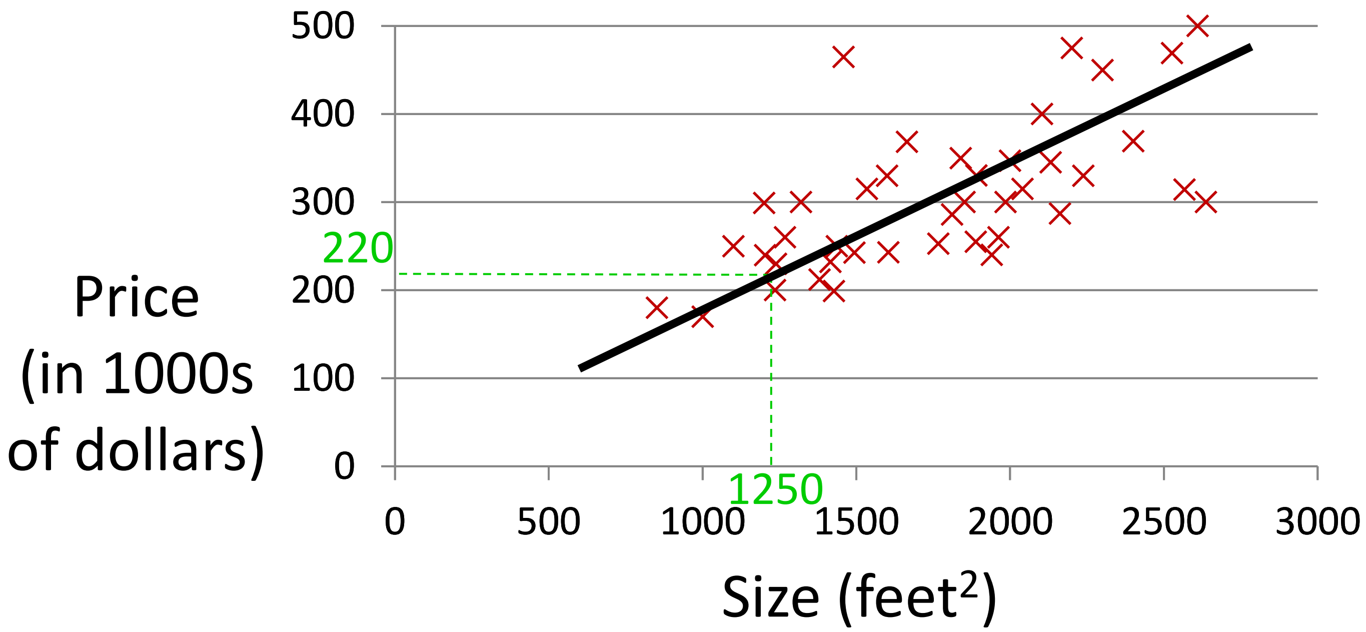
Linear Regression

By Vipul Goyal

House Price from Size



Notation

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

m

Notation:

m = Number of training examples
x's = "input" variable (features)
y's = "output" variable ("target" variable)

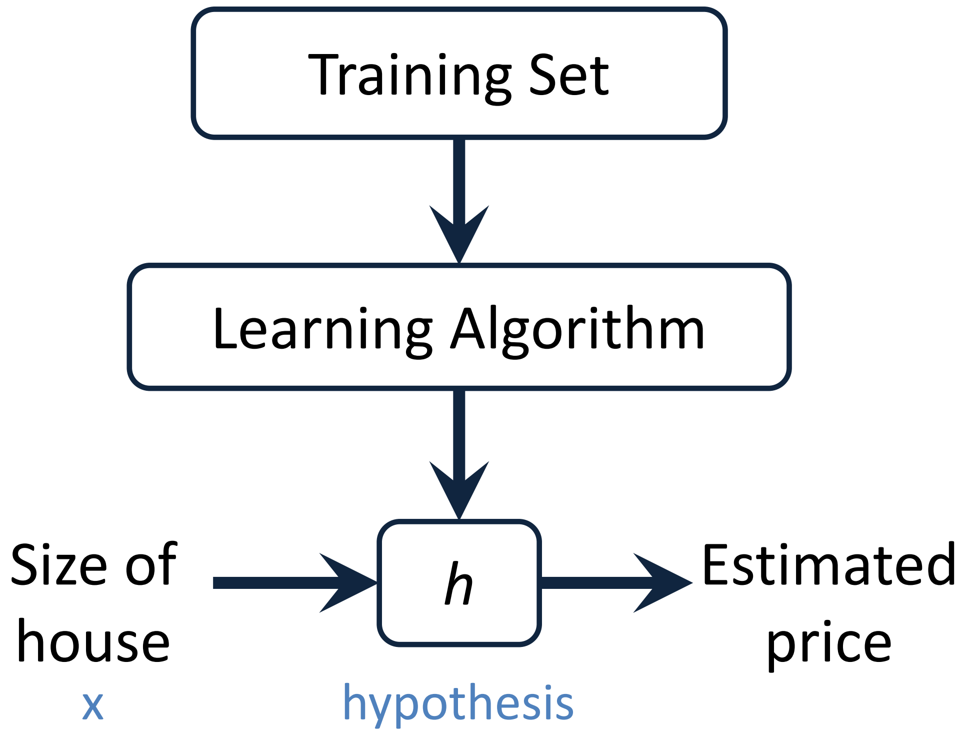
(x, y) – one training example
($x^{(i)}$, $y^{(i)}$) – i-th training example

