LINEAR REGRESSION

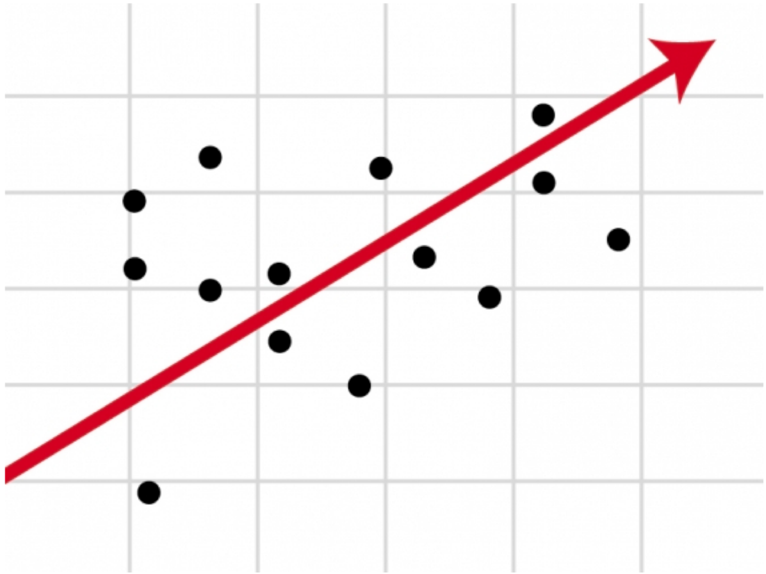
Narges Norouzi

CONTENT

- Introduction
- Regression – Definition
- Linear Regression
- Least Square Method

INTRODUCTION

- Analyze the specific relationships between two or more variables
- This is done to gain the information about one through knowing values of the other variables



REGRESSION

- A statistical measure that attempts to determine the strength of the relationship between one dependent variable (usually denoted by Y) and a series of other changing variables (known as independent variables)

- Forecast value of a dependent variable (Y) from the value of independent variables (X_1, X_2, X_3, …)

- It is widely used for prediction, estimation, hypothesis testing, and modeling causal relationships

DEPENDENT & INDEPENDENT VARIABLES

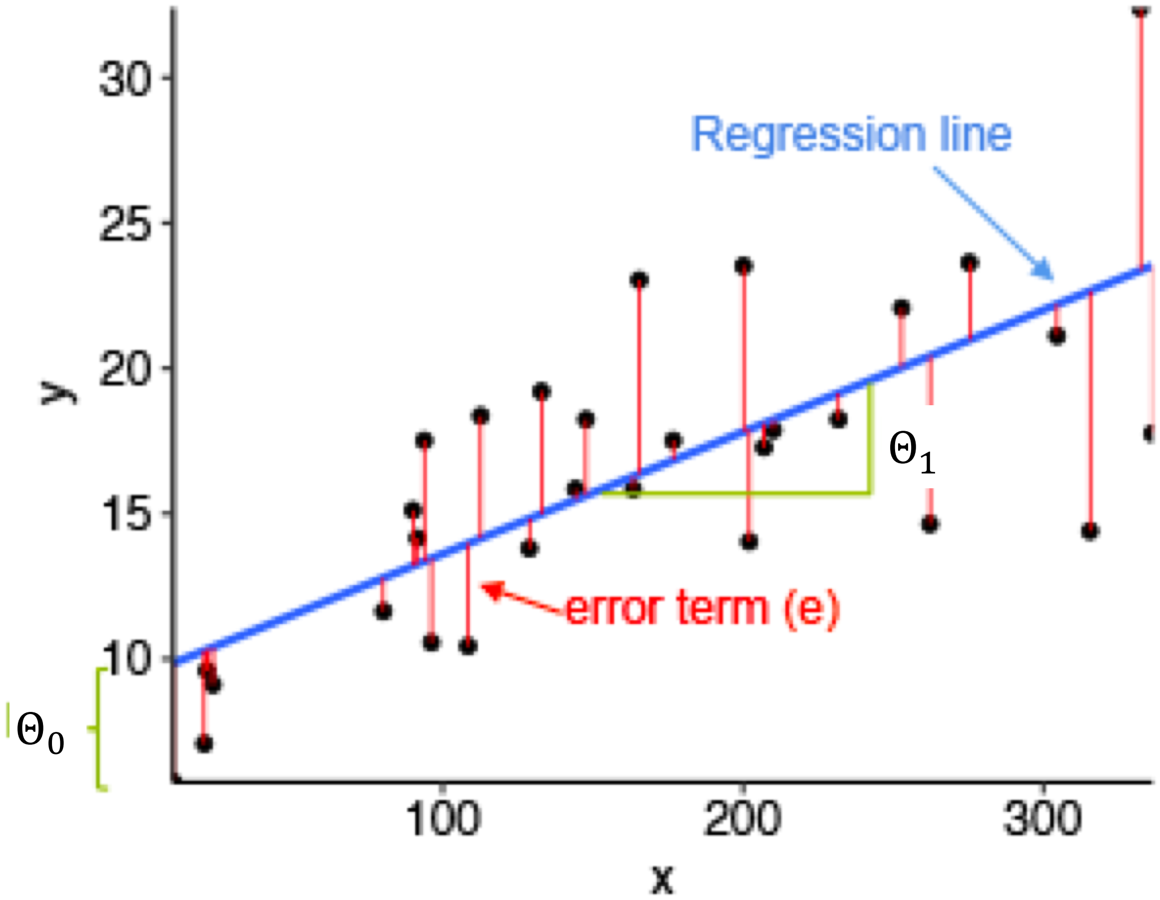
- Independent variables are regarded as inputs to a system and may take on different values freely.
- Dependent variables are those values that change as a consequence of changes in other values in the system.
- Independent variable is also called as predictor or explanatory variable and it is denoted by X.
- Dependent variable is also called as response variable and it is denoted by Y.

