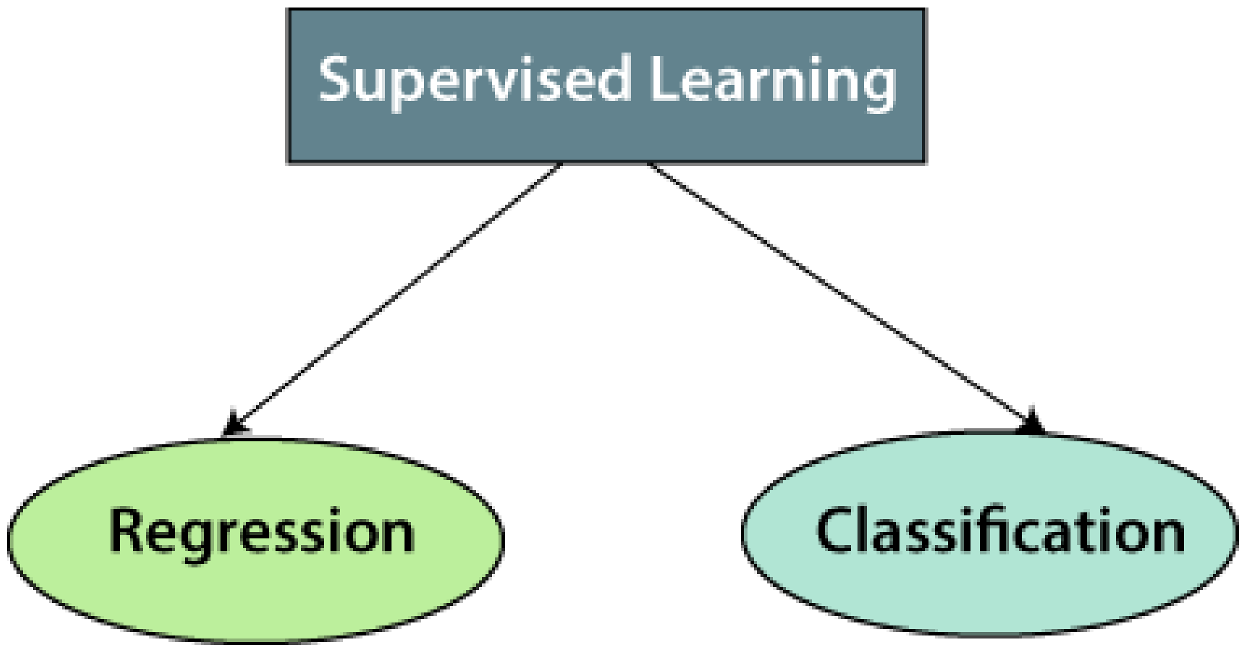
Machine Learning

Dr. Muhammad Adeel Nisar

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

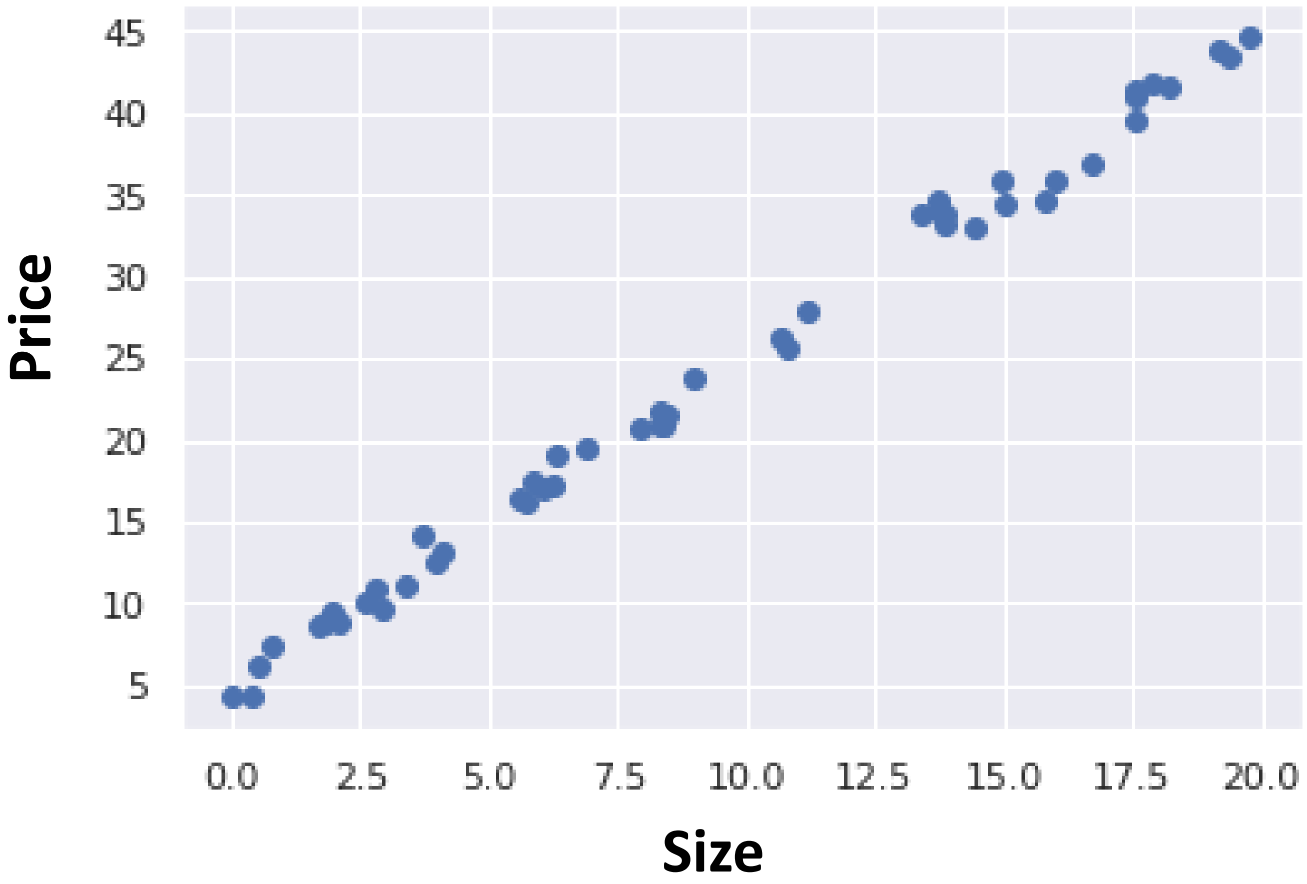

Supervised Learning
Regression
Classification

Regression

- A Supervised learning algorithm
- Taking input variables and trying to fit the output onto a continuous values.
- Linear regression with one variable is also known as “Univariate linear regression”.
- Univariate linear regression is used when you want to predict a single output value y from a single input value x.
- The Hypothesis Function $y' = h\theta(x) = \theta_0 + \theta_1 x$
- Given the training data with right answers, predict the real-valued output for the test data.

Dataset and Plotted Graph

Input Data (x)	Correct Answer (y)
8.3	20.99
14.4	32.89
6.05	17.08
..	..