$x^{(1)} = 2104$
$x^{(2)} = 1416$
$y^{(1)} = 460$

Hypothesis

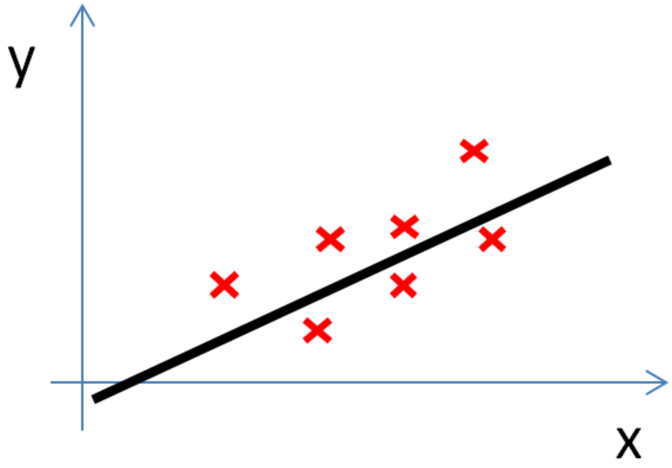
- In machine learning, many times the goal is simply to compute a "hypothesis" denoted as h
- This hypothesis h can take as input x (feature) and produce output y (target). That is $h(x) = y$
- For example, if $h(2100) = 457$, we know the predicted price of a house of 2100 sq feet is \$457k
- How can we compute h?

Hypothesis



h computed using the training dataset

Representing *h*?



Maybe *h* is a straight line? Linear Regression!

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

Later: *h* is more complex

Computing h?

Training Set

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

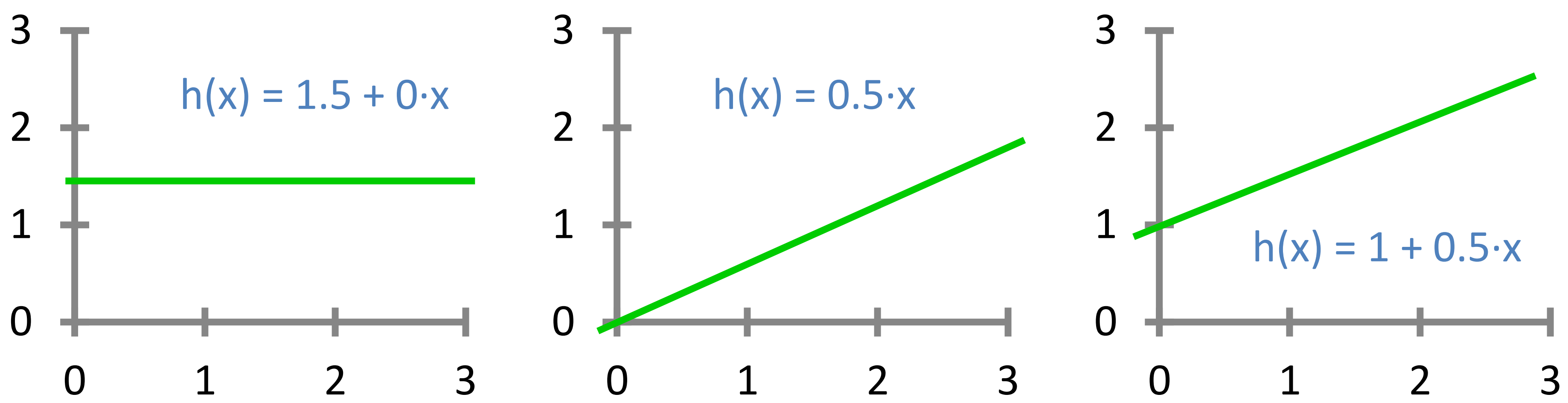
Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

$\theta_i' s$: coefficients or parameters

How to choose $\theta_i' s$ given training dataset?

