Data Science

Dr. Muhammad Adeel Nisar

Assistant Professor – Department of IT,
Faculty of Computing and Information Technology,
University of the Punjab, Lahore

Recap of the Last Lecture

- Univariate Linear Regression

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

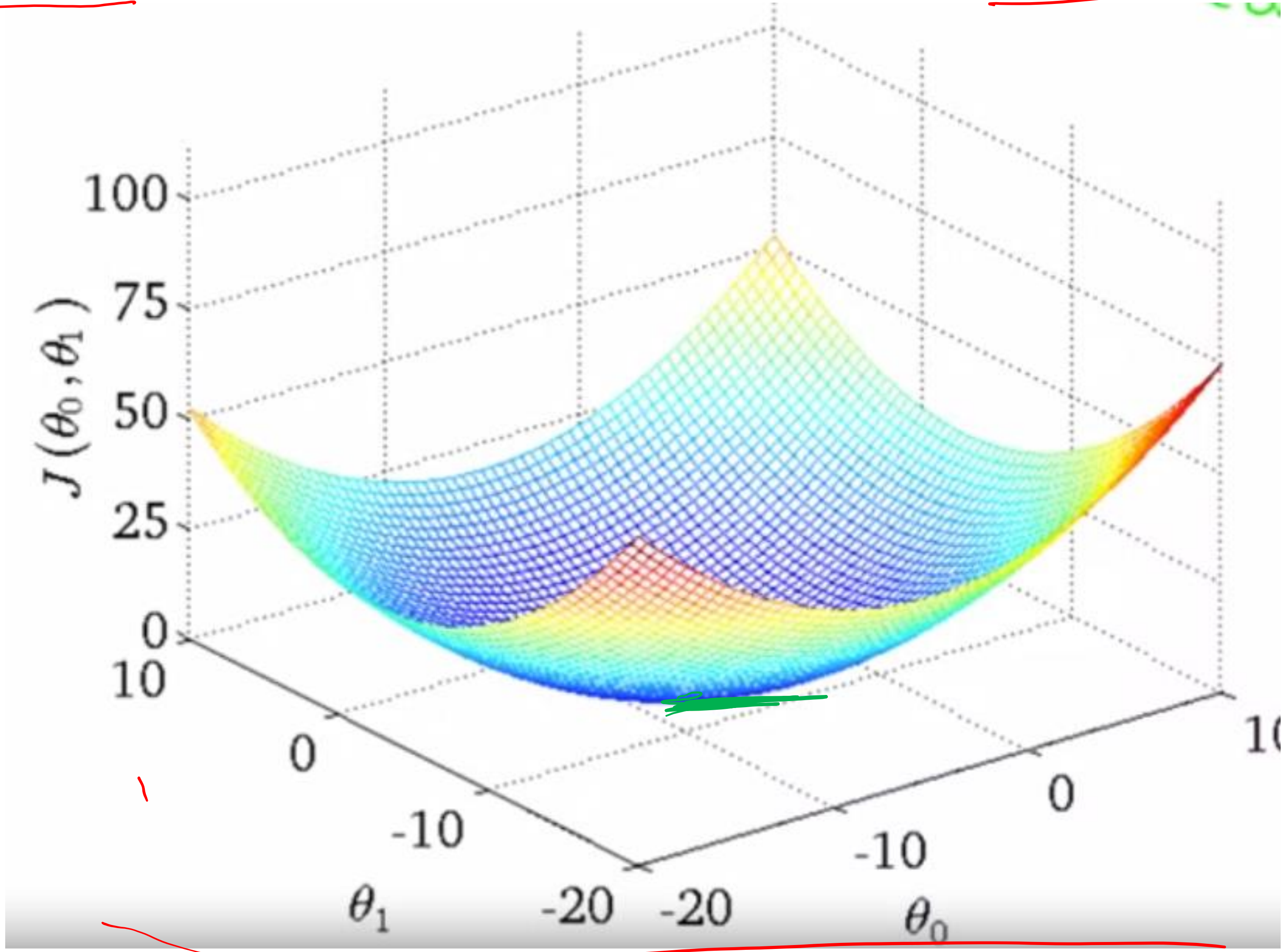
- Cost/Loss function, Mean Squared Error

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

Graph of $J(\vartheta_0, \vartheta_1)$

Convex

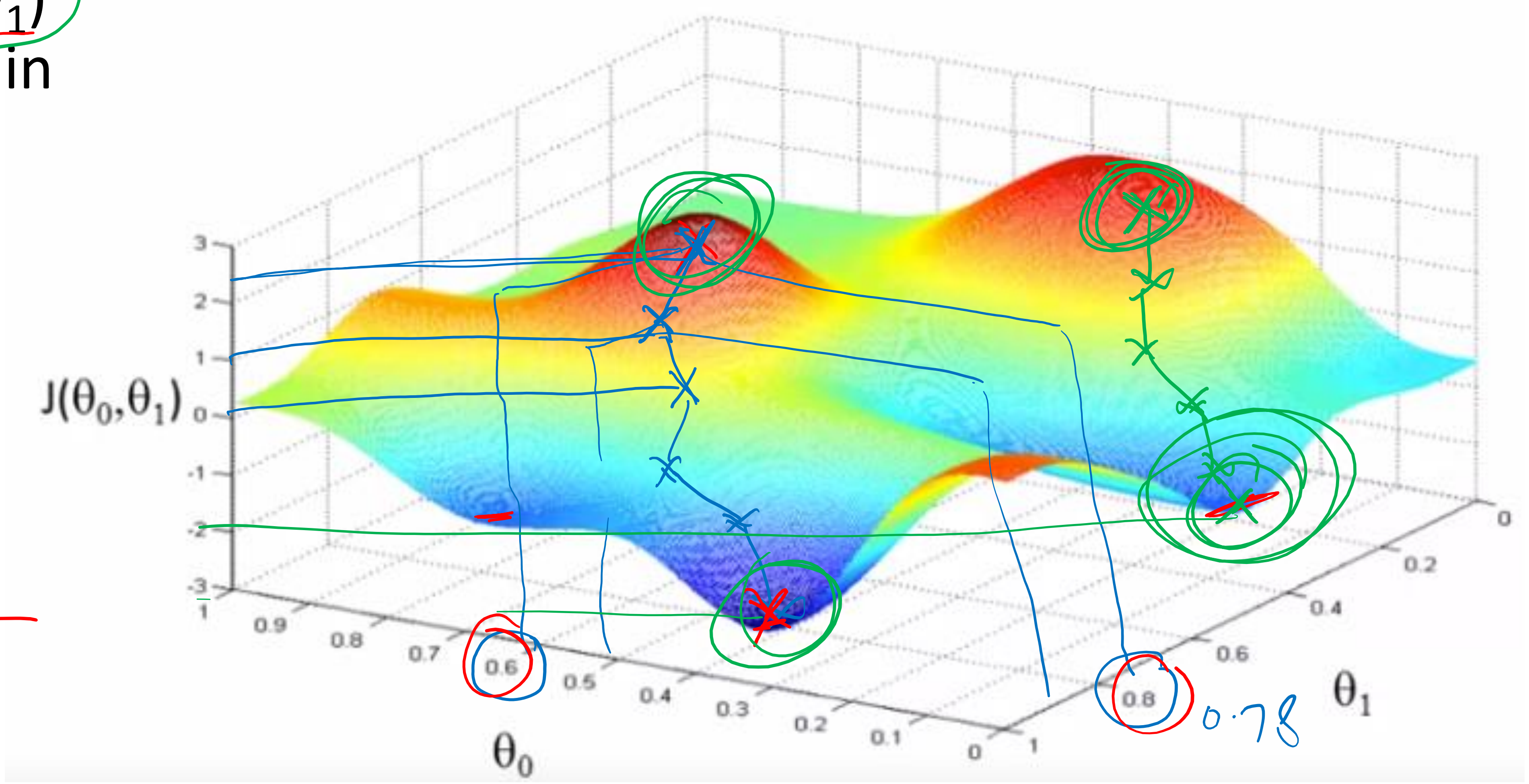
cost



Gradient Descent Algorithm

- We have $J(\vartheta_0, \vartheta_1)$ and we want min $J(\vartheta_0, \vartheta_1)$

cost



Gradient descent algorithm

repeat until convergence {

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

}

(simultaneously update $j = 0$ and $j = 1$)