THE FIRST ORDER LINEAR MODEL

$$Y = \theta_0 + \theta_1 X$$

- Y = dependent/outcome/response variable
- X = independent/predictor/explanatory variable
- θ_0 = Y-intercept
- θ_1 = slope of the line

LINEAR REGRESSION EXAMPLE



https://nextjournal.com/intelrefinery/simple-linear-regression

REGRESSION DATA

- Given
  - Data:

$$X = \{x^{(1)}, x^{(2)}, x^{(3)}, \dots., x^{(n)}\} \text{ where } x^{(i)} \in \mathcal{R}^d$$

  - Corresponding labels:

$$Y = \{y^{(1)}, y^{(2)}, y^{(3)}, \dots., y^{(n)}\} \text{ where } y^{(i)} \in \mathcal{R}$$

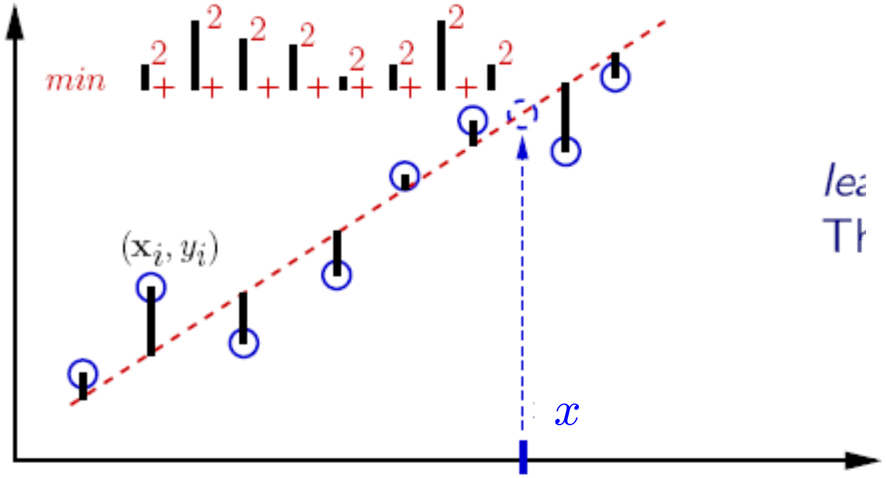
REGRESSION HYPOTHESIS

- Hypothesis:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ \ldots + \theta_d x_d = \sum_{j=0}^{d} \theta_j x_j$$

Assume $x_0 = 1$

- Fit a model by minimizing sum or mean of squared errors



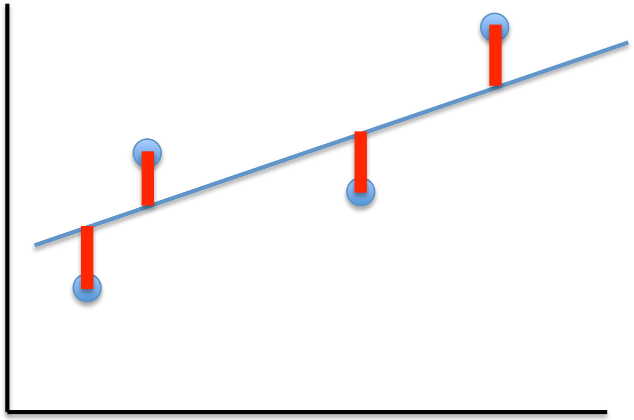
LEAST SQUARES LINEAR REGRESSION

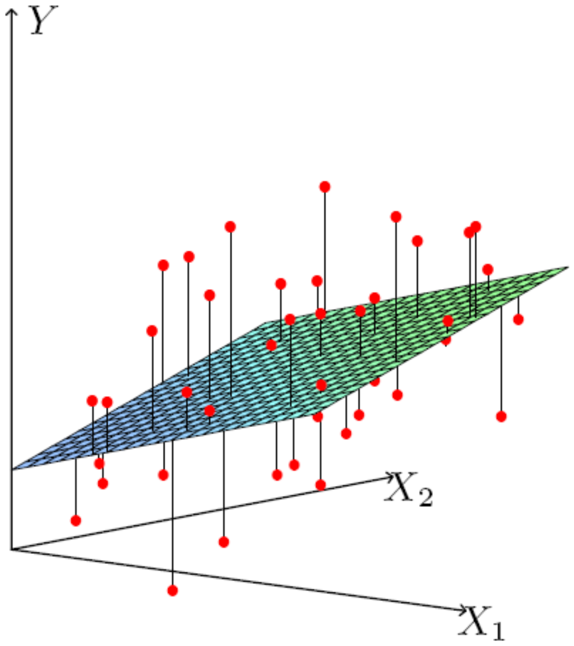
- Cost Function

$$Cost(\theta) = \frac{1}{2\times n}\sum_{i=0}^{n}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