Examples of Coefficients and h

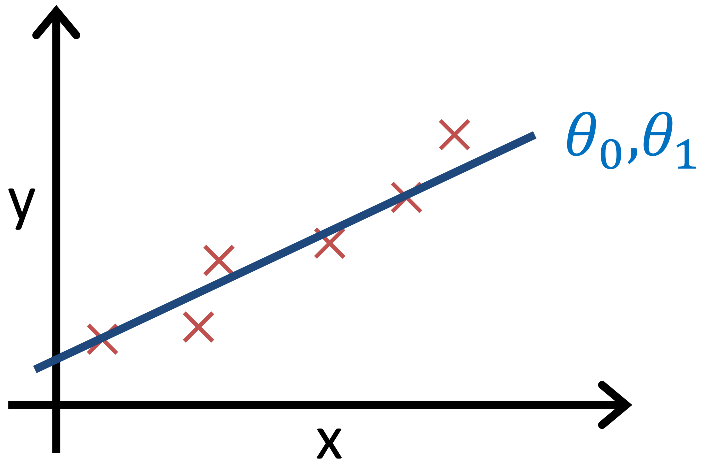
$$h_\theta(x) = \theta_0 + \theta_1 x$$



$\theta_0 = 1.5;\ \theta_1 = 0$ $\theta_0 = 0;\ \theta_1 = 0.5$ $\theta_0 = 1;\ \theta_1 = 0.5$

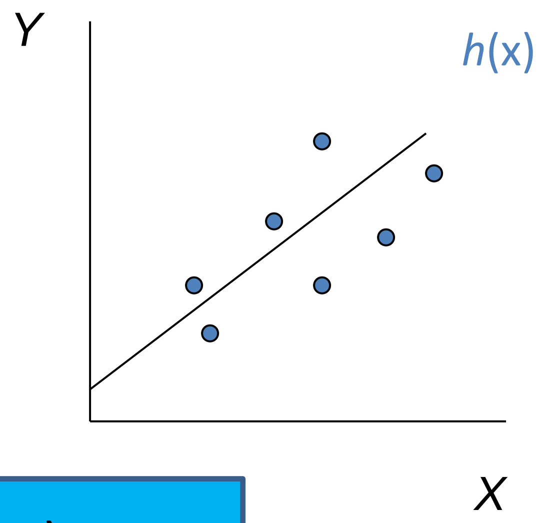
- θ_0 is the starting point of the line
- θ_1 represents the “slope” or angle of the line

Linear Regression



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

Cost Function



Hypothesis:

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

Parameters:

$$\theta_0, \theta_1$$

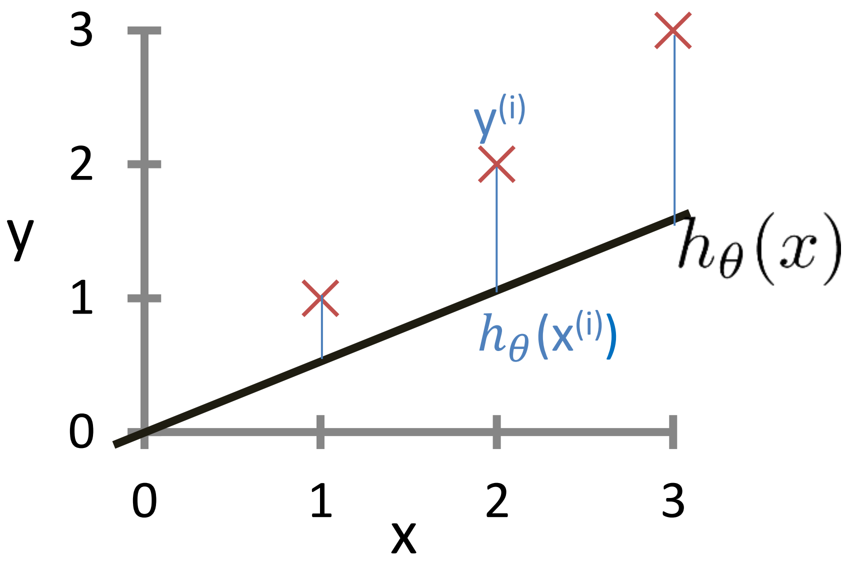
Cost Function (squared error function):

