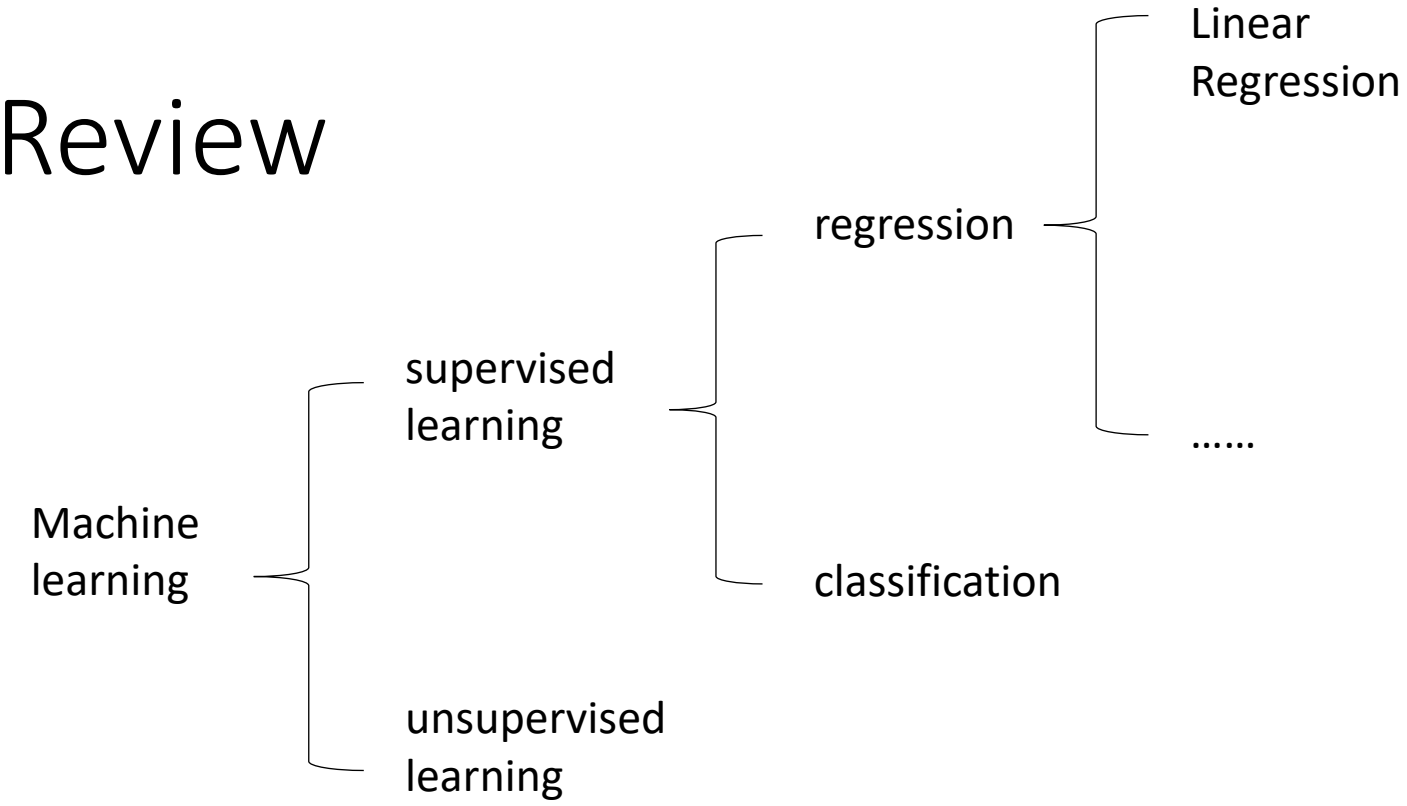
L7: Linear Regression

Prof. Xun Jiao

Before class

- HW 1 due
- HW2 will be posted

Review


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

Given **labeled training data** of the form

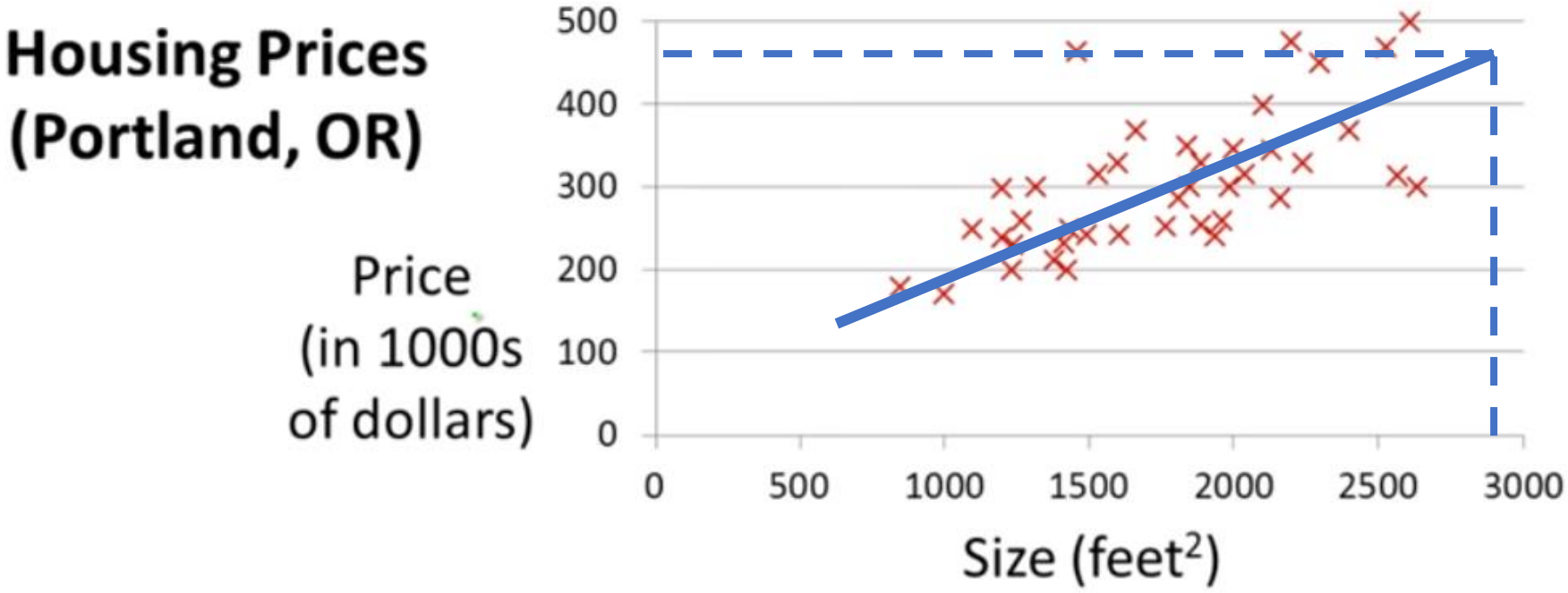
