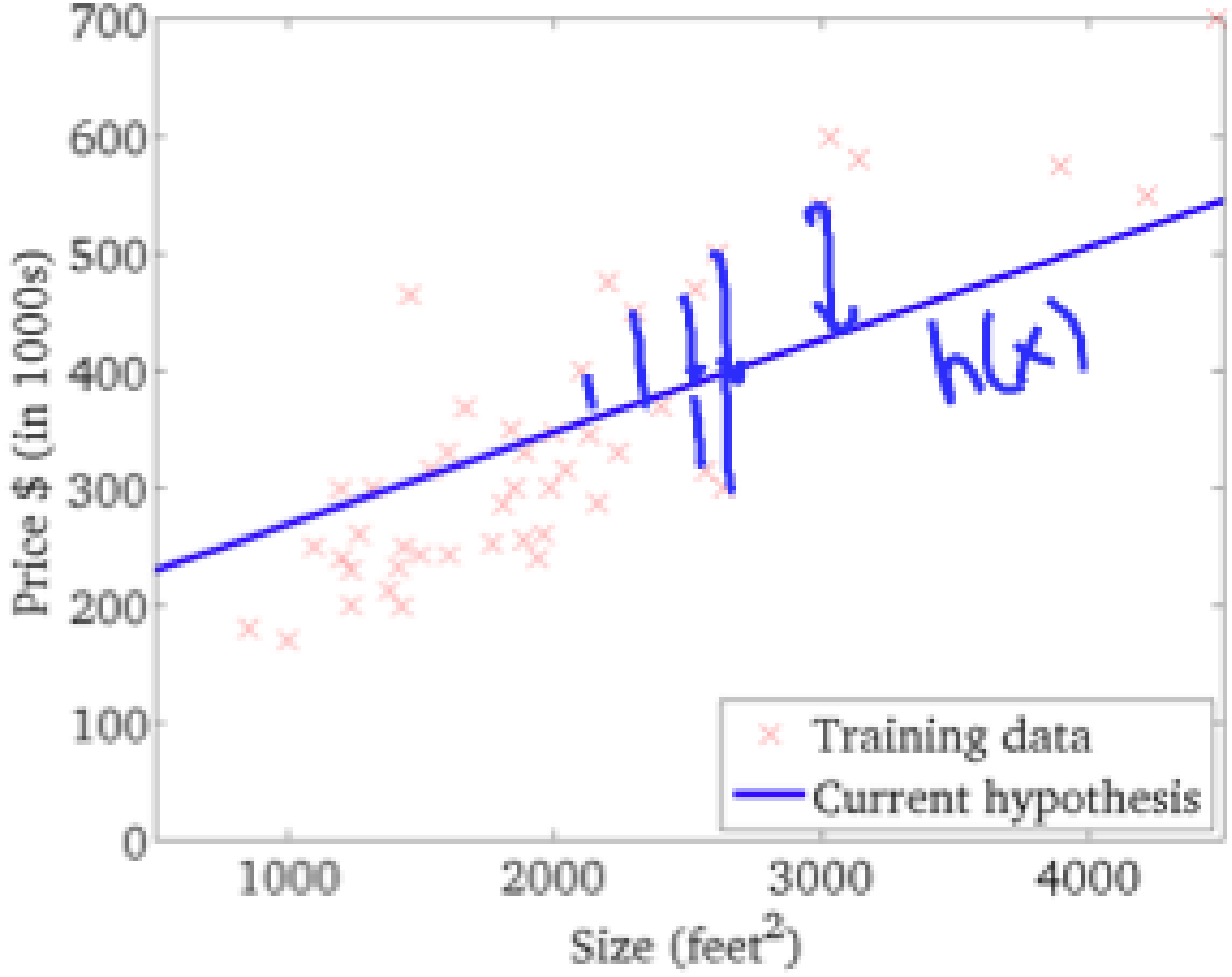
(function of the parameters θ_0, θ_1)