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

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

Visualizing Cost Function

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



J can be thought of as the sum of square of all the blue lines (divided by 2m)

Keep in mind: J is a function of θ_0 and θ_1. Why?

Because θ_0 and θ_1 define the line (hypothesis *h*). Cost function J is different for different lines.

Cost Function

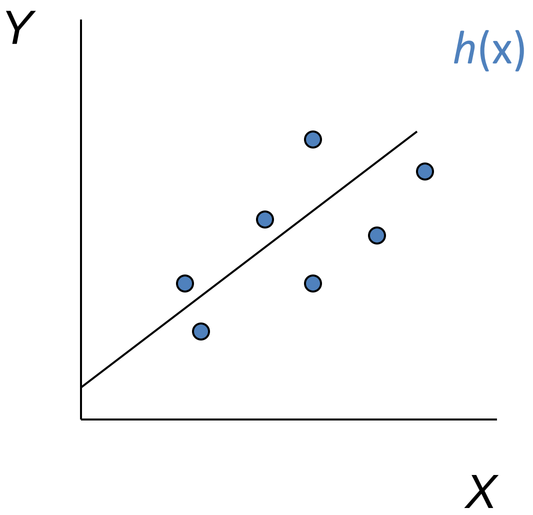
Why squares?

- Negatives and positives don’t cancel each other out. Suppose:

$$h(x^{(1)}) - y^{(1)} = 500$$

$$h(x^{(2)}) - y^{(2)} = -500$$

Then summation will be 0 even though h is not predicting y well at all!



- minimizes squared distance between training data and predicted line

- Math works nicely

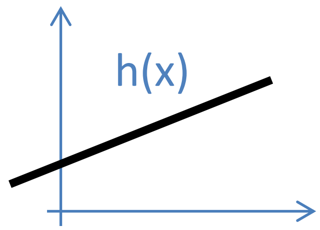
Simplified Cost Function

Hypothesis:

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

Parameters:

$$\theta_0 + \theta_1$$



Cost Function:

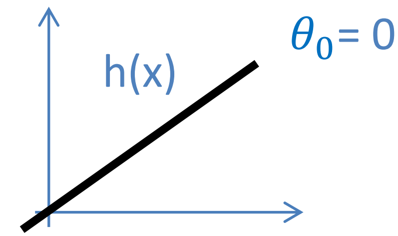
$$J\,(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

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

<u>Simplified</u>

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

$$\theta_1$$

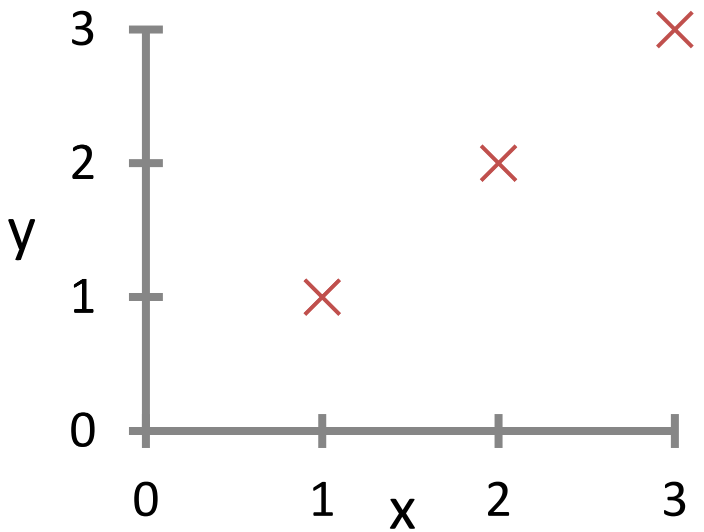


$$J\,(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

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

Plotting the Cost Function

We plotted h, but can we plot J?