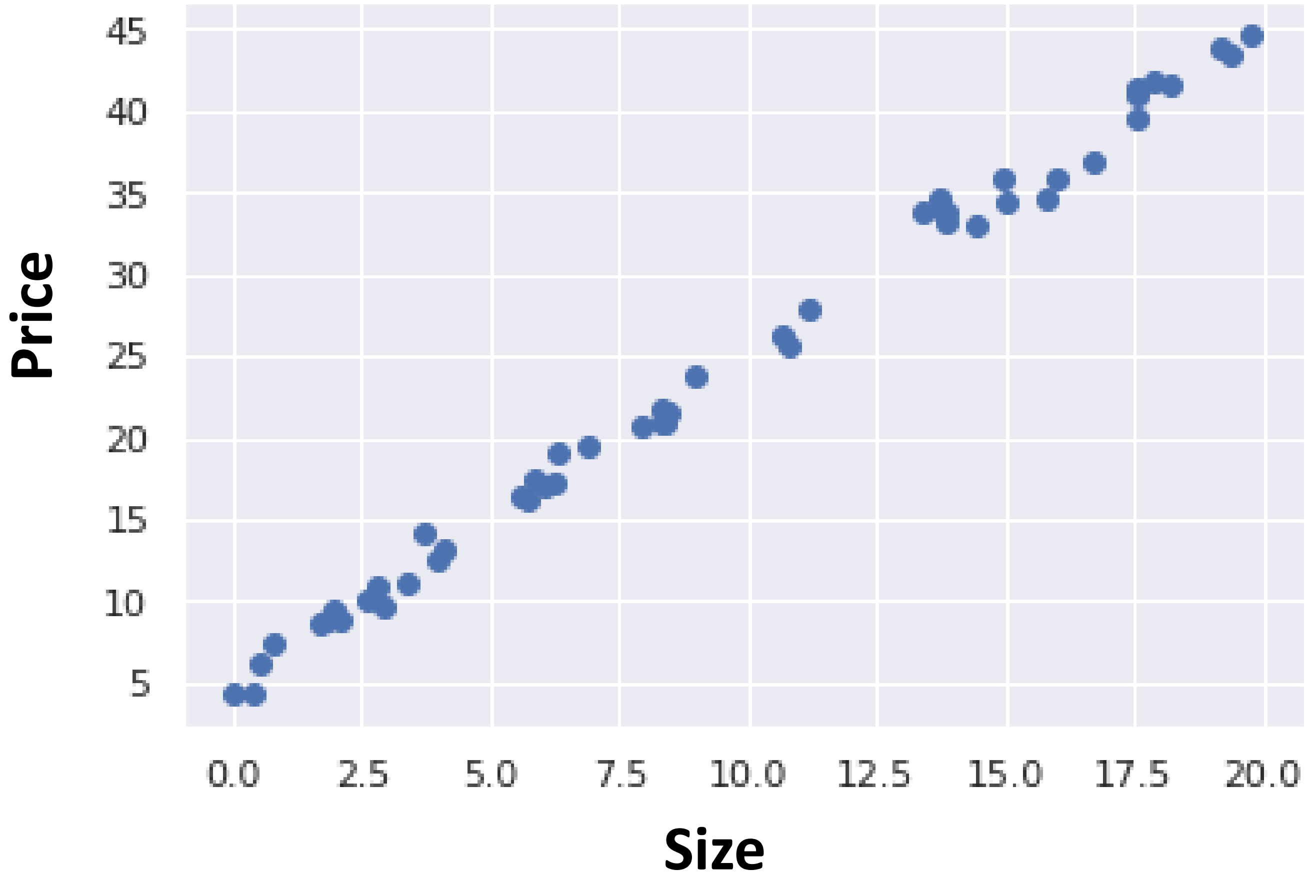
Dataset and Notations

- Notations
- m = Number of Training Examples
- **x** = Input Variable/ Features
- **y** = Output Variable / Target Value
- **(x,y)** is one training example
- **$(x^{(i)},y^{(i)})$** is i^{th} training example
- **$(x^{(1)},y^{(1)})$ = (8.3, 20.99)**