- Fit by solving

$$\min_\theta Cost(\theta)$$

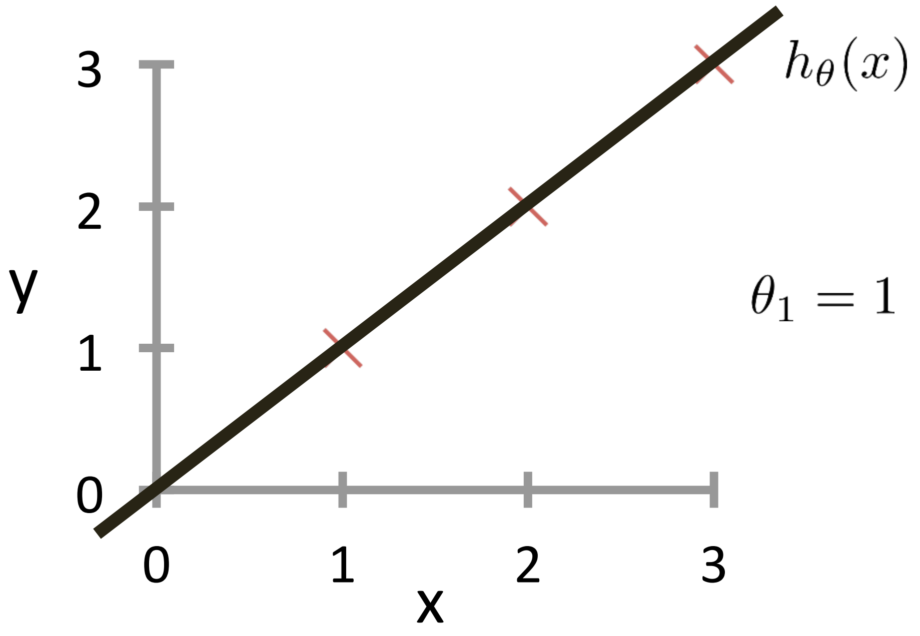




INTUITION BEHIND COST FUNCTION

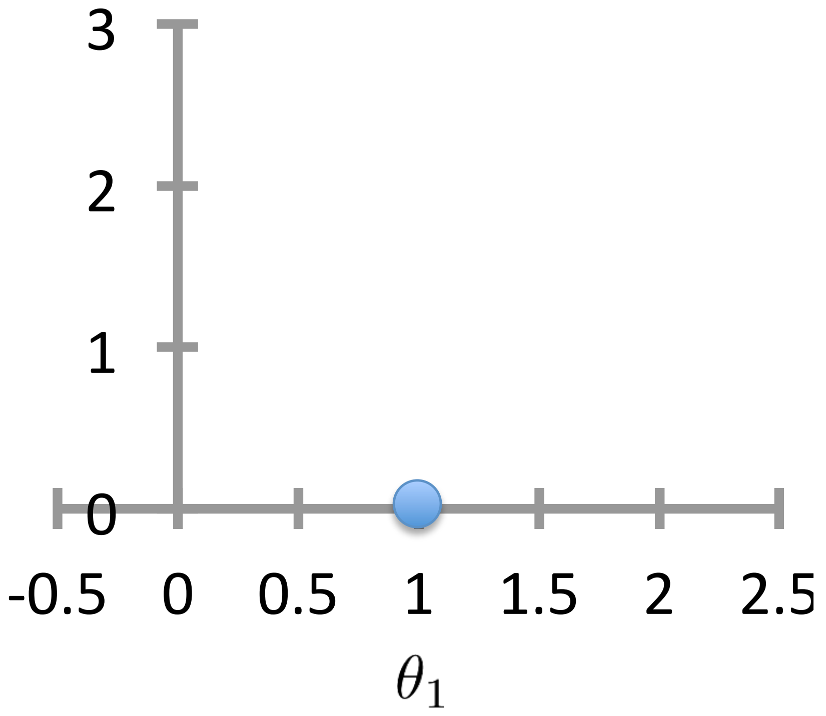
$h_\theta(x)$

For a fixed θ_1. This is a function of x



$Cost(\theta_1)$

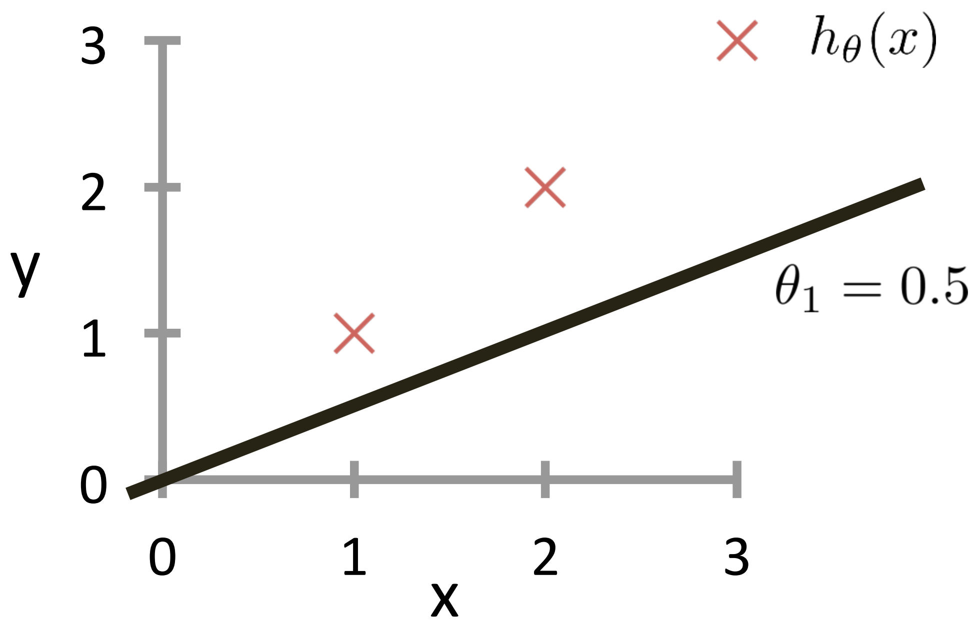
Function of a parameter θ_1



INTUITION BEHIND COST FUNCTION