$\theta_0 = 360$
$\theta_1 = 0$

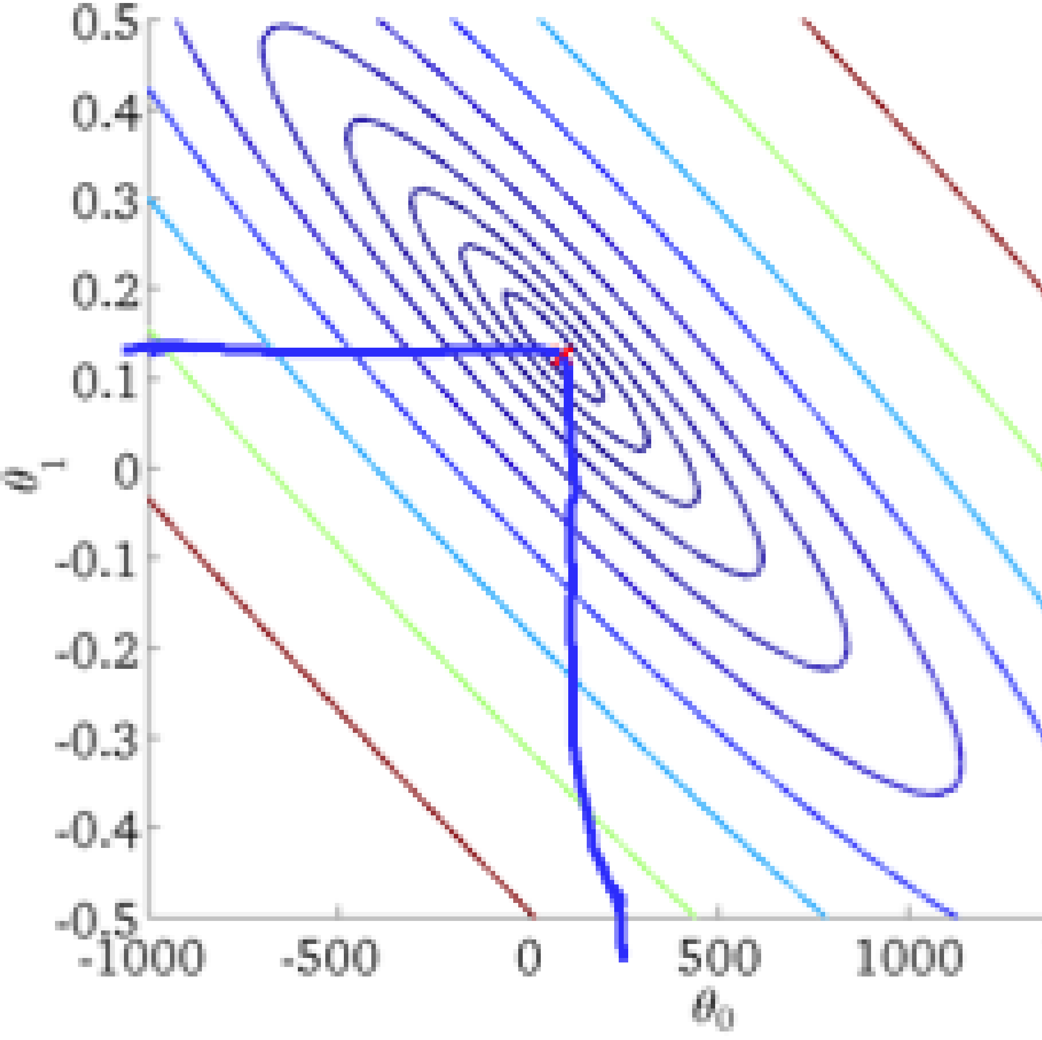
$h_\theta(x)$

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



$J(\theta_0, \theta_1)$

(function of the paramete



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

Repeat until convergence:{

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h\left(x^{(i)}\right) - y^{(i)}\right) x_0^{(i)}$$

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

}

x_1	x_2	y
12	2	39
23	-5	50
45	44	165
0	12	10
123	-12	360
67	-20	190

More Notations

$x_j^{(i)}$ = value of the feature j in the i^{th} training example
$x^{(i)}$ = the input features of the i^{th} training example
m = number of training examples
n = number of features

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

$$h(x) = [\theta_0 \;\; \theta_1 \;\; \dots \theta_n] \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \theta^T x$$

$$H = X\theta = \begin{bmatrix} x_0^1 & x_1^1 & \cdots & x_n^1 \\ x_0^2 & x_1^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_0^m & x_1^m & \cdots & x_n^m \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} = \begin{bmatrix} h(x^{(1)}) \\ h(x^{(2)}) \\ \vdots \\ h(x^{(m)}) \end{bmatrix}$$

$$H - y = \begin{bmatrix} h(x^{(1)}) \\ h(x^{(2)}) \\ \vdots \\ h(x^{(m)}) \end{bmatrix} - \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

Repeat until convergence:{

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h\left(x^{(i)}\right) - y^{(i)} \right) x_0^{(i)}$$

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

$$\vdots$$

$$\theta_n := \theta_n - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h\left(x^{(i)}\right) - y^{(i)} \right) x_n^{(i)}$$

}

Repeat until convergence $(for\ j\ in\ [0, n])$: {

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

}