APPLIED MACHINE LEARNING

BY
MESFIN DIRO

July 30, 2021

Linear Regression

- Regression refers to a set of methods for modeling the relationship between one or more independent variables and a dependent variable.
- Regression problems pop up whenever we want to predict a numerical value.
- Linear regression is mainly used to predict: the issue of the presence of a linear relationship dependent variable and independent variable
- For example, buy a house problem: house prices are determined by the number and size of the house room, and this can be roughly linear regression to predict prices.

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

Linear Regression

- With a training set defined:
  - Take training set
  - Pass into al learning algorithm
  - algorithm outputs functions(h= hypothesis)
    - this function takes an input
    - tires to output the estimated value of Y
- How do we represent hypothesis h?
  - the h is represented as : $h_\theta(x) = \theta_0 + \theta_1 x$

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

Linear Regression

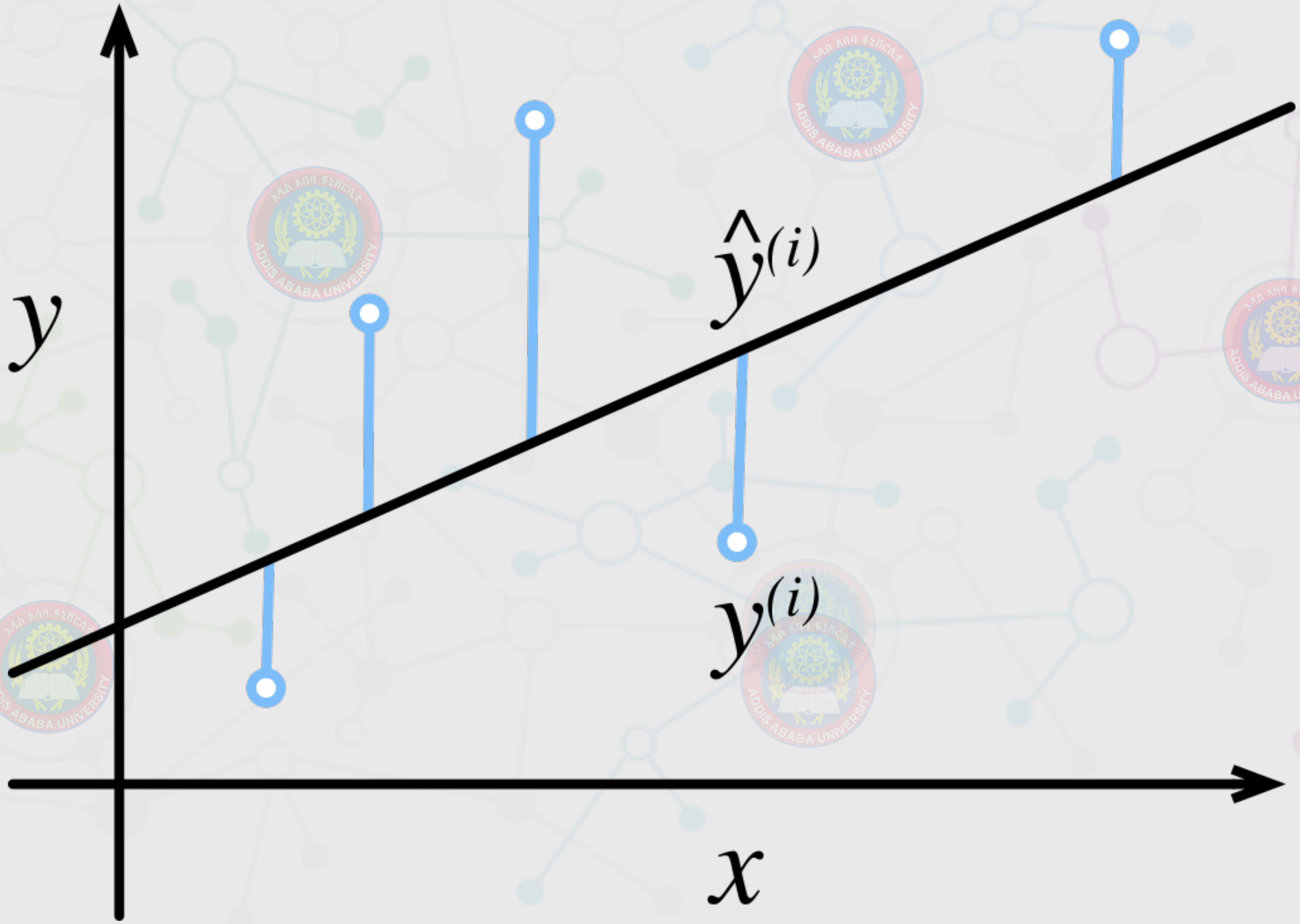
$$h_\theta(x) = \theta_0 + \theta_1 x = \theta^T x$$

- ‘x' is the **independent variable** on which the hypothesis depends
- ‘**theta 0**’ is our **bias** variable.
- ‘**theta 1**’ is our **weight** variable.
- theta 0 and theta 1 together

Loss Function

- Before we start thinking about how to fit data with our model, we need to determine a measure of fitness.
- The loss function quantifies the distance between the real and predicted value of the target.
- The loss will usually be a non-negative number where smaller values are better and perfect predictions incur a loss of zero.
- The most popular loss function in regression problems is the squared error.