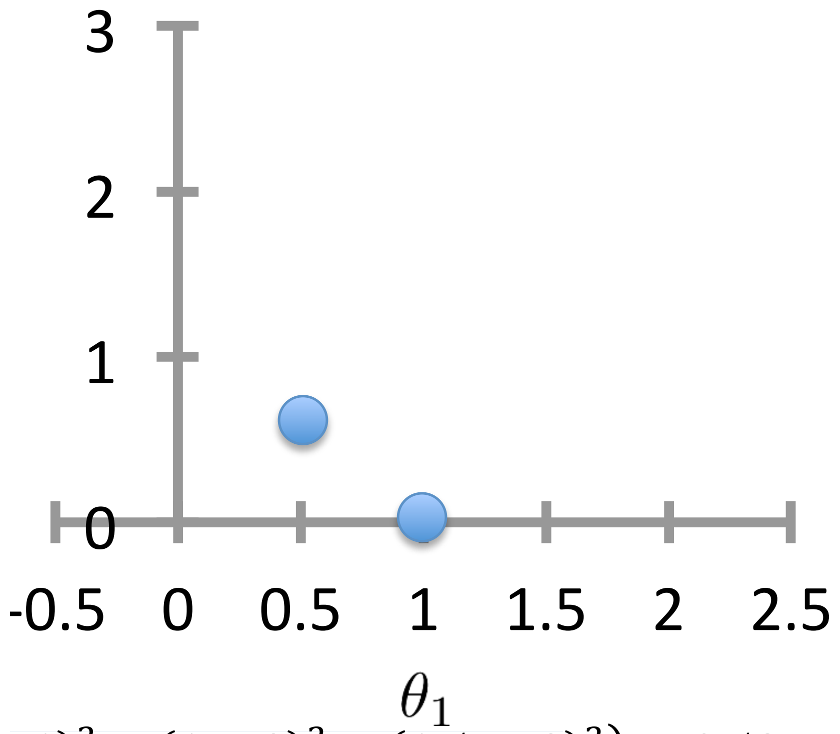
$h_\theta(x)$

For a fixed θ_1. This is a function of x



$Cost(\theta_1)$

Function of a parameter θ_1

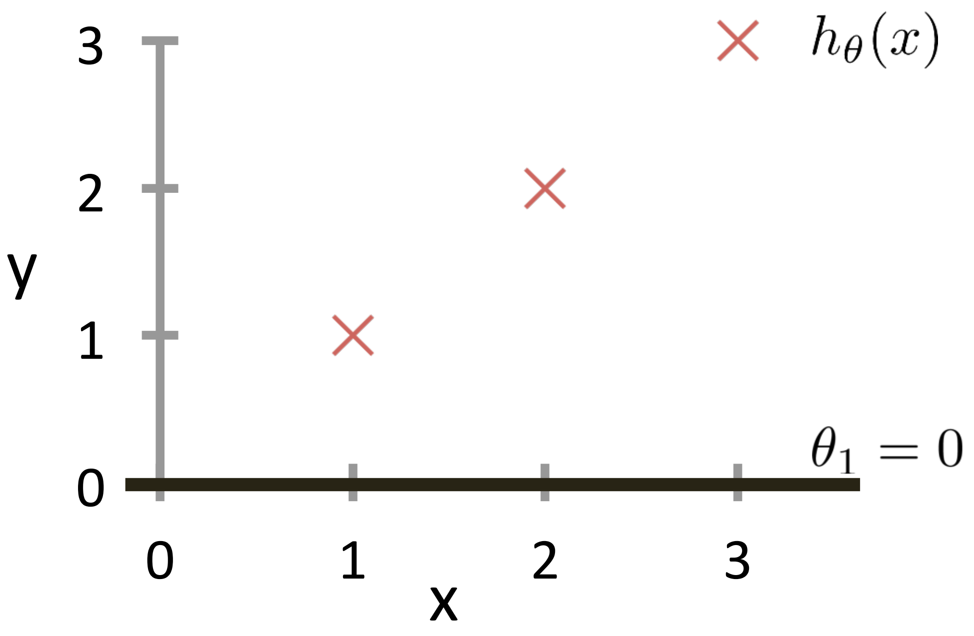


$$Cost(\theta) = Cost([\theta_0, \theta_1]) = Cost([0, 0.5]) = \frac{1}{2\times 3}\left((0.5 - 1)^2 + (1-2)^2 + (1.5 - 3)^2\right) = 0.58$$