$$\{(\mathrm{data}_1, \mathrm{label}_1), \ldots, (\mathrm{data}_n, \mathrm{label}_n)\}$$

Infer the function

$$f(\mathrm{data}) \xrightarrow{?} \mathrm{labels}$$

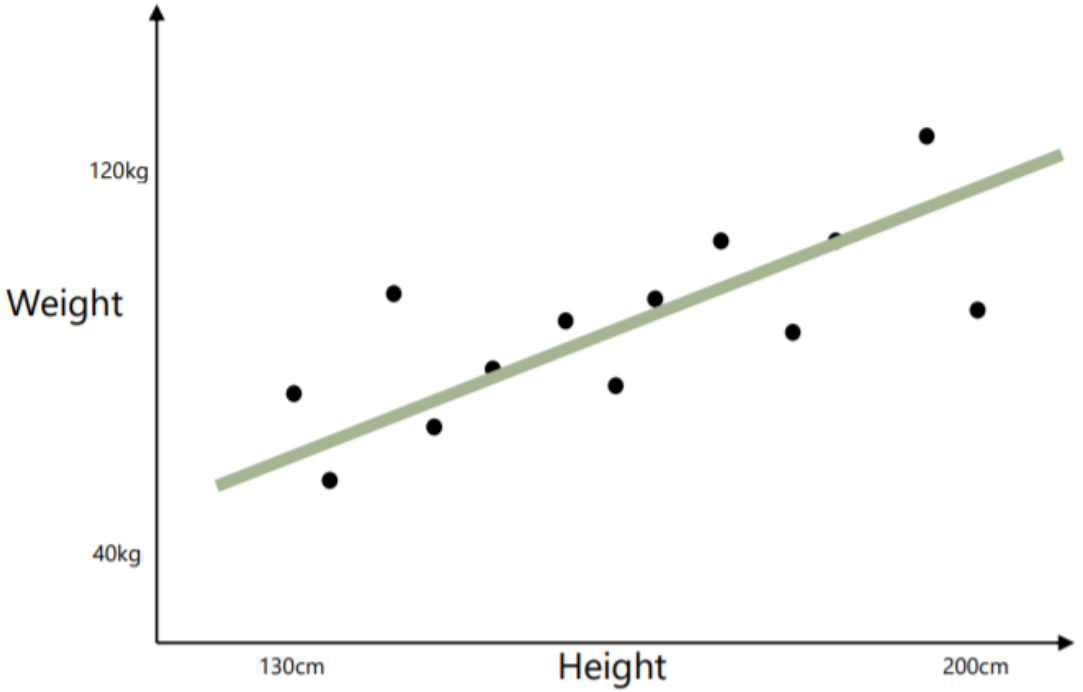
The easiest assumption: linear relationship