Say you the given the training data set such that:

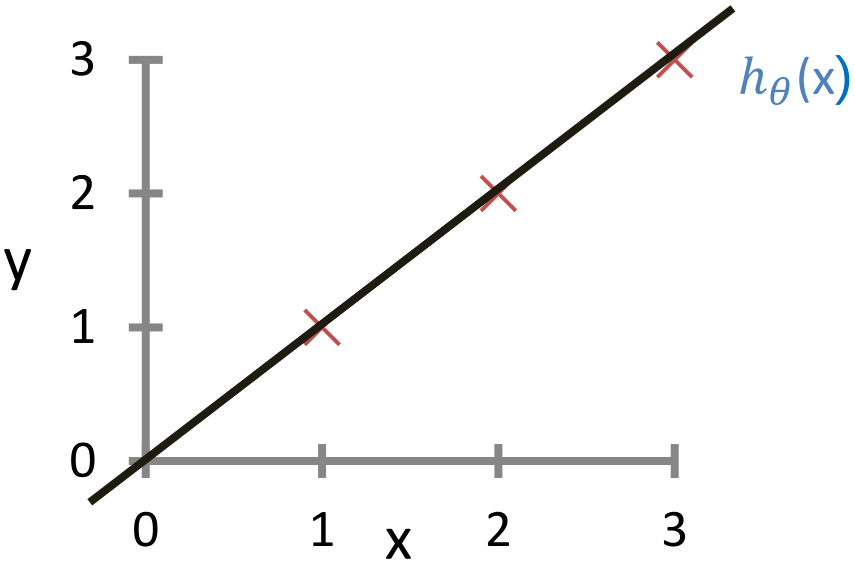
$x^{(i)} = y^{(i)}$ for every i

Of course, correct hypothesis h for this set is h(x) = x.
Hence θ_1 = 1 (and θ_0 = 0).

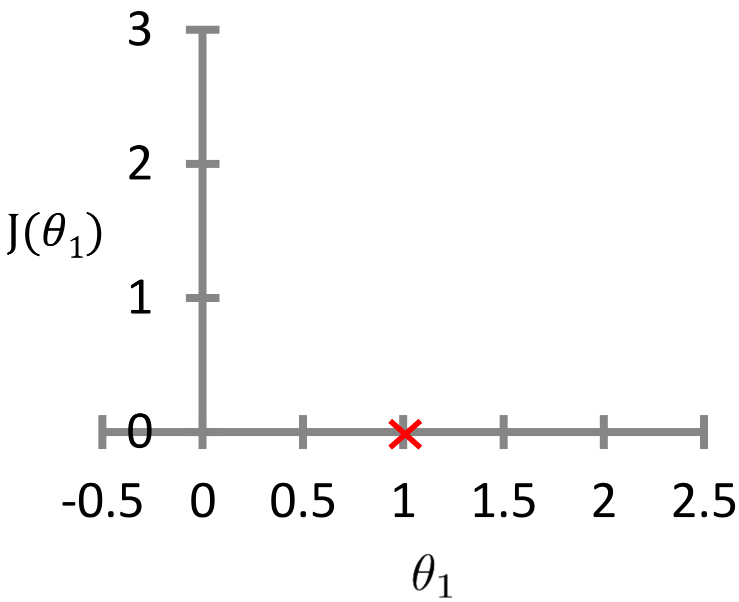
But let's compute J for different values of θ_1.

Plotting the Cost Function

$\theta_1 = 1$



$J(1) = 0$



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

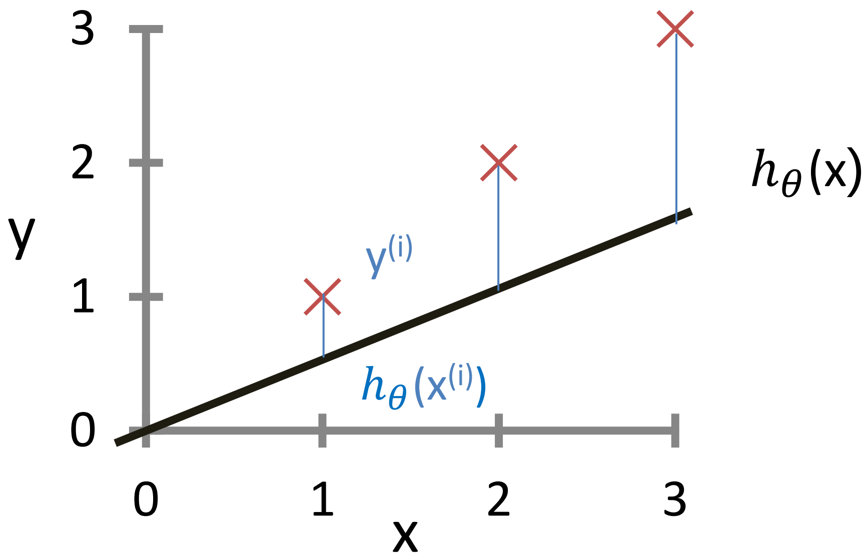
$$= \frac{1}{2m} \sum_{i=1}^{m} (\theta_1 x^{(i)} - y^{(i)})^2$$