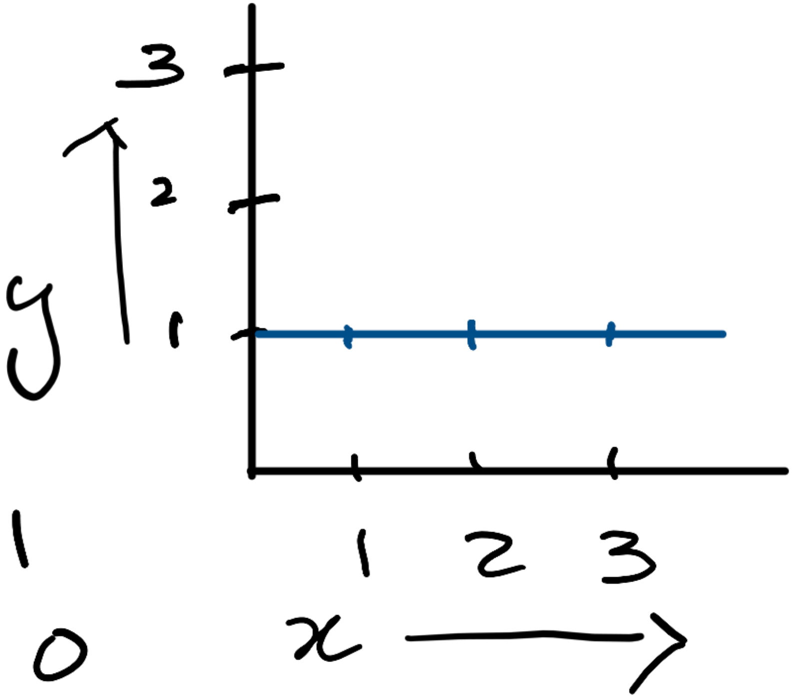
Input Data (x)	Correct Answer (y)
8.3	20.99
14.4	32.89
6.05	17.1
..	..

Linear Regression with One Variable

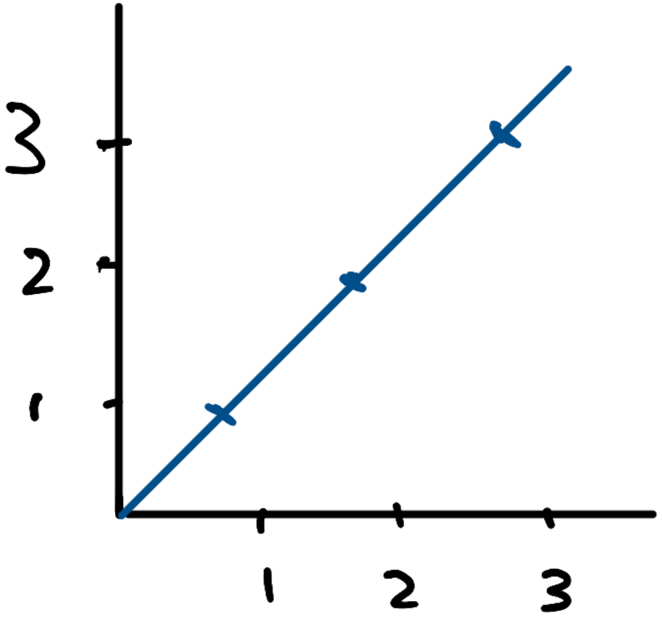
- This is like the equation of a straight line.
- We give $h\theta(x)$ values for θ_0 and θ_1 to get our estimated output y'.
- We are trying to create a function that will map out input data to our output data.