INTUITION BEHIND COST FUNCTION

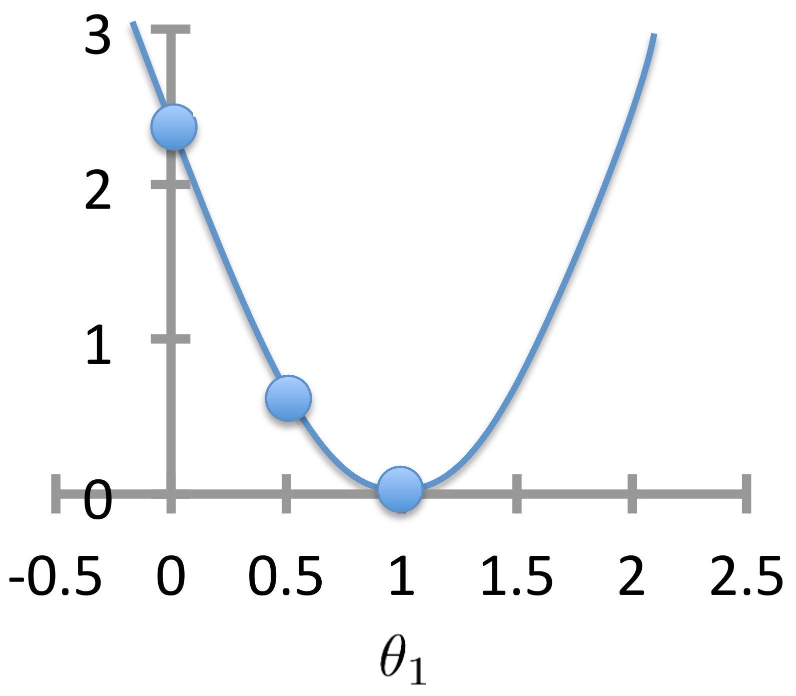
$h_\theta(x)$

For a fixed θ_1. This is a function of x



$Cost(\theta_1)$

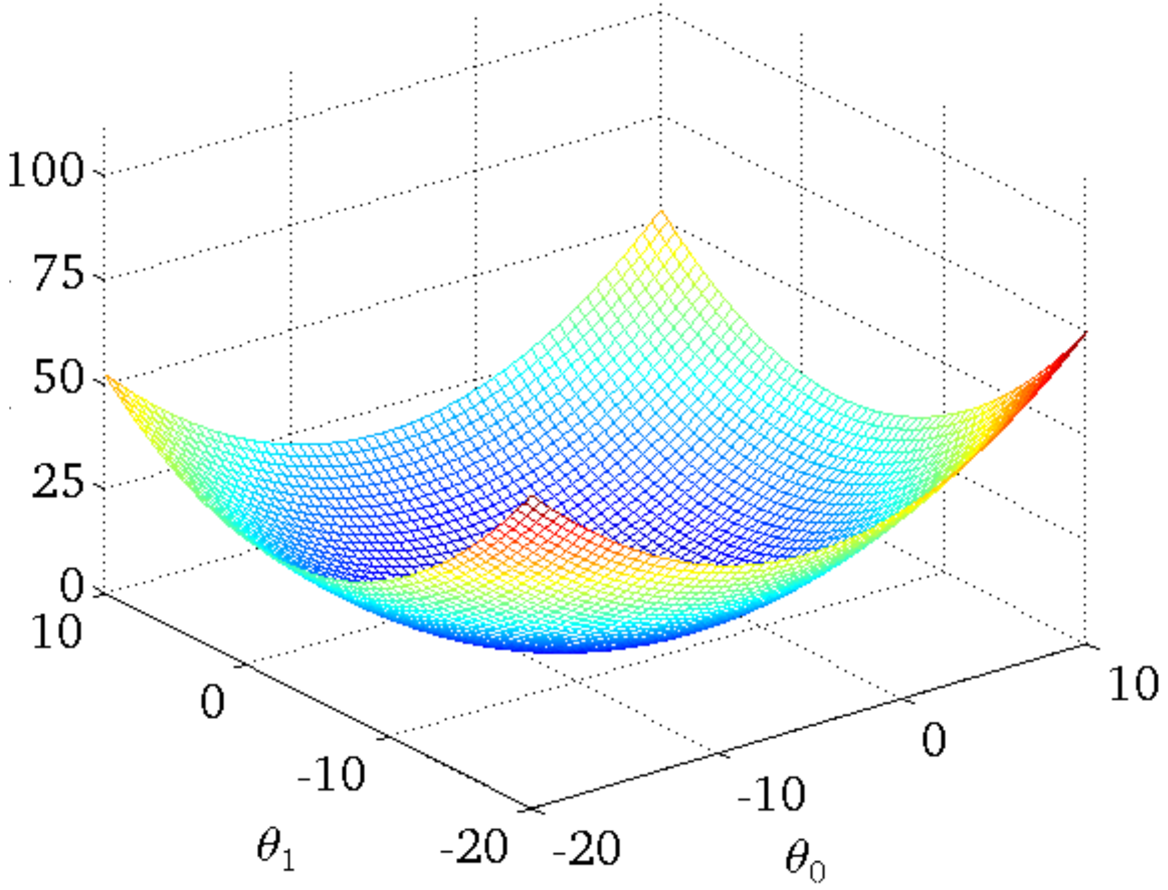
Function of a parameter θ_1



$$Cost(\theta) = Cost([\theta_0, \theta_1]) = Cost([0,0]) = \frac{1}{2\times 3}\left((0-1)^2 + (0-2)^2 + (0\ -3)^2\right) = 2.33$$