α: learning rate

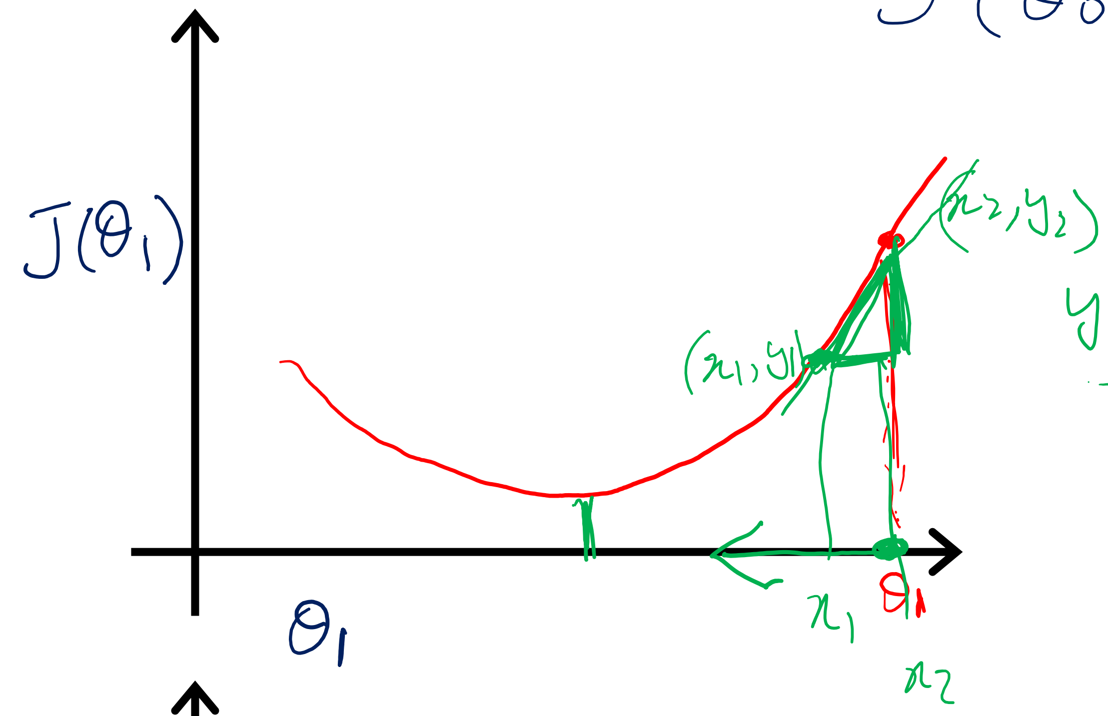
$temp_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$

$temp_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$

$\theta_0 := temp_0$

$\theta_1 := temp_1$

$J(\theta_0, \theta_1)$, $\theta_0 = 0$ $J(\theta_1)$

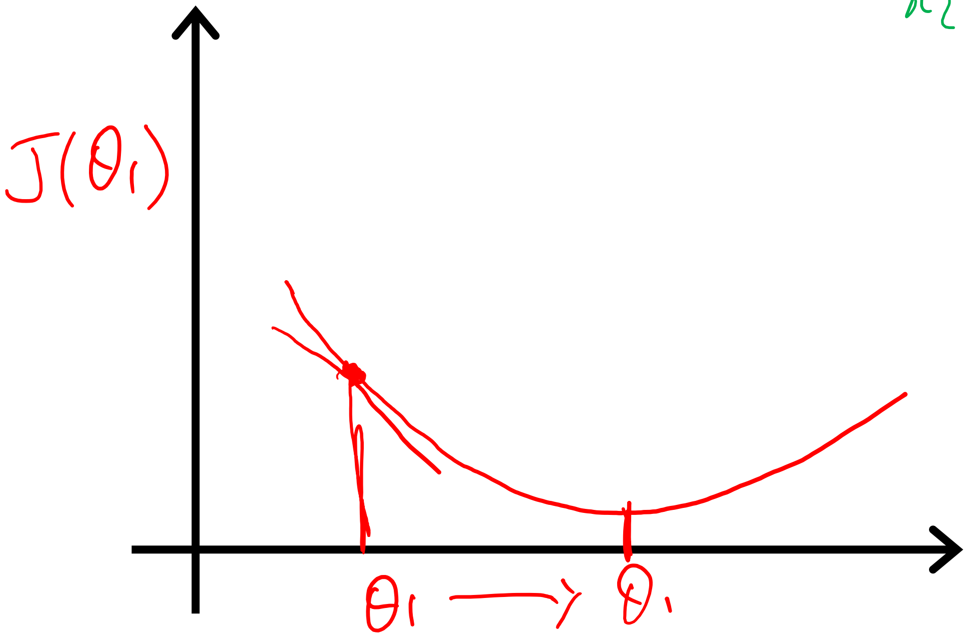


$$\theta_1 := \theta_1 - \boxed{\alpha} \boxed{\frac{d}{d\theta_1}(J(\theta_1))}$$

$$\frac{y_2 - y_1}{x_2 - x_1}$$

$$\theta_1 := \theta_1 - \alpha(+ve\ S)$$

$$\theta_1 := \theta_1 - S$$



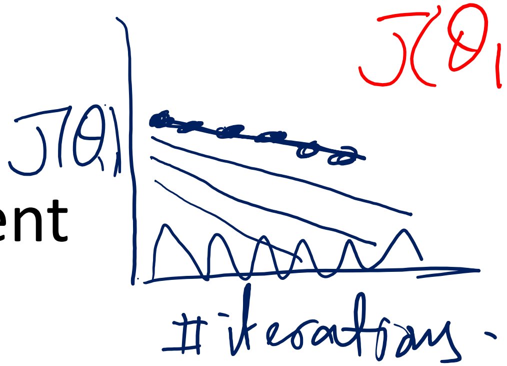
$$\theta_1 := \theta_1 - \alpha \underbrace{\frac{d}{d\theta_1}(J(\theta_1))}_{-ve}$$