$$= \frac{1}{2m} (0^2 + 0^2 + 0^2)$$

Left: plot of h, Right: plot of J

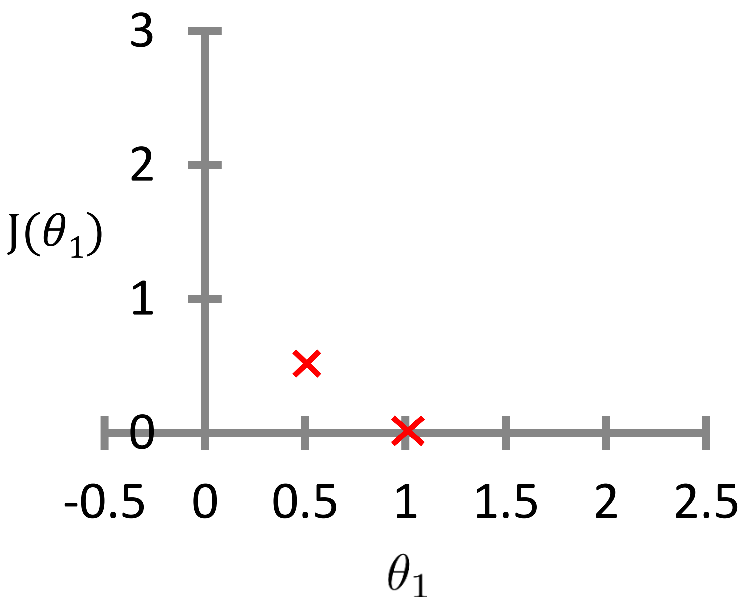
Plotting the Cost Function

$\theta_1 = 0.5$



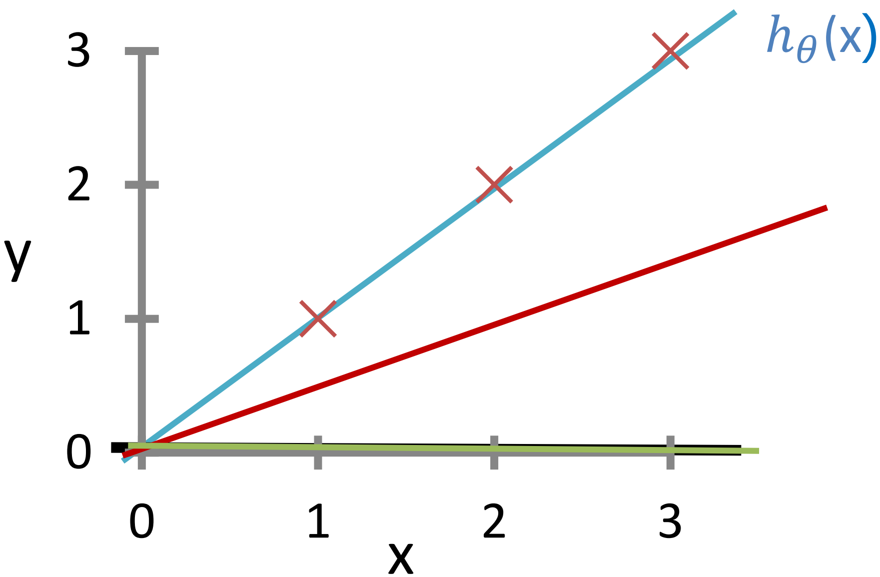
$$J(0.5) = \frac{1}{2\cdot 3}\sum_{i=1}^{3}\left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right]$$