Loss Function



Loss Function

- When the for prediction i is $\hat{y}^{(i)}$ and the corresponding true label is $y^{(i)}$ the squared error is given by:

$$j^{(i)}(\theta) = \frac{1}{2}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

The constant $\frac{1}{2}$ makes no real difference but will prove notationally convenient, canceling out when we take the derivative of the loss.

- A typical **Mean Squared Error** cost function looks like this:

Loss Function: OLS

- A typical **Mean Squared Error** cost function looks like this:

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

- Here:
  - **'J'** is our Cost Function.
  - 'm' is the number of data points in our data set.
  - **'y'** is the actual value (the value our data tells us) of our observation.

Loss Function

- Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$
- Parameters: θ_0, θ_1
- Cost Function: $J^{(i)}(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}) = \dfrac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^i) - y^{(i)}\right)^2$
- Goal: $\underset{\theta_0,\theta_1}{miniminize} J(\theta_0, \theta_1)$

Gradient Desent

- A **Gradient Descent** is a method by which we shall minimize the loss(Cost function)
- It is an optimization function that changes the values of theta 0 and theta 1 based on the slope of the cost function curve at that point.
- It is an optimization function that changes the values of theta 0 and theta 1 based on the slope of the cost function curve at that point.

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

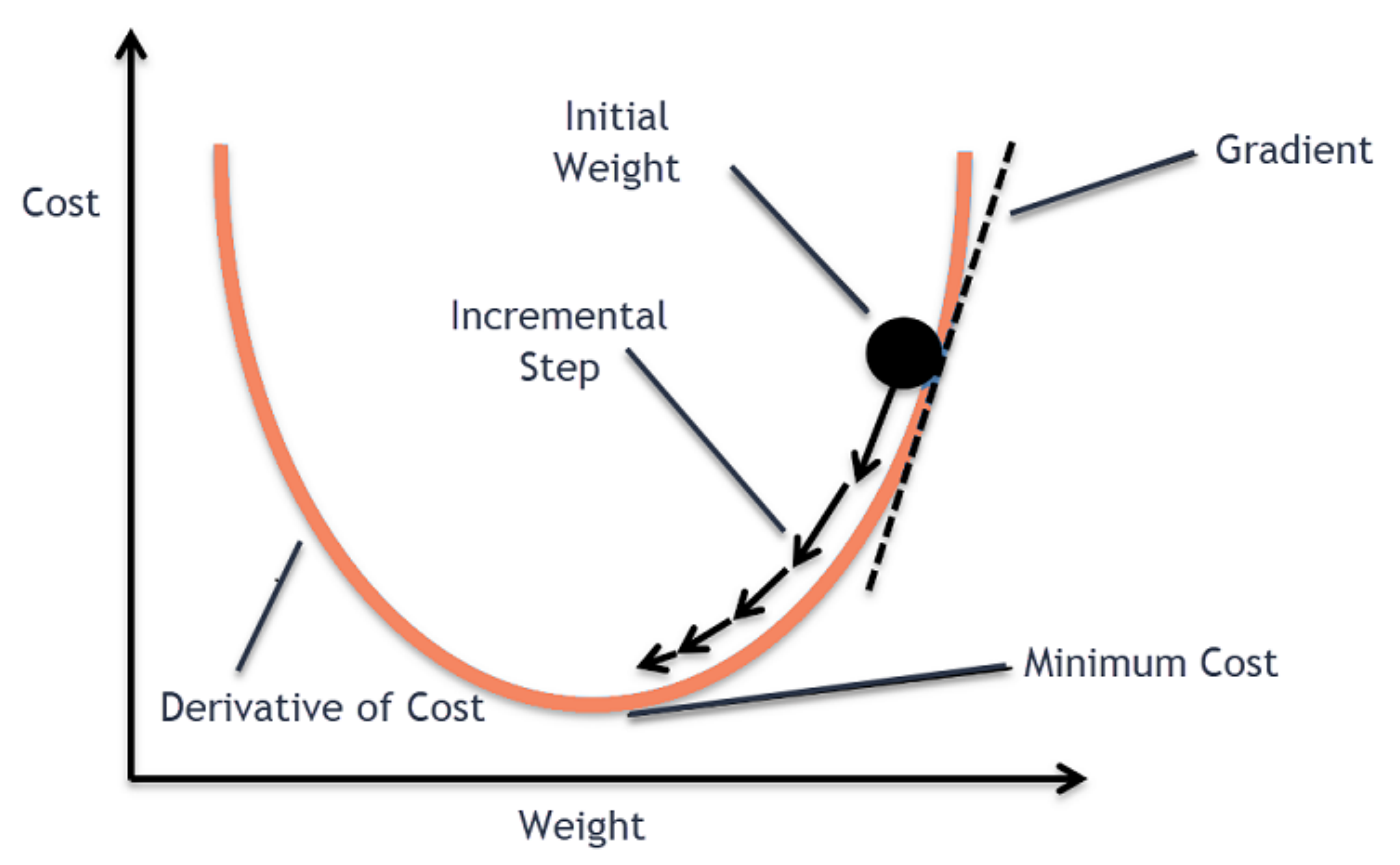
- Here:
  - **'J'** is our Cost Function.
  - 'm' is the number of data points in our data set.
  - **'y'** is the actual value (the value our data tells us) of our observation.
  - α is learning rate