- The easiest assumption would be a linear relationship (mapping) from data to label.



Another linear example

Q: Can we find a line that (approximately) fits the data?



What is a linear equation?

Linear regression assumes a predictor of the form

$$X\theta = y$$

matrix of features (data)

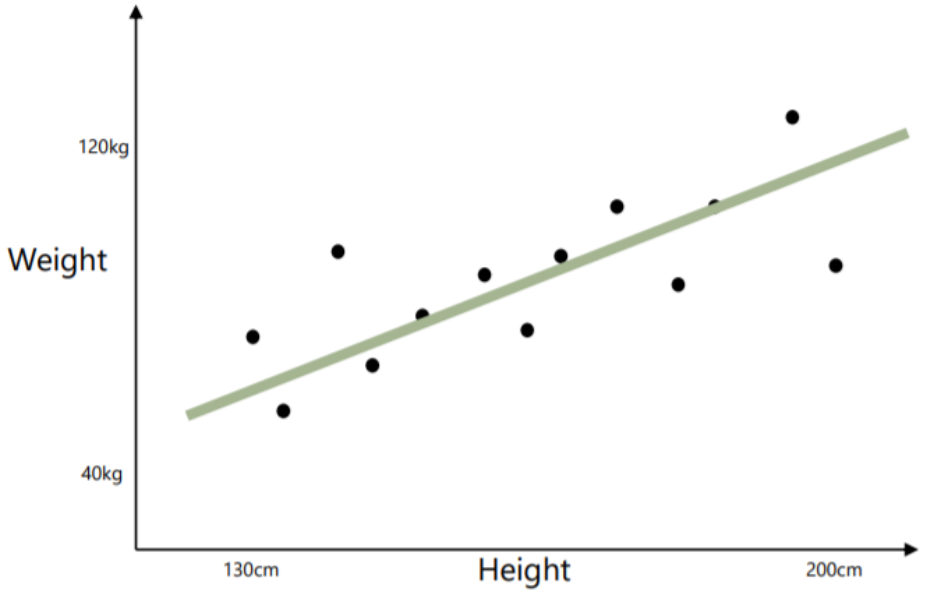
unknowns (which features are relevant)

vector of outputs (labels)

(or $Ax = b$ if you prefer)

Recap: equation for a line

What is the formula describing the line?