$\theta_1 \uparrow$

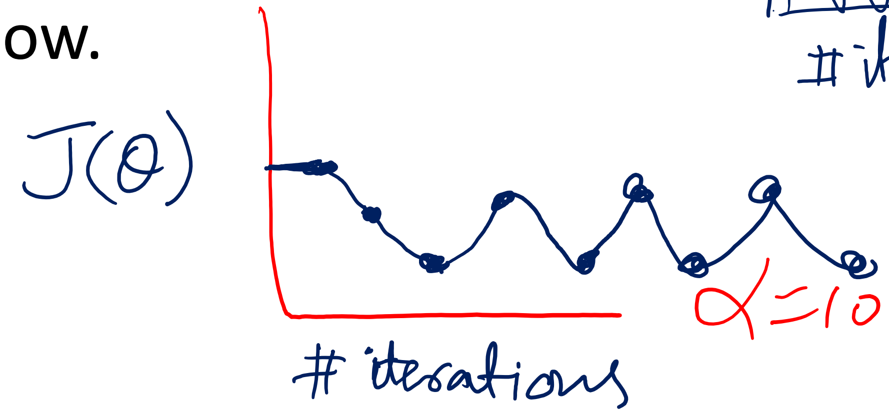
0.0001

0.01 -0.2

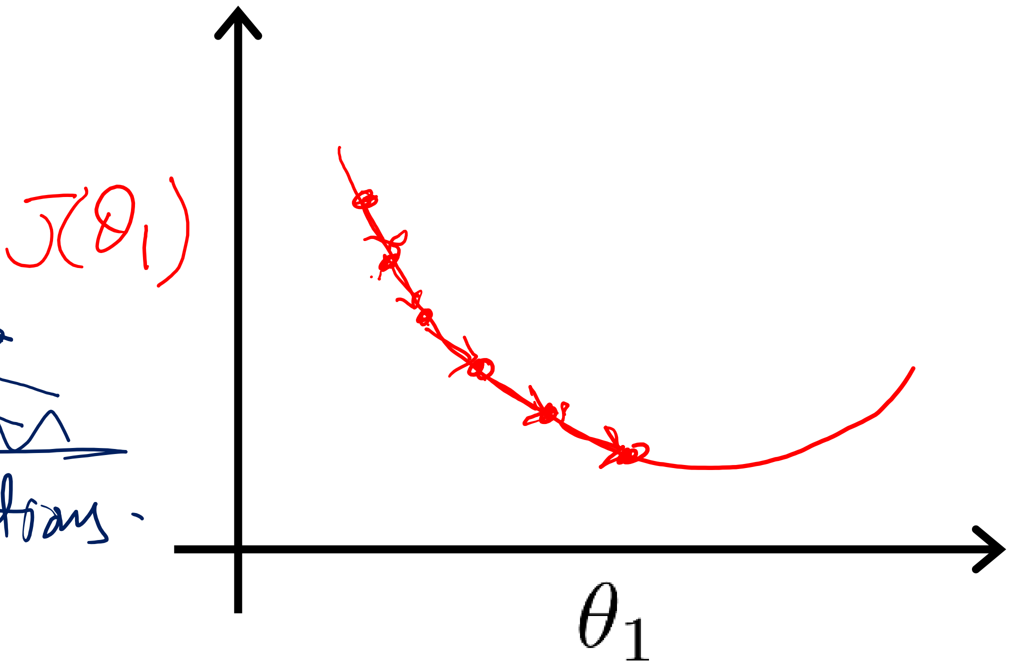
$$\rightarrow \quad \theta_1 := \theta_1 - \alpha \boxed{\frac{\partial}{\partial\theta_1} J(\theta_1)}$$

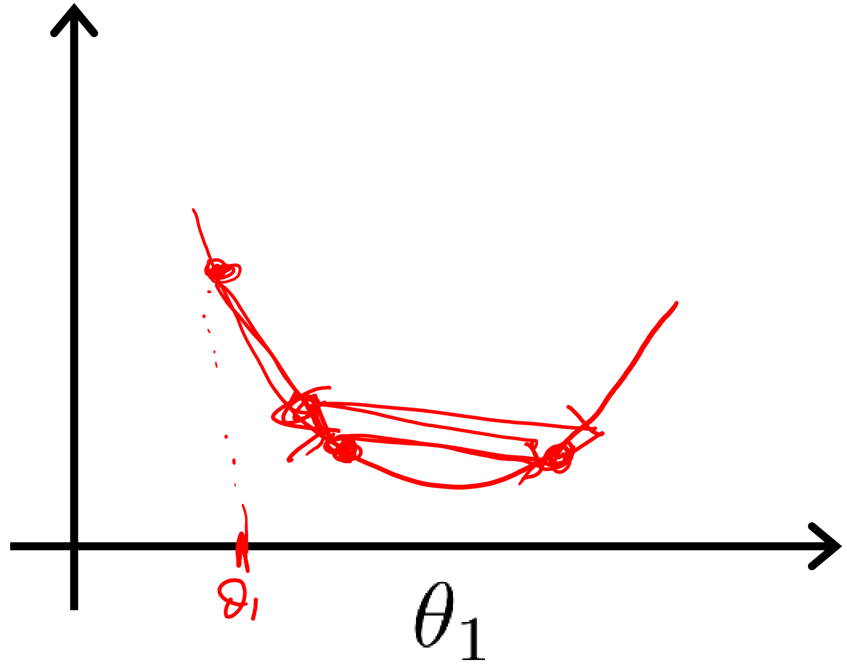


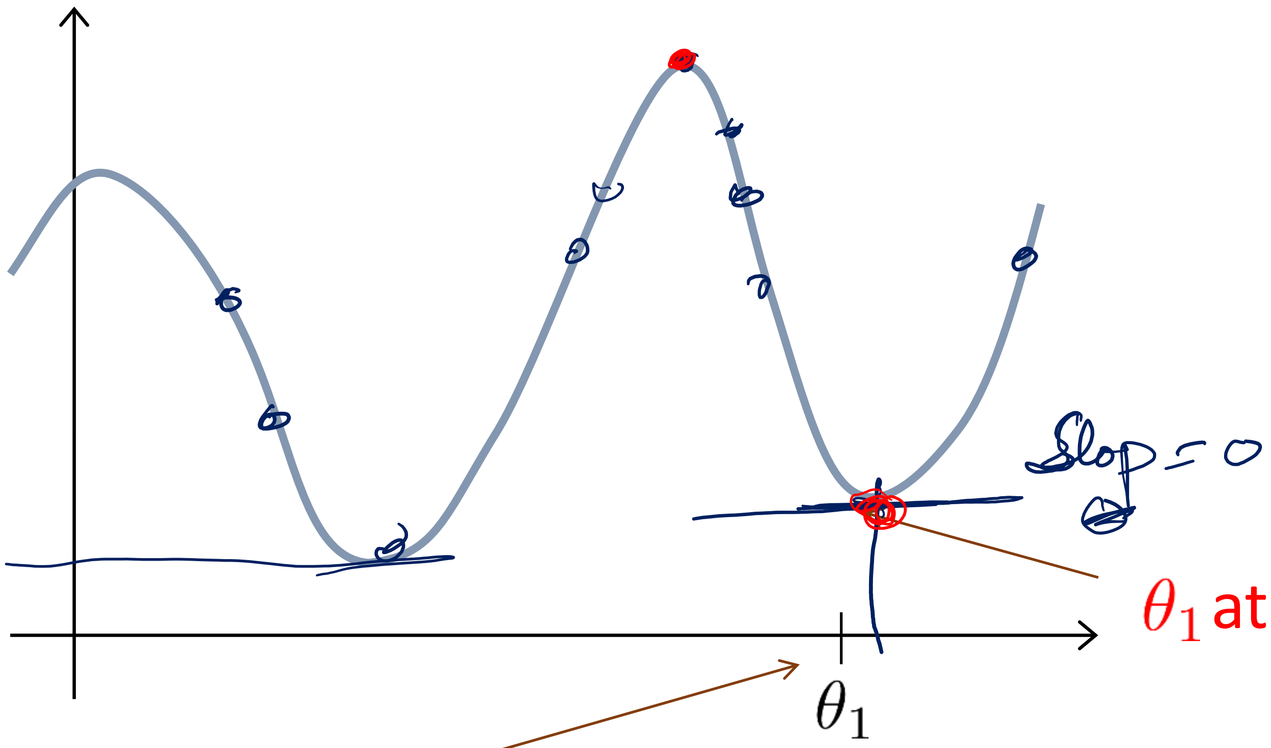
If α is too small, gradient descent can be slow.



If α is too large, gradient descent can overshoot the minimum. It may fail to converge, or even diverge.