$$= \frac{1}{6}\cdot(3.5) = 0.58$$



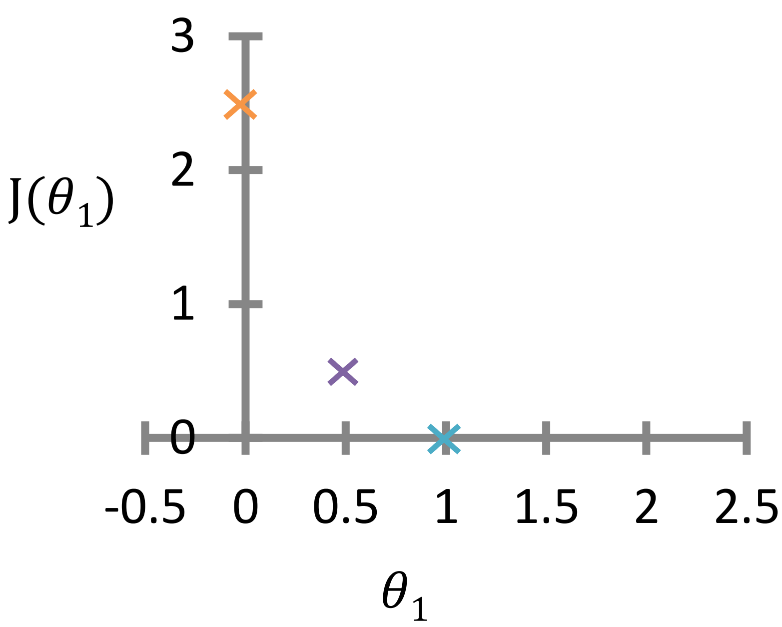
Plotting the Cost Function

$\theta_1 = 0$

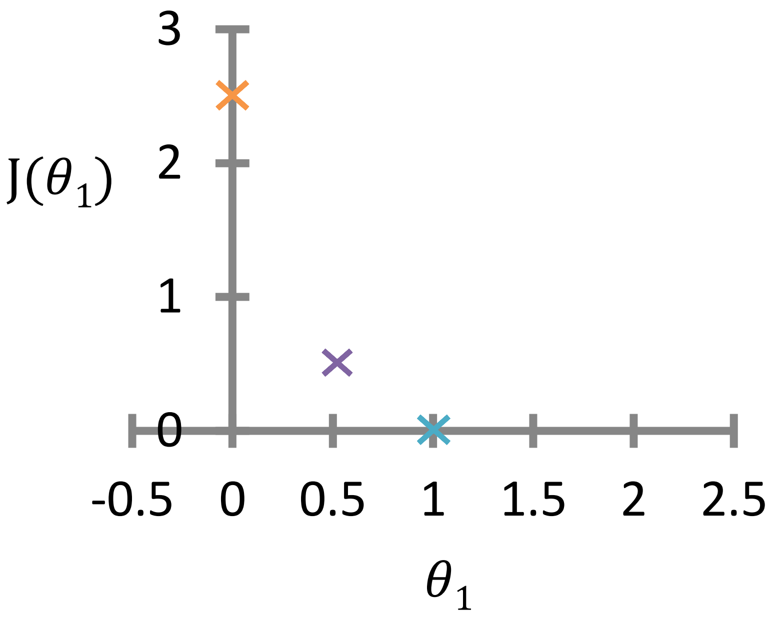


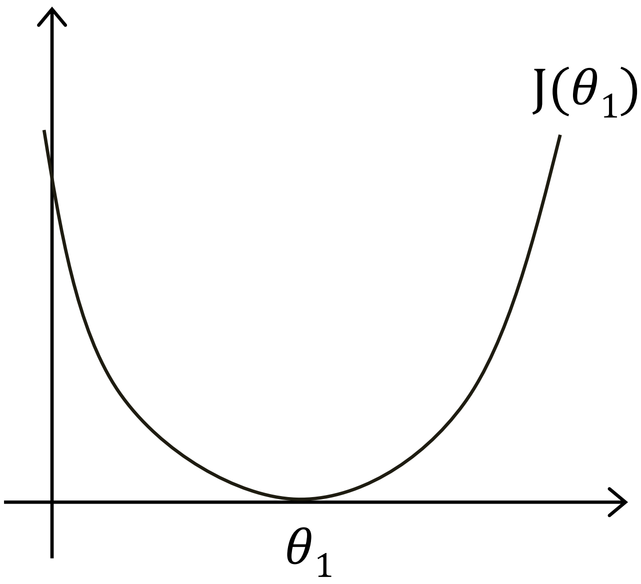
$$J(0) = \frac{1}{2 \cdot 3} \sum_{i=1}^{3} [1^2 + 2^2 + 3^2]$$