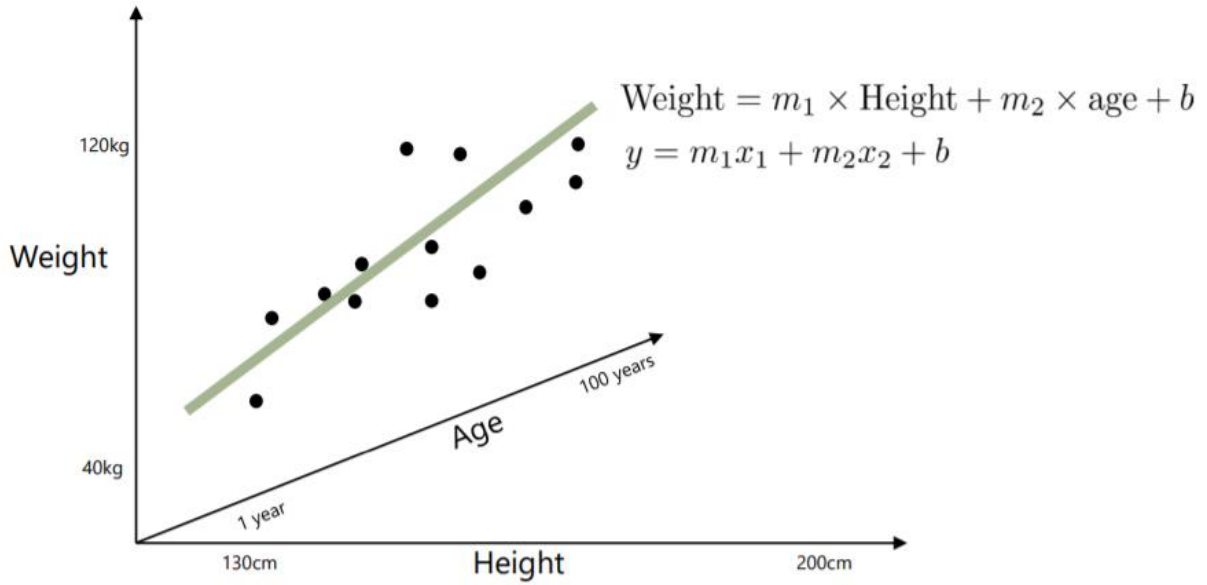


$$y = mx + b$$

$$\text{Weight} = m \times \text{Height} + b$$

Recap: equation for a line

What about in more dimensions?

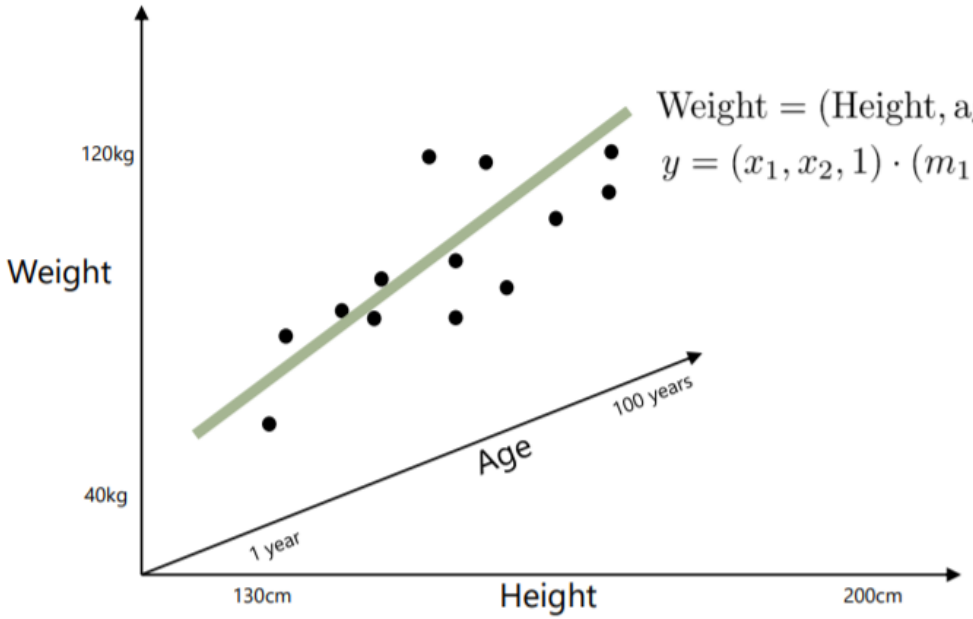


$$\text{Weight} = m_1 \times \text{Height} + m_2 \times \text{age} + b$$

$$y = m_1 x_1 + m_2 x_2 + b$$

Recap: equation for a line as an inner product

What about in more dimensions?



$$\text{Weight} = (\text{Height}, \text{age}, 1) \cdot (m_1, m_2, b)$$

$$y = (x_1, x_2, 1) \cdot (m_1, m_2, b)$$

Our goal: find what is m???