θ_1 at local optima

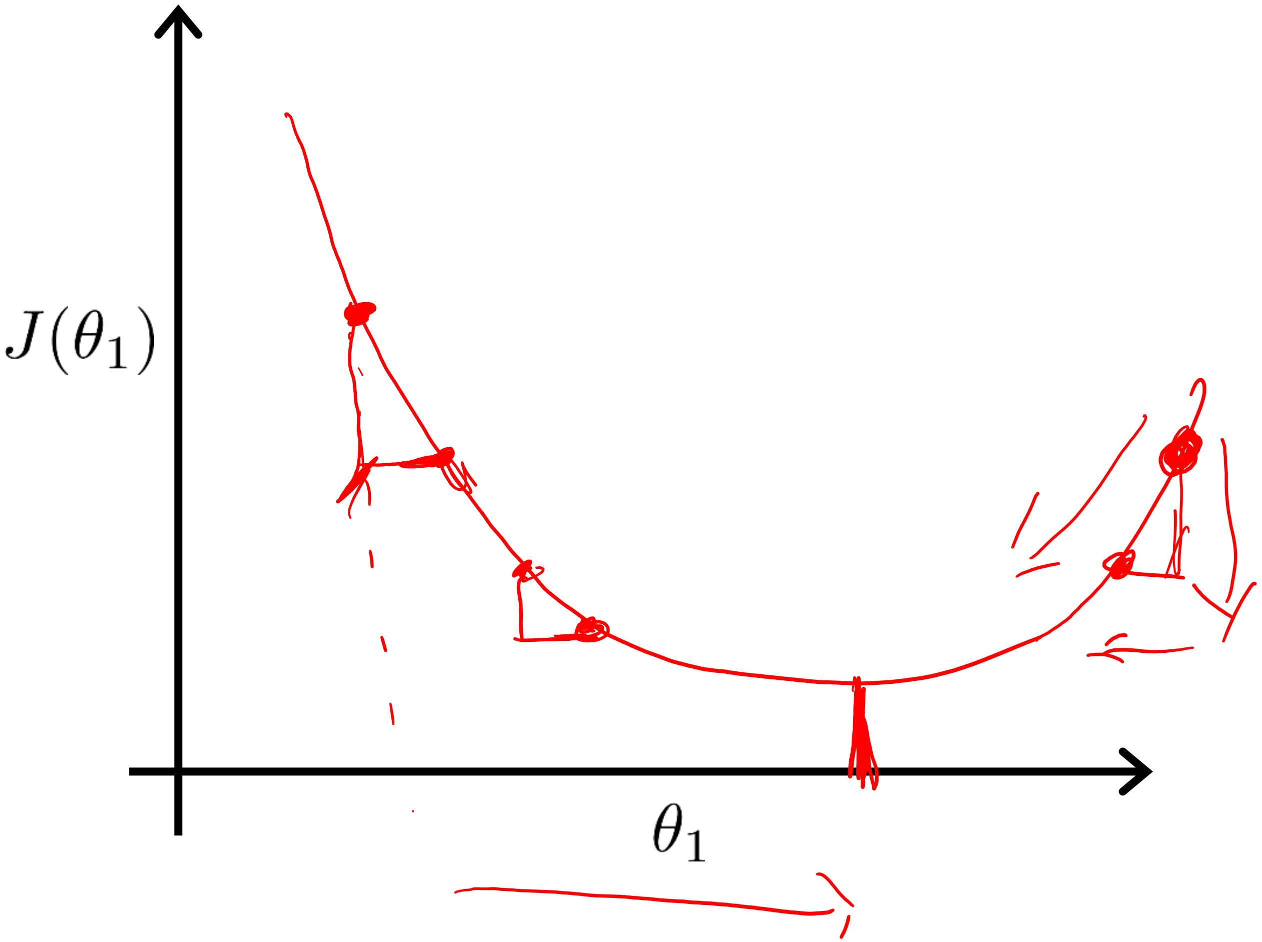
Current value of θ_1

$$\underline{\theta_1} := \underline{\theta_1} - \alpha \cancel{\frac{d}{d\theta_1}} J(\theta_1)$$

Gradient descent can converge to a local minimum, even with the learning rate α fixed.

$$\theta_1 := \theta_1 - \alpha \frac{d}{d\theta_1} J(\theta_1)$$

As we approach a local minimum, gradient descent will automatically take smaller steps. So, no need to decrease α over time.



Derivatives

$f(x) = 4x$ $= 4$

$f(x) = x^3$ $= 3x^2$

$f(x) = (x + 2)^4$ $= 4(x+2)^3 \cdot \cancel{\tfrac{d}{dx}(x+2)}$

$f(x,y) = (3x + 2y + 2)^2$ $= \dfrac{\partial f(x,y)}{\partial x} = 2(3x+2y+2)\cdot 3 = 6(3x+ \ldots)$

$= 4(3x+2y+2)$

Gradient Descent Algorithm

- $\vartheta_j := \vartheta_j - \alpha \frac{\partial}{\partial \vartheta_j} J(\vartheta_0, \vartheta_1)$ (for j = 0 and j = 1)
- $\frac{\partial}{\partial \vartheta_j} J(\vartheta_0, \vartheta_1)$ is a partial derivative term
- α : (Alpha) is learning rate
- Simultaneous Update
- temp0 = $\vartheta_0 - \alpha \frac{\partial}{\partial \vartheta_0} J(\vartheta_0, \vartheta_1)$
- temp1 = $\vartheta_1 - \alpha \frac{\partial}{\partial \vartheta_1} J(\vartheta_0, \vartheta_1)$ ✓
- ϑ_0 := temp0
- ϑ_1 := temp1

Linear Regression with Gradient Descent

$$\vartheta_j := \vartheta_j - \alpha \frac{\partial}{\partial \vartheta_j} J(\vartheta_0, \vartheta_1)$$

$$\frac{\partial}{\partial \vartheta_j} J(\vartheta_0, \vartheta_1) = \frac{\partial}{\partial \vartheta_j} \left(\frac{1}{2m} \sum_{i=1}^{m} (\boxed{h\vartheta(x^{(i)})} - y^{(i)})^2 \right)$$

$$\frac{\partial}{\partial \vartheta_j} J(\vartheta_0, \vartheta_1) = \frac{\partial}{\partial \vartheta_j} \left(\frac{1}{2m} \sum_{i=1}^{m} (\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)})^2 \right)$$

Linear Regression with Gradient Descent

$$\frac{\partial}{\partial \vartheta_j} J(\vartheta_0, \vartheta_1) = \frac{\partial}{\partial \vartheta_j} \left(\frac{1}{2m} \sum_{i=1}^{m} \left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)} \right)^2\right)$$

$$\frac{\partial}{\partial \vartheta_0} J(\vartheta_0, \vartheta_1) = \frac{\partial}{\partial \vartheta_0} \left(\frac{1}{2m} \sum_{i=1}^{m} \left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)} \right)^2\right)$$

$$\frac{\partial}{\partial \vartheta_0} J(\vartheta_0, \vartheta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)} \right)$$

$$\frac{\partial}{\partial \vartheta_1} J(\vartheta_0, \vartheta_1) = \frac{\partial}{\partial \vartheta_1} \left(\frac{1}{2m} \sum_{i=1}^{m} \left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)} \right)^2\right)$$

$$\frac{\partial}{\partial \vartheta_1} J(\vartheta_0, \vartheta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(\left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)} \right) x^{(i)}\right)$$