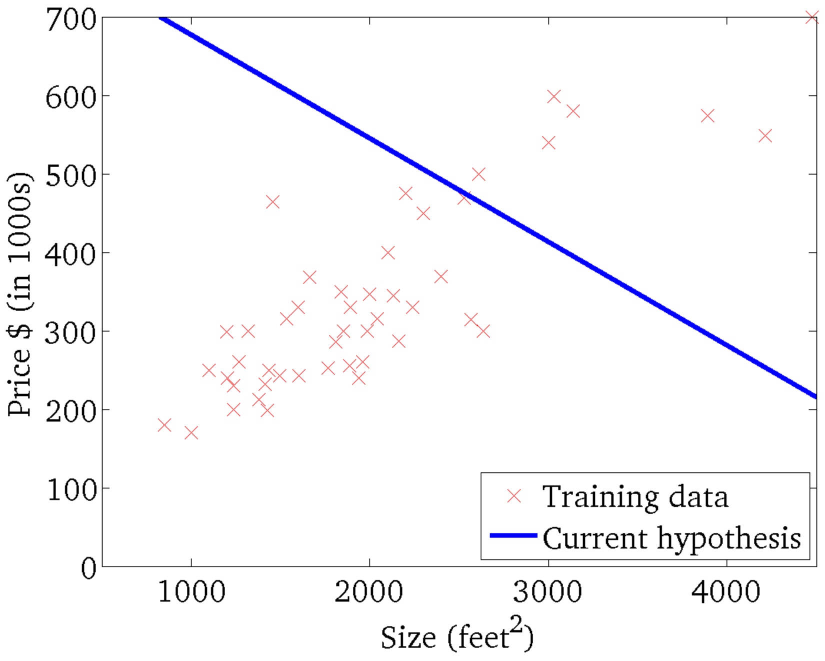
INTUITION BEHIND COST FUNCTION

$$Cost([\theta_0, \theta_1])$$

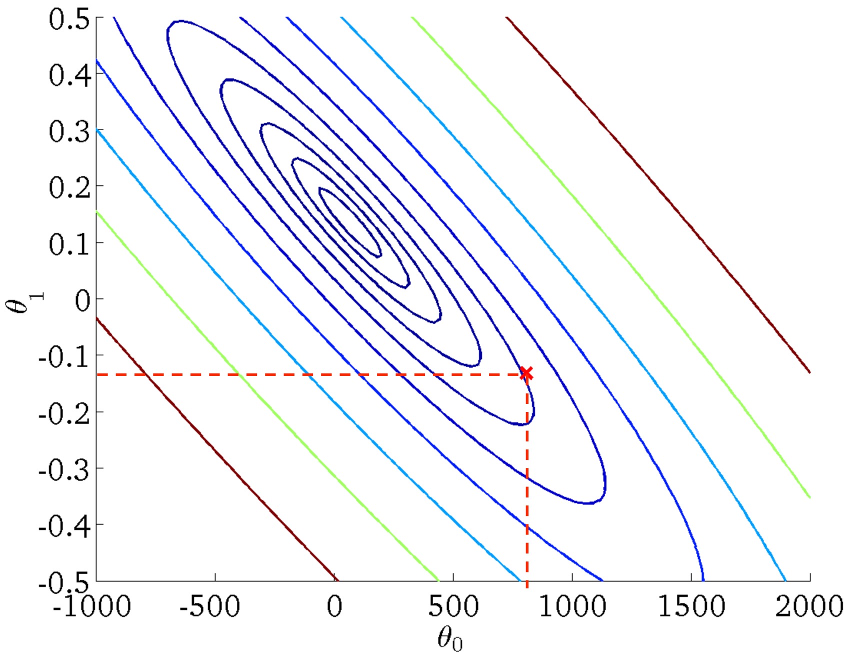


INTUITION BEHIND COST FUNCTION

$$h_\theta(x)$$



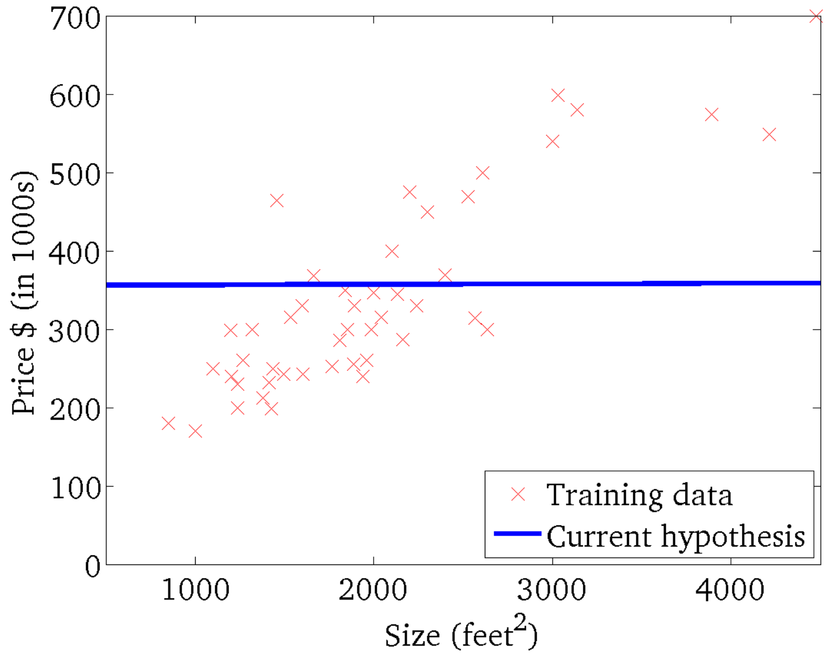
$$Cost([\theta_0, \theta_1])$$

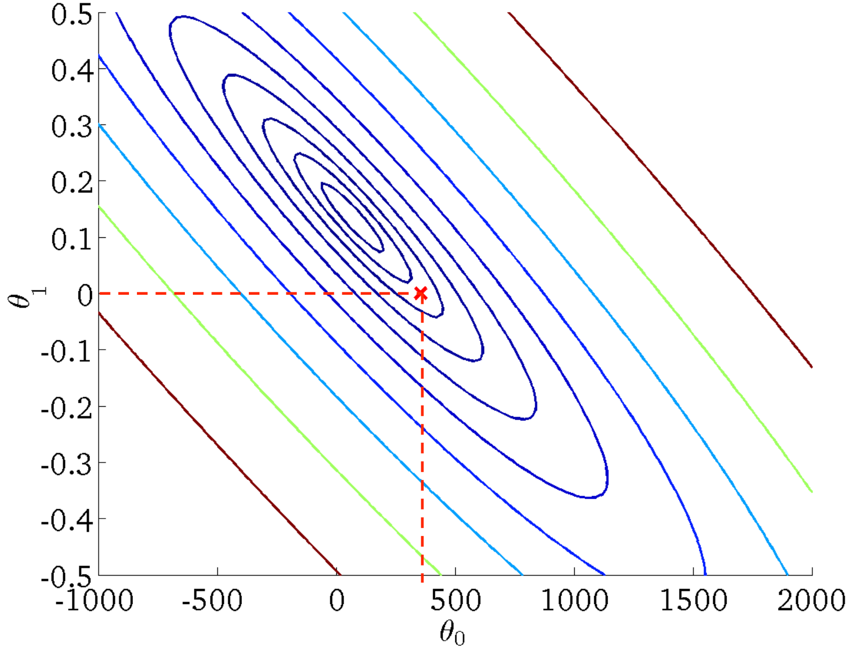


INTUITION BEHIND COST FUNCTION

$h_\theta(x)$