Gradient Desent

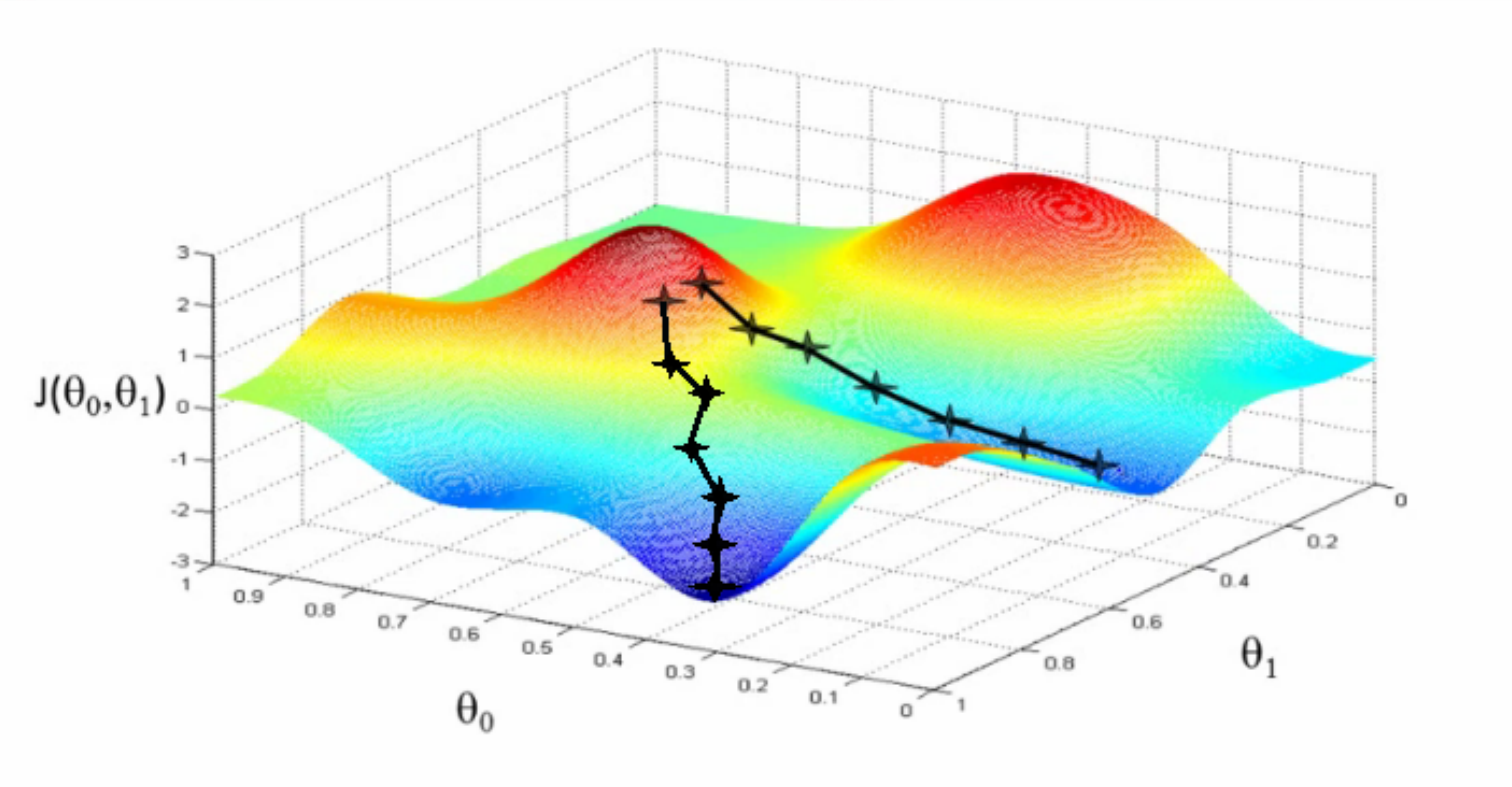
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_i} J(\theta_0, \theta_1)$$

$$:= \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} [(h_\theta(x_i) - y_i)] x_i$$

Gradient Desent



Gradient Desent



Gradient Desent