Linear Regression with Gradient Descent

θ_0, θ_1

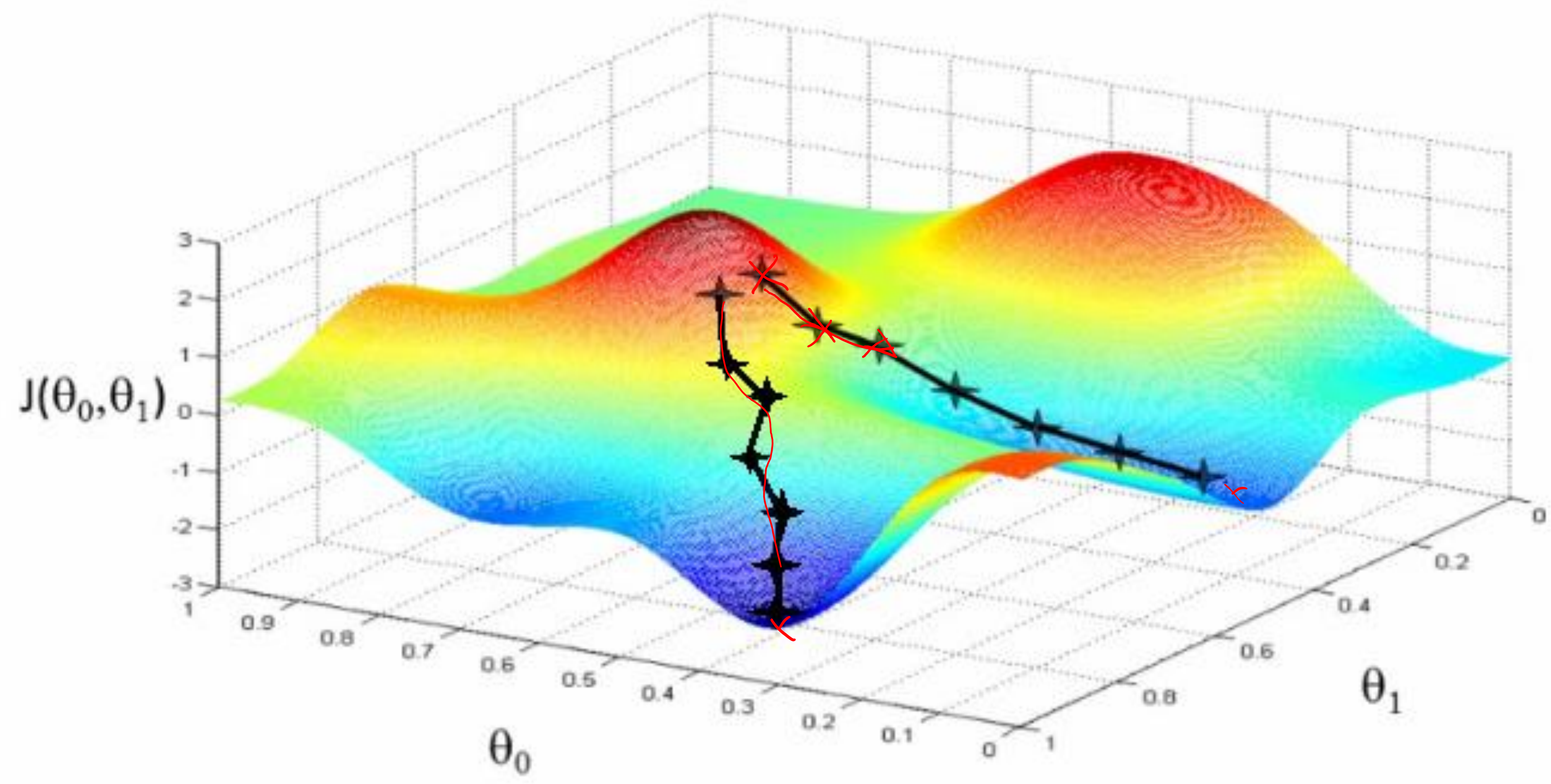
- Repeat until **converge**

$$\vartheta_0 := \vartheta_0 - \alpha \left(\frac{1}{m}\sum_{i=1}^{m}\left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)}\right)\right) \checkmark$$

$$\vartheta_1 := \vartheta_1 - \alpha \left(\frac{1}{m}\sum_{i=1}^{m}\left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)}\right) x^{(i)}\right) \checkmark$$

- Simultaneous update

Solving Minimization Problem



Review of Univariate Linear Regression

- Univariate Linear Regression

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

- Cost/Loss function, Mean Squared Error

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

- Gradient Descent Algorithm

$$\vartheta_j := \vartheta_j - \alpha \frac{\partial}{\partial \vartheta_j} J(\vartheta_0, \vartheta_1) \quad \text{(for } j = 0 \text{ and } j = 1\text{)}$$

Linear Regression with Multiple Features

Size of Plot x_1	Locality Value x_2	Facing Park x_3	Distance from School x_4	Price y	
5	1.0	1	2	15	$\mathbf{x}^{(1)}$
10	0.9	0	2.5	25	
7	1.5	1	1.9	35	
…	…	…	…	…	
4	0.5	0	10	5	$\mathbf{x}^{(m)}$

n

Multivariate Linear Regression

- **Hypothesis Function (Uni-variate)**

$$y' = h_{\vartheta}(x) = \vartheta_0 + \vartheta_1 x_1$$

- **Hypothesis Function (multivariate)**

$$y' = h_{\vartheta}(x) = \vartheta_0 + \vartheta_1 x_1 + \vartheta_2 x_2 + \ldots + \vartheta_n x_n$$

where, x_0 x_j *is jth feature*

n is the number of features

x —

$$y' = h_{\vartheta}(x) = \vartheta_0 x_0 + \vartheta_1 x_1 + \vartheta_2 x_2 + \ldots + \vartheta_n x_n \text{ where } x_0 = 1$$

$$\boldsymbol{\theta}^T = [\vartheta_0, \vartheta_1, \vartheta_2, \ldots, \vartheta_n]^T, \qquad \mathbf{x} = [x_0, x_1, x_2, \ldots, x_n]^T$$

$$y' = h_{\vartheta}(x) = \boldsymbol{\theta}^T \mathbf{x}$$

Gradient Descent For Multivariate Linear Regression

- **Hypothesis Function**

  $y' = h_\vartheta(x) = \boldsymbol{\theta}^T\mathbf{x} = \vartheta_0 x_0 + \vartheta_1 x_1 + \vartheta_2 x_2 + \ldots + \vartheta_n x_n$ *where* $x_0 = 1$

  Parameters $= \vartheta_0, \vartheta_1, \vartheta_2, \ldots, \vartheta_n = \boldsymbol{\theta}$ *n+1 feature vector*

- **Cost Function**

$$J(\vartheta_0, \vartheta_1, \vartheta_2, \ldots, \vartheta_n) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

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

- **Gradient Descent**

  Repeat until convergence:

$$\vartheta_j := \vartheta_j - \alpha \frac{\partial}{\partial \vartheta_j} J(\vartheta_0, \vartheta_1, \ldots, \vartheta_n) = \vartheta_j - \alpha \frac{\partial}{\partial \vartheta_j} J(\boldsymbol{\theta})$$

Simultaneous update for each *j = 0, 1, 2, …, n*

Gradient Descent for Multivariate Regression

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

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

$$= \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x) - y^{(i)}\right) x_0^{(i)}$$

$$\frac{\partial J(\theta)}{\partial \theta_1} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x) - y^{(i)}\right) x_1^{(i)}$$

$$\frac{\partial J(\theta)}{\partial \theta_2} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x) - y^{(i)}\right) x_2^{(i)}$$

Gradient Descent for Multivariate Regression

- Repeat until **converge** (for n = 1) {

$$\vartheta_0 := \vartheta_0 - \alpha \left(\frac{1}{m}\sum_{i=1}^{m}(h_{\vartheta}(x^{(i)}) - y^{(i)})\right)$$

$$\vartheta_1 := \vartheta_1 - \alpha \left(\frac{1}{m}\sum_{i=1}^{m}(h_{\vartheta}(x^{(i)}) - y^{(i)})) \quad x^{(i)}\right)$$

}

- Repeat until **converge** (for n >= 1) {

$$\vartheta_j := \vartheta_j - \alpha \left(\frac{1}{m}\sum_{i=1}^{m}(h_{\vartheta}(x^{(i)}) - y^{(i)})) \quad x_j^{(i)}\right)$$

$$\vartheta_0 := \vartheta_0 - \alpha \left(\frac{1}{m}\sum_{i=1}^{m}(h_{\vartheta}(x^{(i)}) - y^{(i)})) \quad x_0^{(i)}\right)$$

$$\vartheta_1 := \vartheta_1 - \alpha \left(\frac{1}{m}\sum_{i=1}^{m}(h_{\vartheta}(x^{(i)}) - y^{(i)})) \quad x_1^{(i)}\right)$$

}

Feature Scaling

- Gradient Descent:

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

- Different Strategies
- Specific Range: -1 <= x <= +1
- Mean Normalization: (x – mean) / max or (x – mean) / (max-min)

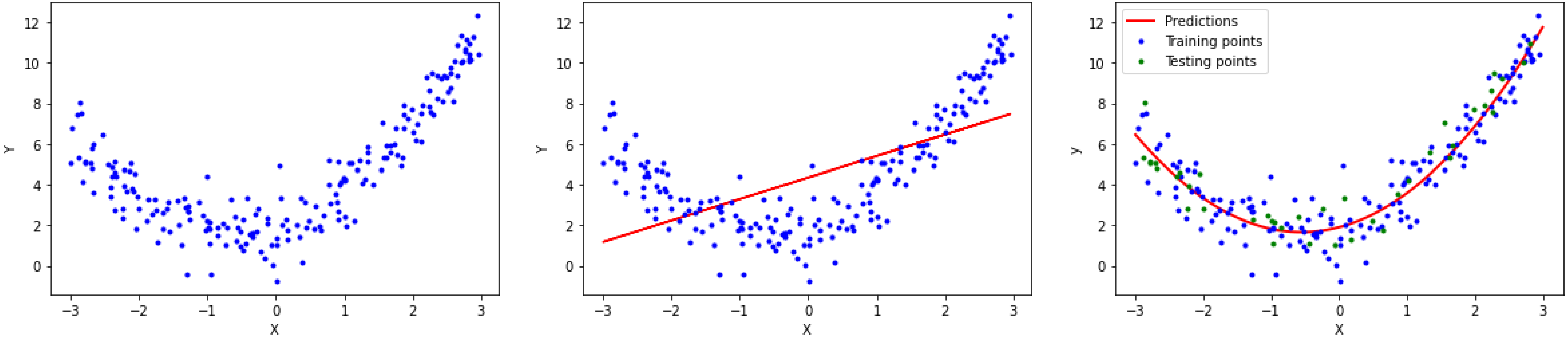
  -0.5 <= x <= +0.5

Selection of Learning Rate

- When Gradient Descent works properly then the cost $J(\boldsymbol{\theta})$ should decrease after every iteration.
- A model is assumed to be converged if it decreases the cost less than a threshold value in subsequent iterations.
- Gradient Descent doesn’t work properly if cost increases or fluctuates in subsequent iterations. Solution: try a smaller learning rate.
- If learning rate is too small: slow convergence
- If learning rate is too large: Gradient descent might not converge
- Solution: Try a range of values for the learning rate and then pick the best

Polynomial Regression

- If the relationship between data is not linear:



- $y' = h_{\vartheta}(x) = \vartheta_0 + \vartheta_1 x_1 + \vartheta_2 x_2^{\,2} + \vartheta_3 x_3^{\,3}$

Normal Eqautions

- Method to solve for θ analytically
- $J(\vartheta) = a\vartheta^2 + b\vartheta + c$
- $\frac{d}{d(\vartheta)} J(\vartheta) = 0$

Linear Regression with One Variable

The Hypothesis Function / Model

- $y' = h_\vartheta(x) = \vartheta_0 + \vartheta_1 x$
- $y'^{(i)} = h_\vartheta(x^{(i)}) = \vartheta_0 + \vartheta_1 x^{(i)}$

Cost function:

$$J(\vartheta_0, \vartheta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(y'^{(i)} - y^{(i)} \right)^2$$

Objective:

$$\min_{\vartheta_0, \vartheta_1} J(\vartheta_0, \vartheta_1)$$

$$\frac{\partial}{\partial \vartheta_0} J(\vartheta_0, \vartheta_1) = 0$$

$$\frac{\partial}{\partial \vartheta_1} J(\vartheta_0, \vartheta_1) = 0$$

$$\frac{\partial}{\partial \vartheta_0} J(\vartheta_0, \vartheta_1) = \frac{\partial}{\partial \vartheta_0}\left(\frac{1}{2m}\sum_{i=1}^{m}\left(y'^{(i)} - y^{(i)} \right)^2\right)$$

$$= \frac{\partial}{\partial \vartheta_0}\left(\frac{1}{2m}\sum_{i=1}^{m}\left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)} \right)^2\right)$$

$$= \frac{2}{2m}\sum_{i=1}^{m}\left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)}\right) \frac{\partial}{\partial \vartheta_0}\left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)}\right)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)}\right)$$

$$\frac{1}{m}\sum_{i=1}^{m}\left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)}\right) = 0$$

$$\Rightarrow \sum_{i=1}^{m}\left(\vartheta_0 + \vartheta_1 x^{(i)} - y^{(i)}\right) = 0$$

$$\Rightarrow \sum_{i=1}^{m}\vartheta_0 + \sum_{i=1}^{m}\vartheta_1 x^{(i)} - \sum_{i=1}^{m} y^{(i)} = 0$$

$$\Rightarrow \sum_{i=1}^{m}\vartheta_0 + \sum_{i=1}^{m}\vartheta_1 x^{(i)} = \sum_{i=1}^{m} y^{(i)} \quad \text{---------------> A}$$

$$\sum_{i=1}^{m}\vartheta_0 x^{(i)} + \sum_{i=1}^{m}\vartheta_1 x^{2(i)} = \sum_{i=1}^{m} y^{(i)} x^{(i)} \quad \text{-----------> B}$$

Linear Regression using simultaneous equation

- $\sum_{i=1}^{m} \vartheta_0 + \sum_{i=1}^{m} \vartheta_1 x^{(i)} = \sum_{i=1}^{m} y^{(i)}$ -----------------------> A
- $\sum_{i=1}^{m} \vartheta_0 x^{(i)} + \sum_{i=1}^{m} \vartheta_1 x^{2(i)} = \sum_{i=1}^{m} y^{(i)} x^{(i)}$ -----------------------> B

8x + 2y = 46
7x + 3y = 47

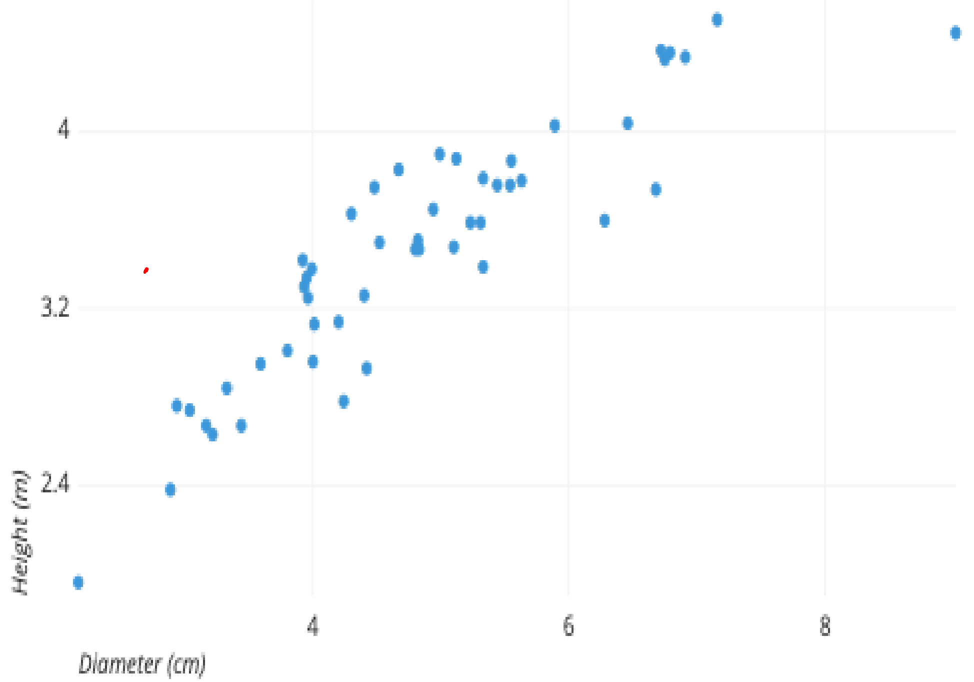
- $\begin{bmatrix} \sum_{i=1}^{m} 1 & \sum_{i=1}^{m} x^{(i)} \\ \sum_{i=1}^{m} x^{(i)} & \sum_{i=1}^{m} x^{2(i)} \end{bmatrix} \begin{bmatrix} \vartheta_0 \\ \vartheta_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y^{(i)} \\ \sum_{i=1}^{m} y^{(i)} x^{(i)} \end{bmatrix}$
- $\begin{bmatrix} m & \sum_{i=1}^{m} x^{(i)} \\ \sum_{i=1}^{m} x^{(i)} & \sum_{i=1}^{m} x^{2(i)} \end{bmatrix} \begin{bmatrix} \vartheta_0 \\ \vartheta_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y^{(i)} \\ \sum_{i=1}^{m} y^{(i)} x^{(i)} \end{bmatrix}$