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

Training Set

	Size in feet² (x)	Price ($) in 1000's (y)	
x^1	2104	460	y^1
x^2	1416	232	y^2
x^3	1534	315	y^3
	852	178	
	...	...	

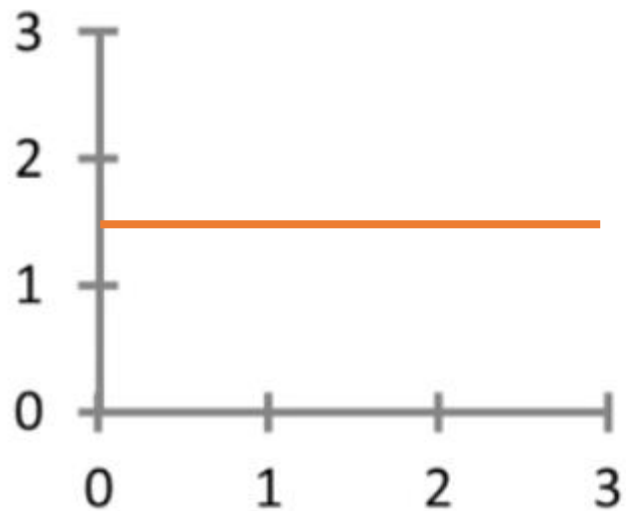
m training samples

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

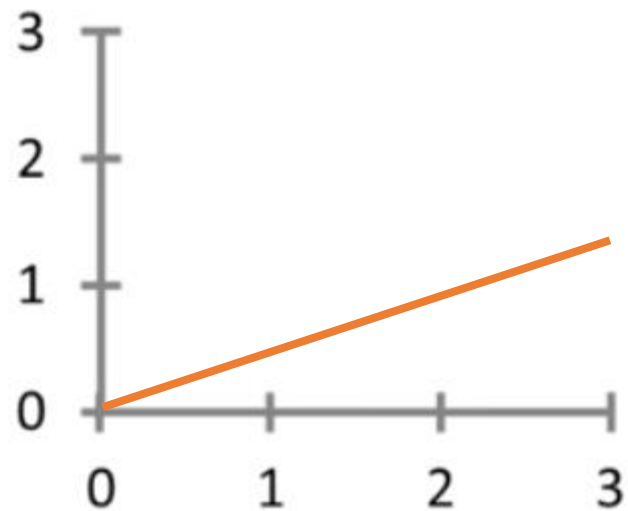
θ_i's: Parameters

How to find them???