$$= \frac{1}{6} \cdot 14 = 2.3$$



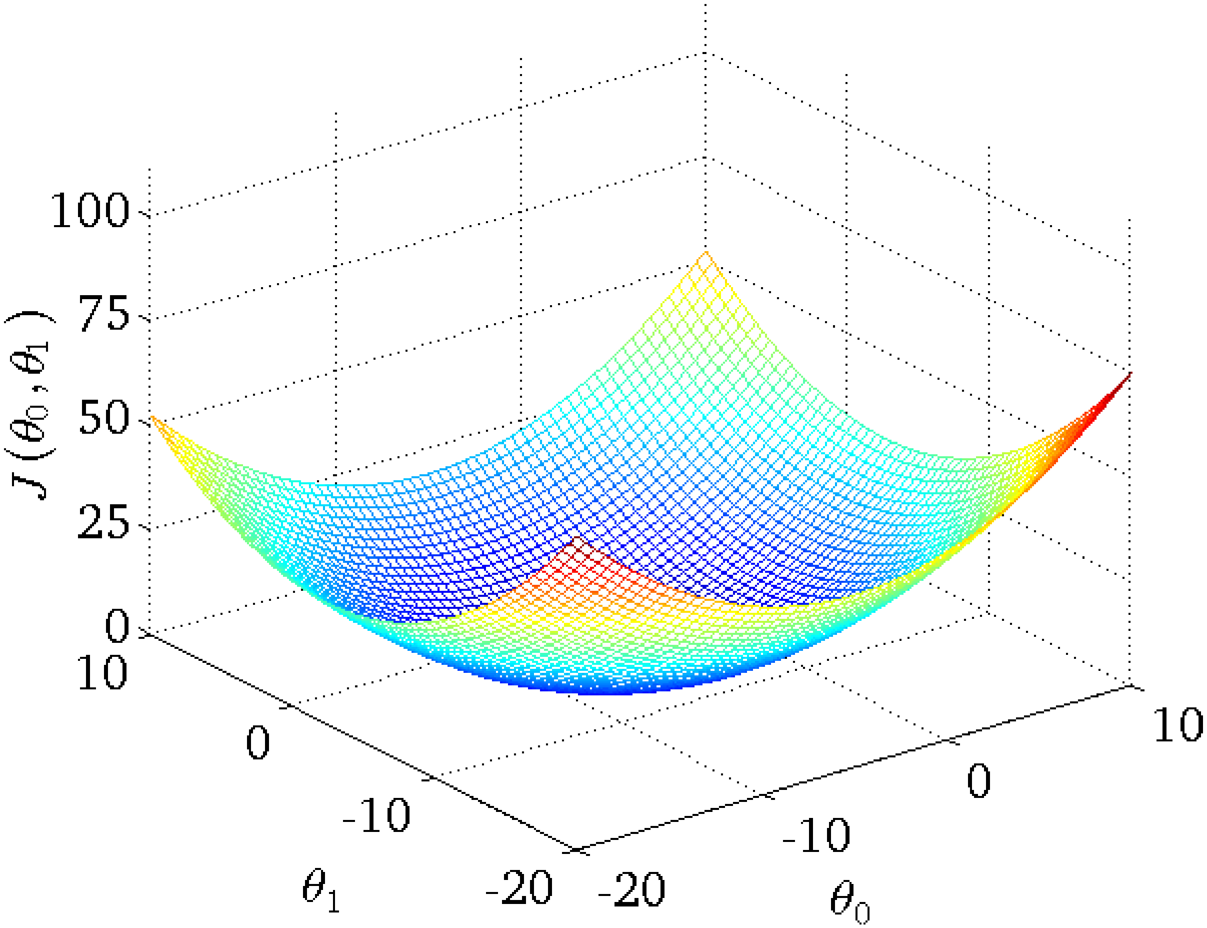
Completing the Plot





This is a 1-dimensional cost function. θ_1 is the only variable here. What if θ_0 was also non-zero?

2-Dimensional Cost Function



Goal: Find θ_0 and θ_1 for which the cost function is minimized. Gradient Descent Algorithm!

Questions?