$$\begin{array}{ccc} A & X & B \\ \begin{bmatrix} 8 & 2 \\ 7 & 3 \end{bmatrix} & \begin{bmatrix} x \\ y \end{bmatrix} = & \begin{bmatrix} 46 \\ 47 \end{bmatrix} \end{array}$$

A X B

$$\begin{bmatrix} m & \sum_{i=1}^{m} x^{(i)} \\ \sum_{i=1}^{m} x^{(i)} & \sum_{i=1}^{m} x^{2(i)} \end{bmatrix} \begin{bmatrix} \vartheta_0 \\ \vartheta_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y^{(i)} \\ \sum_{i=1}^{m} y^{(i)} \; x^{(i)} \end{bmatrix}$$

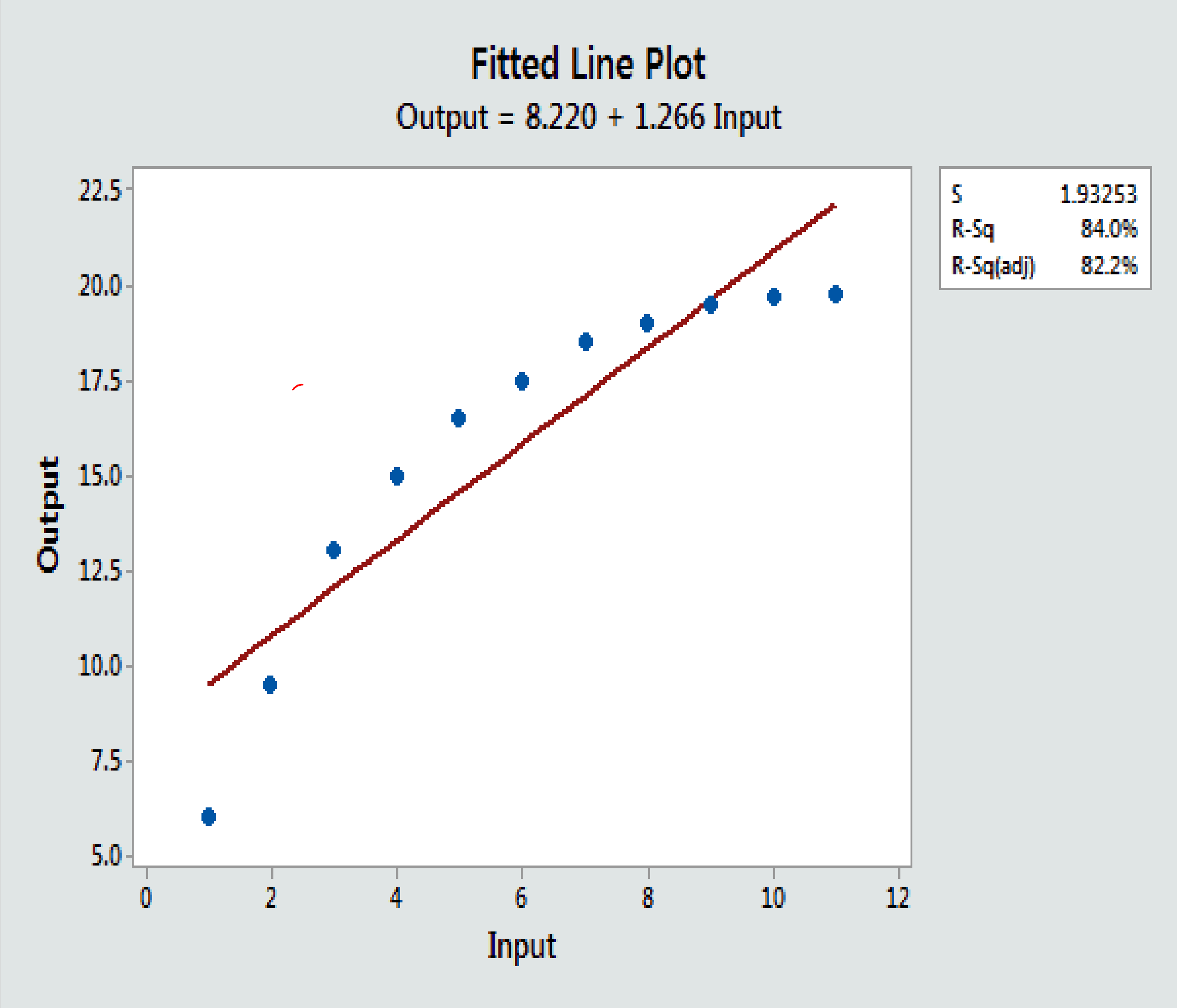
ht (x)	wt (y)
1	10
2	20
3	30
4	40

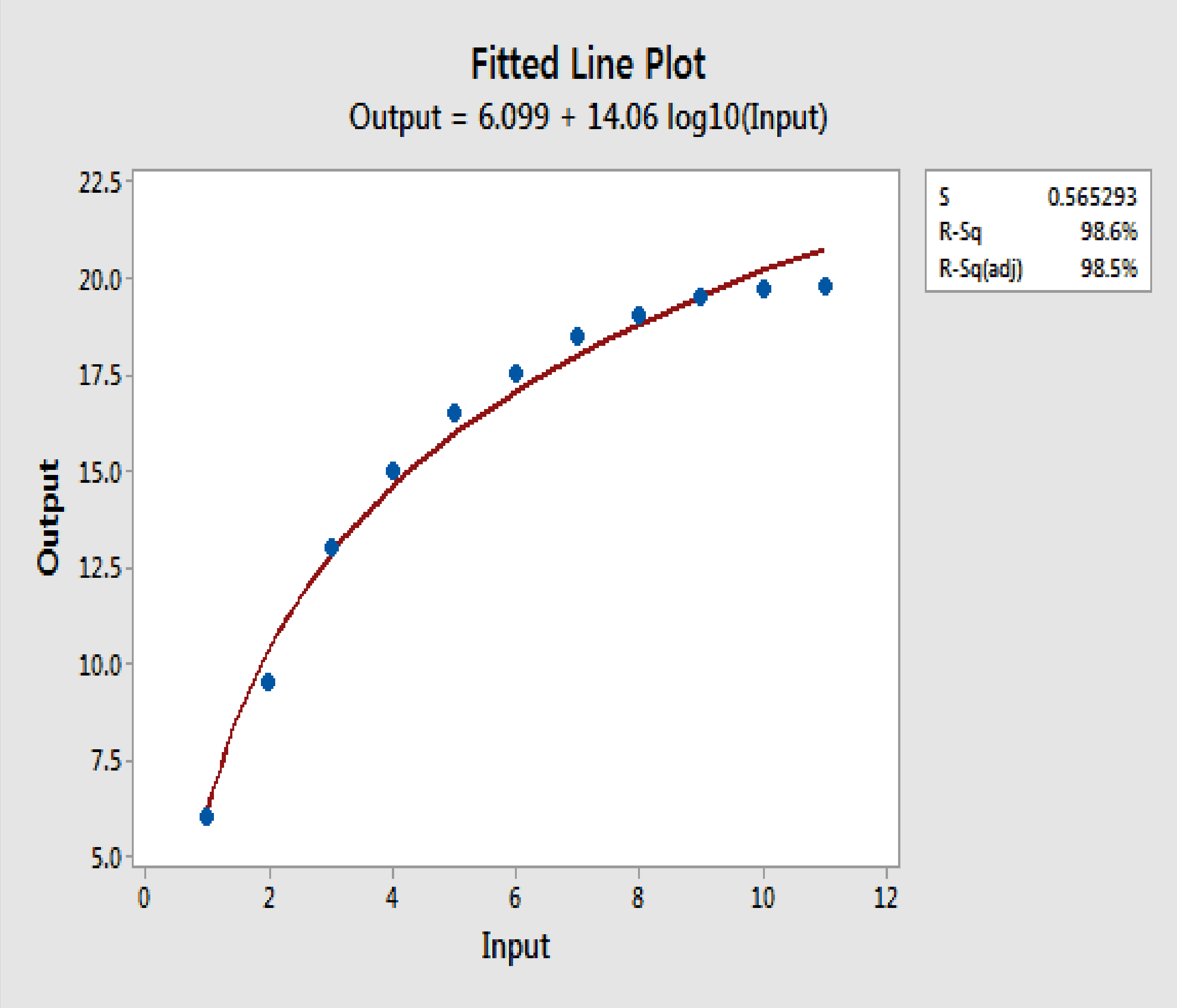


Comparison of Gradient Descent and Normal Equation

Gradient Descent	Normal Equation
• Need to choose learning rate	• No Need to choose learning rate
• Need many iterations	• No iterations required
• Works well for even large number of features	• Works slow if number of features is very large $O(n^3)$

Quadratic Model





Quadratic Model

The Hypothesis Function / Model

- $y' = h_\vartheta(x) = \vartheta_0 + \vartheta_1 x + \vartheta_2 x^2$
- $y'^{(i)} = h_\vartheta(x^{(i)}) = \vartheta_0 + \vartheta_1 x^{(i)} + \vartheta_2 x^{2(i)}$

Cost function:

$$J(\vartheta_0, \vartheta_1, \vartheta_2) = \frac{1}{2m} \sum_{i=1}^{m} (y'^{(i)} - y^{(i)})^2$$

Objective:

$$\min_{\vartheta_0, \vartheta_1, \vartheta_2} J(\vartheta_0, \vartheta_1, \vartheta_2)$$

$$\frac{\partial}{\partial \vartheta_0} J(\vartheta_0, \vartheta_1, \vartheta_2) = 0$$

$$\frac{\partial}{\partial \vartheta_1} J(\vartheta_0, \vartheta_1, \vartheta_2) = 0$$

$$\frac{\partial}{\partial \vartheta_2} J(\vartheta_0, \vartheta_1, \vartheta_2) = 0$$

$$\frac{\partial}{\partial \vartheta_0} J(\vartheta_0, \vartheta_1, \vartheta_2) = \frac{\partial}{\partial \vartheta_0} \left(\frac{1}{2m} \sum_{i=1}^{m} (y'^{(i)} - y^{(i)})^2\right)$$

$$= \frac{\partial}{\partial \vartheta_0} \left(\frac{1}{2m} \sum_{i=1}^{m} (\vartheta_0 + \vartheta_1 x^{(i)} + \vartheta_2 x^{2(i)} - y^{(i)})^2\right)$$

$$= \frac{2}{2m} \sum_{i=1}^{m} (\vartheta_0 + \vartheta_1 x^{(i)} + \vartheta_2 x^{2(i)} - y^{(i)}) \frac{\partial}{\partial \vartheta_0} (\vartheta_0 + \vartheta_1 x^{(i)} + \vartheta_2 x^{2(i)} - y^{(i)})$$