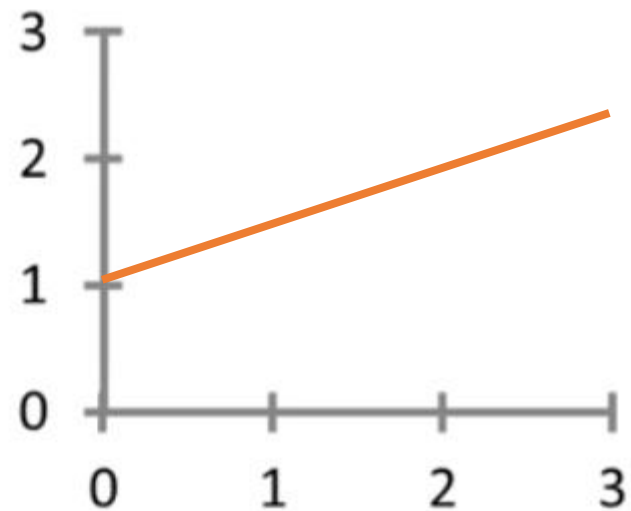
$$h_\theta(x) = \theta_0 + \theta_1 x$$



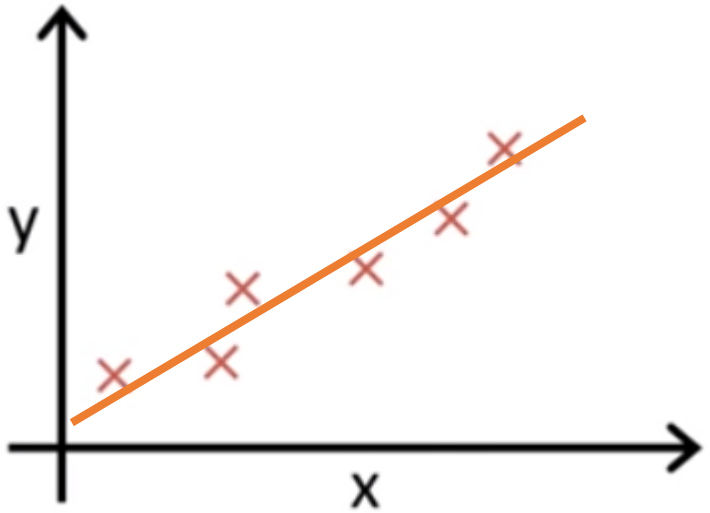
$\theta_0 = 1.5$
$\theta_1 = 0$



$\theta_0 = 0$
$\theta_1 = 0.5$



$\theta_0 = 1$
$\theta_1 = 0.5$



What is this?

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

This is equivalent of saying: we want to minimize the difference (error) between predicted value and true value.

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

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

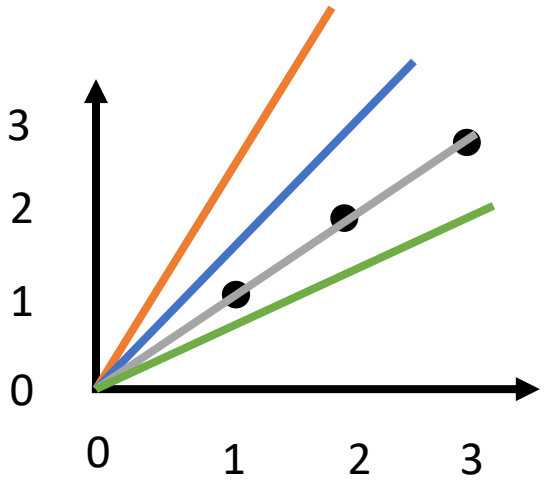
Parameters: θ_0, θ_1

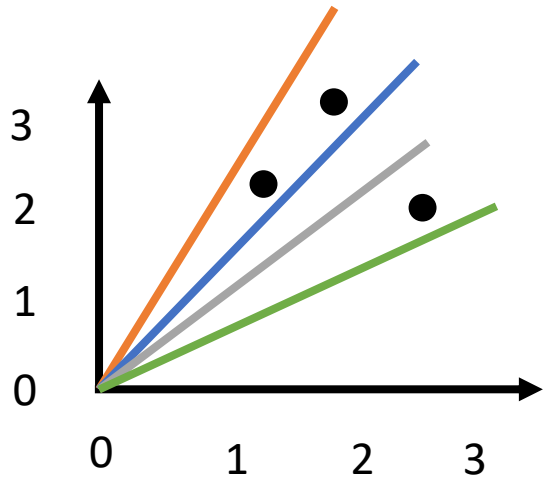
Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$

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



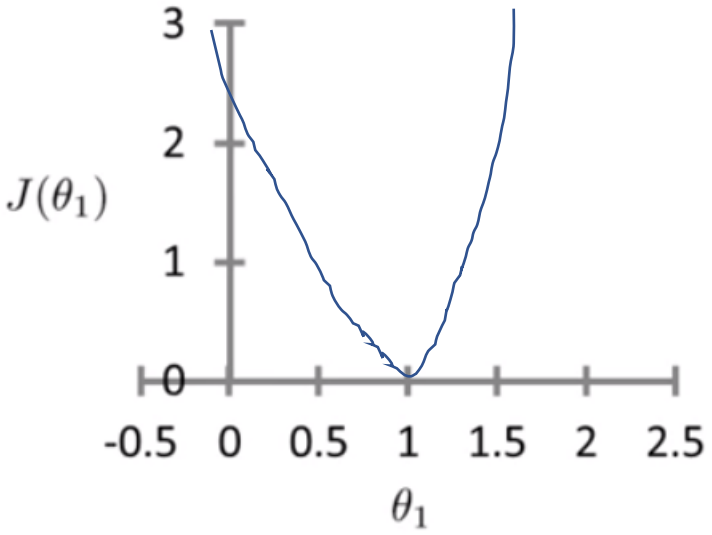


Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

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

$J(\theta_1)$

(function of the parameter θ_1)



What is the J() value under different theta?

- Find theta that can minimize J()