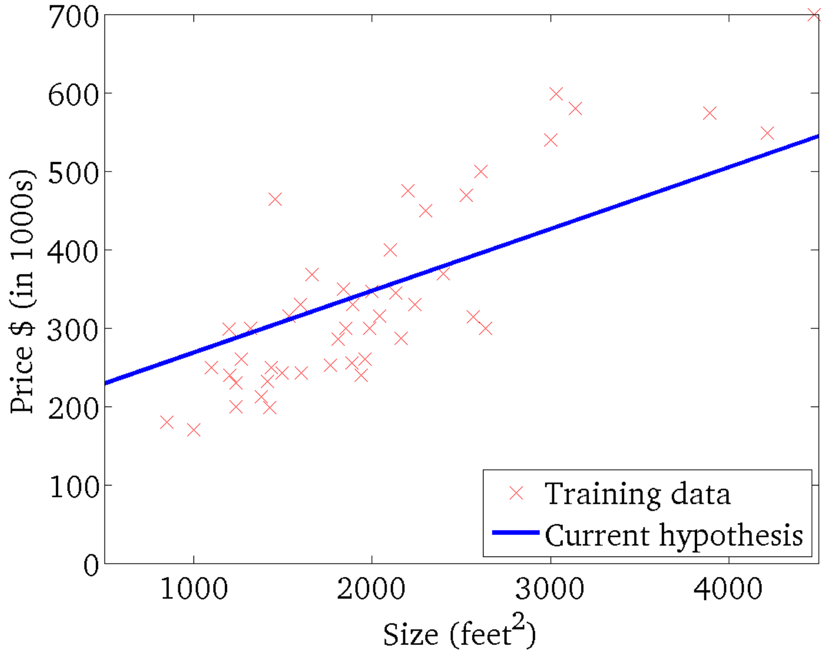
$Cost([\theta_0, \theta_1])$



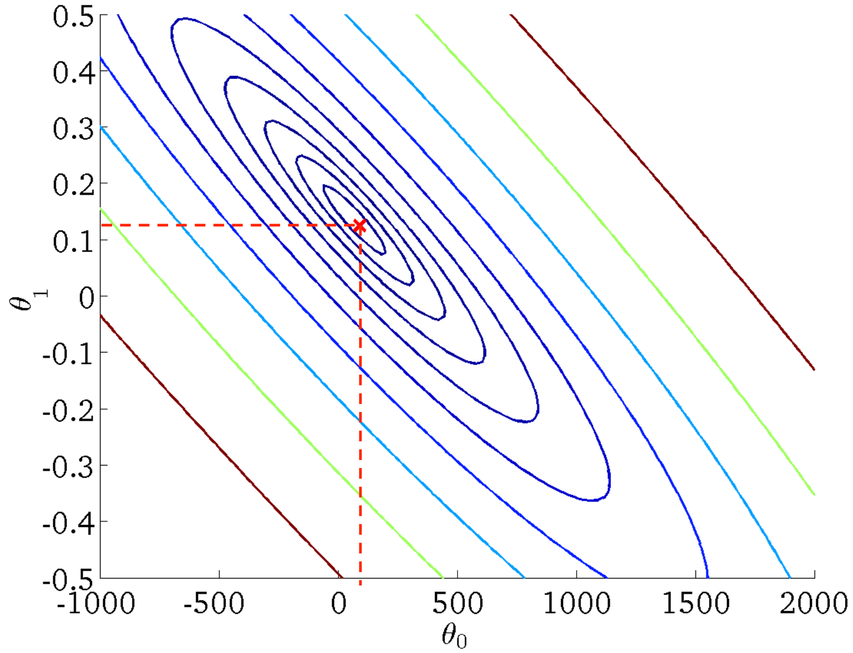


INTUITION BEHIND COST FUNCTION

$$h_\theta(x)$$



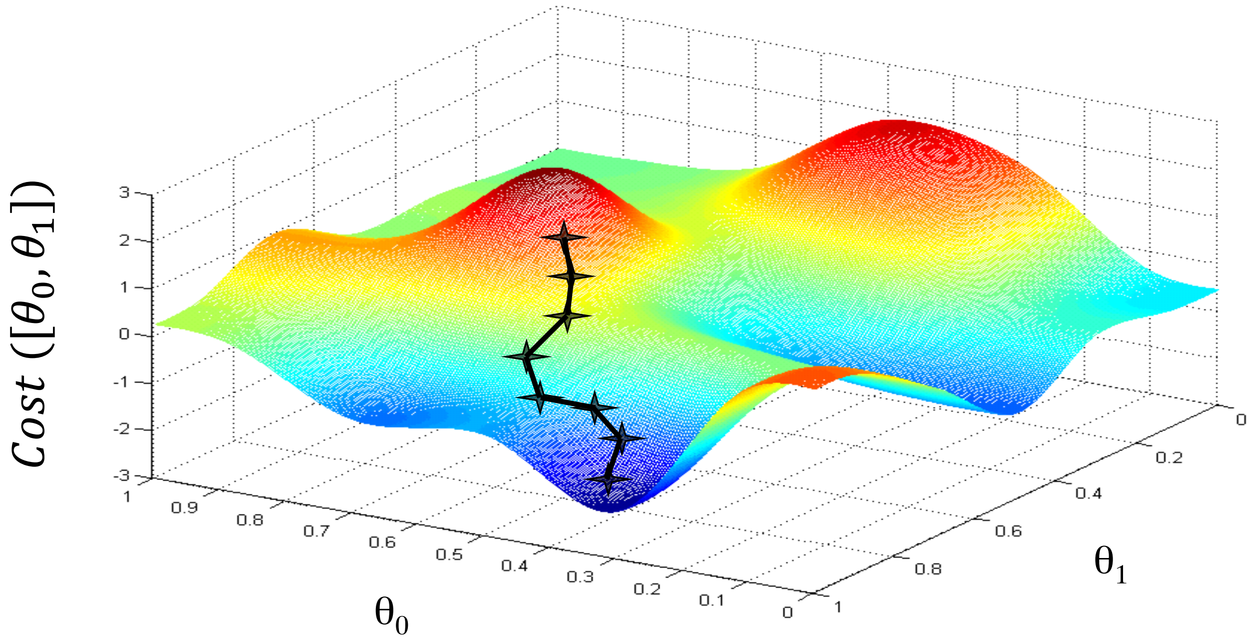
$$Cost([\theta_0, \theta_1])$$



BASIC SEARCH PROCEDURE

- Choose initial value for θ
- Until we reach a minimum
  - Choose a new value for θ to reduce $Cost\ (\theta)$