$$= \frac{1}{m} \sum_{i=1}^{m} (\vartheta_0 + \vartheta_1 x^{(i)} + \vartheta_2 x^{2(i)} - y^{(i)})$$

$$\frac{1}{m} \sum_{i=1}^{m} (\vartheta_0 + \vartheta_1 x^{(i)} + \vartheta_2 x^{2(i)} - y^{(i)}) = 0$$

$$\Rightarrow \sum_{i=1}^{m} (\vartheta_0 + \vartheta_1 x^{(i)} + \vartheta_2 x^{2(i)} - y^{(i)}) = 0$$

$$\Rightarrow \sum_{i=1}^{m} \vartheta_0 + \sum_{i=1}^{m} \vartheta_1 x^{(i)} + \sum_{i=1}^{m} \vartheta_2 x^{2(i)} - \sum_{i=1}^{m} y^{(i)} = 0$$

$$\Rightarrow \sum_{i=1}^{m} \vartheta_0 + \sum_{i=1}^{m} \vartheta_1 x^{(i)} + \sum_{i=1}^{m} \vartheta_2 x^{2(i)} = \sum_{i=1}^{m} y^{(i)} \text{ --------------> A}$$

Linear Regression using simultaneous equation

- $\sum_{i=1}^{m} \vartheta_0 + \sum_{i=1}^{m} \vartheta_1 x^{(i)} + \sum_{i=1}^{m} \vartheta_2 x^{2(i)} = \sum_{i=1}^{m} y^{(i)}$ ---------------> A
- $\sum_{i=1}^{m} \vartheta_0 x^{(i)} + \sum_{i=1}^{m} \vartheta_1 x^{2(i)} + \sum_{i=1}^{m} \vartheta_2 x^{3(i)} = \sum_{i=1}^{m} y^{(i)} x^{(i)}$ ---------------> B
- $\sum_{i=1}^{m} \vartheta_0 x^{2(i)} + \sum_{i=1}^{m} \vartheta_1 x^{3(i)} + \sum_{i=1}^{m} \vartheta_2 x^{4(i)} = \sum_{i=1}^{m} y^{(i)} x^{2(i)}$ ---------------> C

8x + 2y +3z= 46
7x + 3y + 4Z= 47
2x + y + 2z =1

A X B

$$\begin{bmatrix} 8 & 2 & 3 \\ 7 & 3 & 4 \\ 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 46 \\ 47 \\ 1 \end{bmatrix}$$

- $\begin{bmatrix} \sum_{i=1}^{m} 1 & \sum_{i=1}^{m} x^{(i)} & \sum_{i=1}^{m} x^{2(i)} \\ \sum_{i=1}^{m} x^{(i)} & \sum_{i=1}^{m} x^{2(i)} & \sum_{i=1}^{m} x^{3(i)} \\ \sum_{i=1}^{m} x^{2(i)} & \sum_{i=1}^{m} x^{3(i)} & \sum_{i=1}^{m} x^{4(i)} \end{bmatrix} \begin{bmatrix} \vartheta_0 \\ \vartheta_1 \\ \vartheta_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y^{(i)} \\ \sum_{i=1}^{m} y^{(i)} x^{(i)} \\ \sum_{i=1}^{m} y^{(i)} x^{2(i)} \end{bmatrix}$
- $\begin{bmatrix} m & \sum_{i=1}^{m} x^{(i)} & \sum_{i=1}^{m} x^{2(i)} \\ \sum_{i=1}^{m} x^{(i)} & \sum_{i=1}^{m} x^{2(i)} & \sum_{i=1}^{m} x^{3(i)} \\ \sum_{i=1}^{m} x^{2(i)} & \sum_{i=1}^{m} x^{3(i)} & \sum_{i=1}^{m} x^{4(i)} \end{bmatrix} \begin{bmatrix} \vartheta_0 \\ \vartheta_1 \\ \vartheta_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y^{(i)} \\ \sum_{i=1}^{m} y^{(i)} x^{(i)} \\ \sum_{i=1}^{m} y^{(i)} x^{2(i)} \end{bmatrix}$