Since the least squares objective function is convex, we don't need to worry about local minima

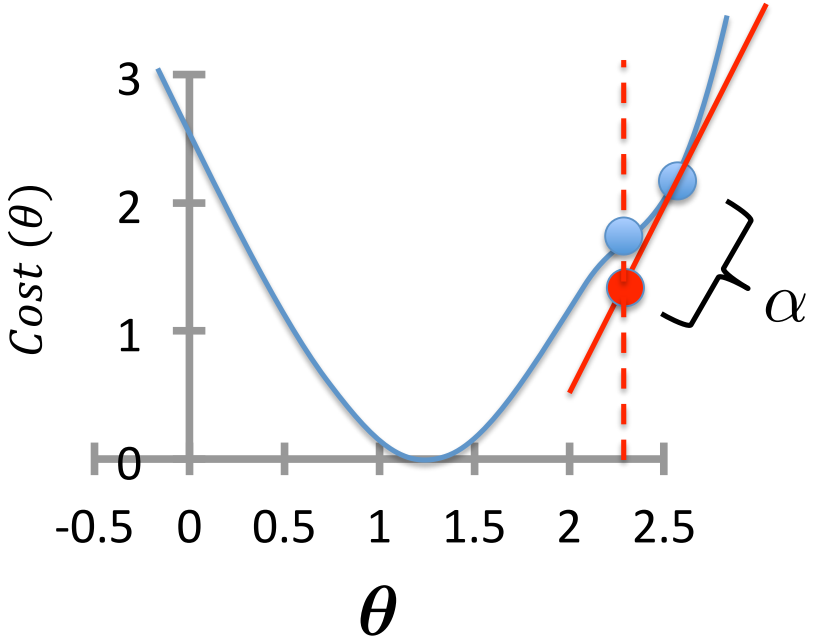


GRADIENT DESCENT

- Initialize θ
- Repeat until convergence

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial Cost(\theta)}{\partial \theta_j} \qquad (simultaneous\ update\ for\ \theta_0, \theta_1, \dots, \theta_d)$$

Learning rate



GRADIENT DESCENT

- Initialize θ
- Repeat until convergence

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial Cost(\theta)}{\partial \theta_j} \qquad (simultaneous\ update\ for\ \theta_0, \theta_1, \dots, \theta_d)$$

- For linear regression:

$$\frac{\partial Cost(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \frac{1}{2n} \sum_{i=1}^{n} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

With $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_d x_d = \sum_{j=0}^{d} \theta_j x_j$

GRADIENT DESCENT

$$\frac{\partial Cost(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \frac{1}{2n} \sum_{i=1}^{n} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

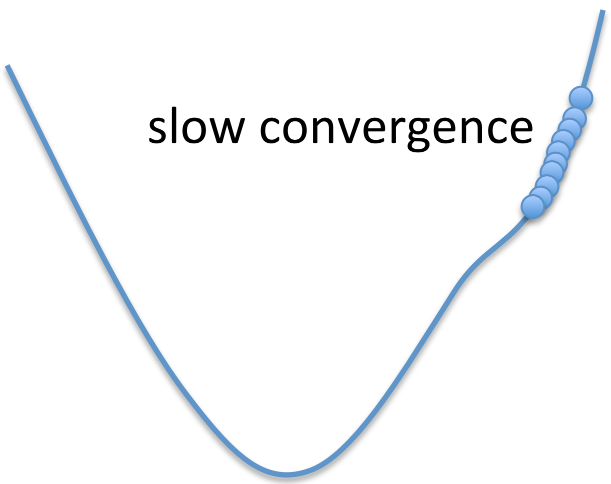
$$= \frac{\partial}{\partial \theta_j} \frac{1}{2n} \sum_{i=1}^{n} \left(\sum_{k=0}^{d} \theta_k x_k^{(i)} - y^{(i)} \right)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} (\sum_{k=0}^{d} \theta_k x_k^{(i)} - y^{(i)}) \frac{\partial}{\partial \theta_j} (\sum_{k=0}^{d} \theta_k x_k^{(i)} - y^{(i)})$$

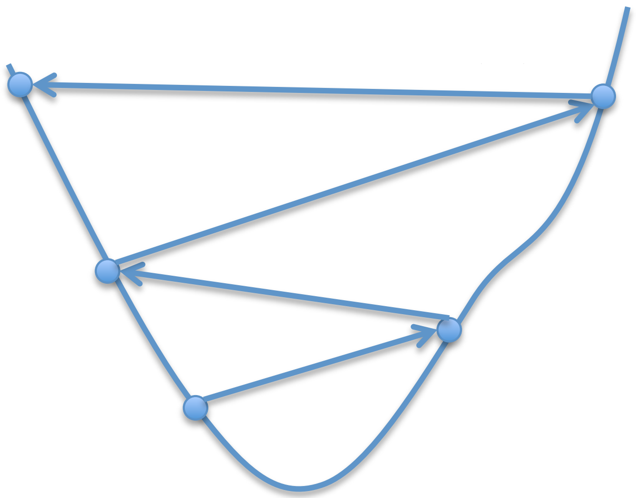
$$= \frac{1}{n} \sum_{i=1}^{n} (\sum_{k=0}^{d} \theta_k x_k^{(i)} - y^{(i)})\, x_j^{(i)}$$

CHOOSE α

Too small



Too large



- May overshoot the minimum
- May fail to converge
- May even diverge

QUESTIONS?