Linear Functions With Varying Values of ϴ



$\theta_0 = 1$
$\theta_1 = 0$

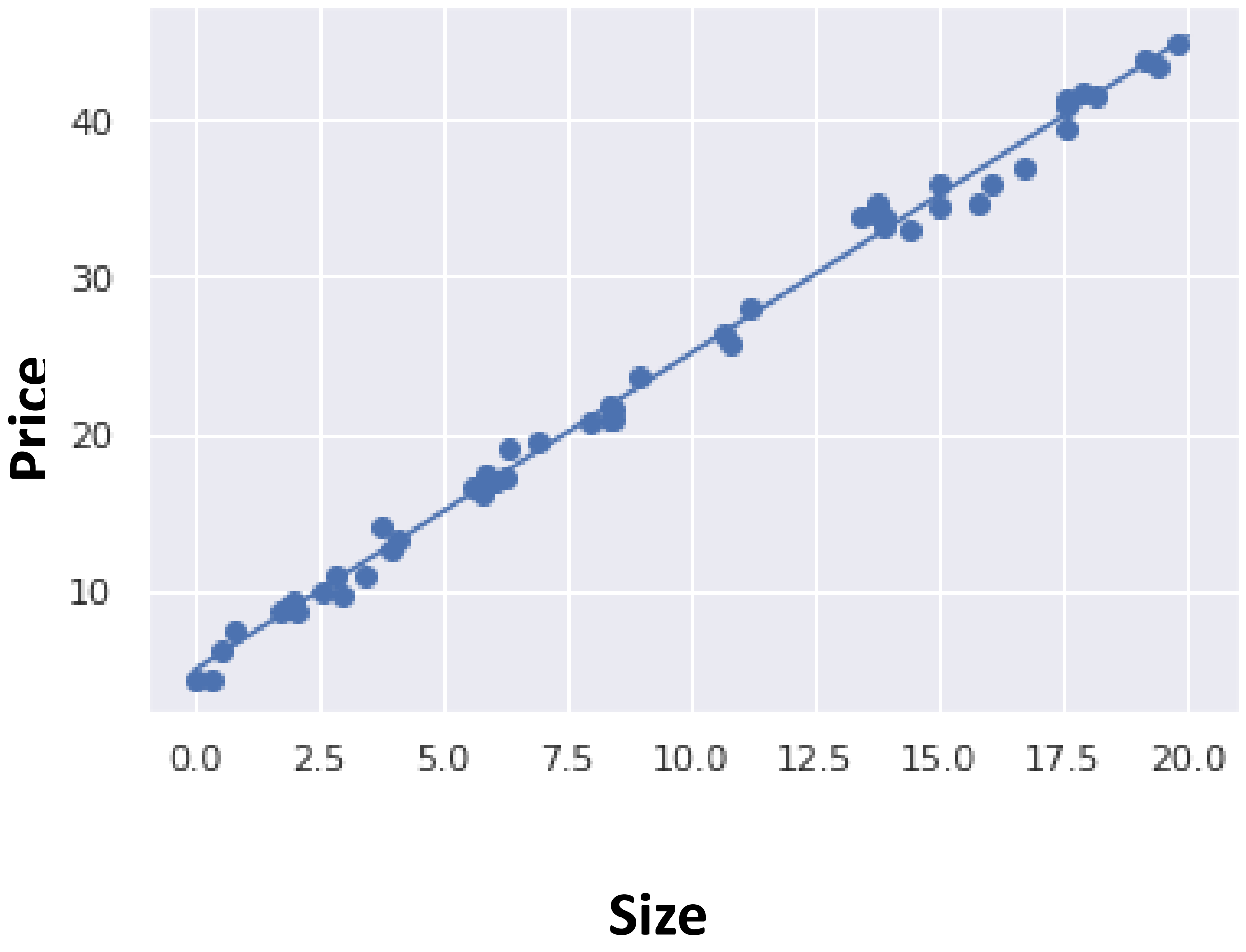


$\theta_0 = 0$
$\theta_1 = 1$

Linear Regression with One Variable

- $y' = h\theta(x) = \theta_0 + \theta_1 x$
- Intercept: $\theta_0 = 5$
- Slope: $\theta_1 = 2$

Input Data (x)	Correct Answer (y)
5	15
10	25
15	35
..	..



Cost Function

- **Hypothesis Function**

$$y' = h\vartheta(x) = \vartheta_0 + \vartheta_1 x$$

- **Cost function** (to measure the performance of hypothesis function)

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

Cost Function

Input Data (x)	Correct Answer (y)	Estimated Answer	Error
8.3	20.99	21.6	-0.61
14.4	32.89	33.8	-0.81
6.05	17.1	17.08	0.02
..	..	..	..

$$Mean\ Square\ Error\ (MSE) = \mathrm{J}(\vartheta_{0,}\vartheta_{1}) = \frac{1}{2m}\sum_{i=1}^{n}(\ y^{(i)} - \ y'^{(i)})^2$$

Cost Function

- **Hypothesis Function**

$$y' = h\vartheta(x) = \vartheta_0 + \vartheta_1 x$$

- **Cost function** (to measure the performance of hypothesis function)

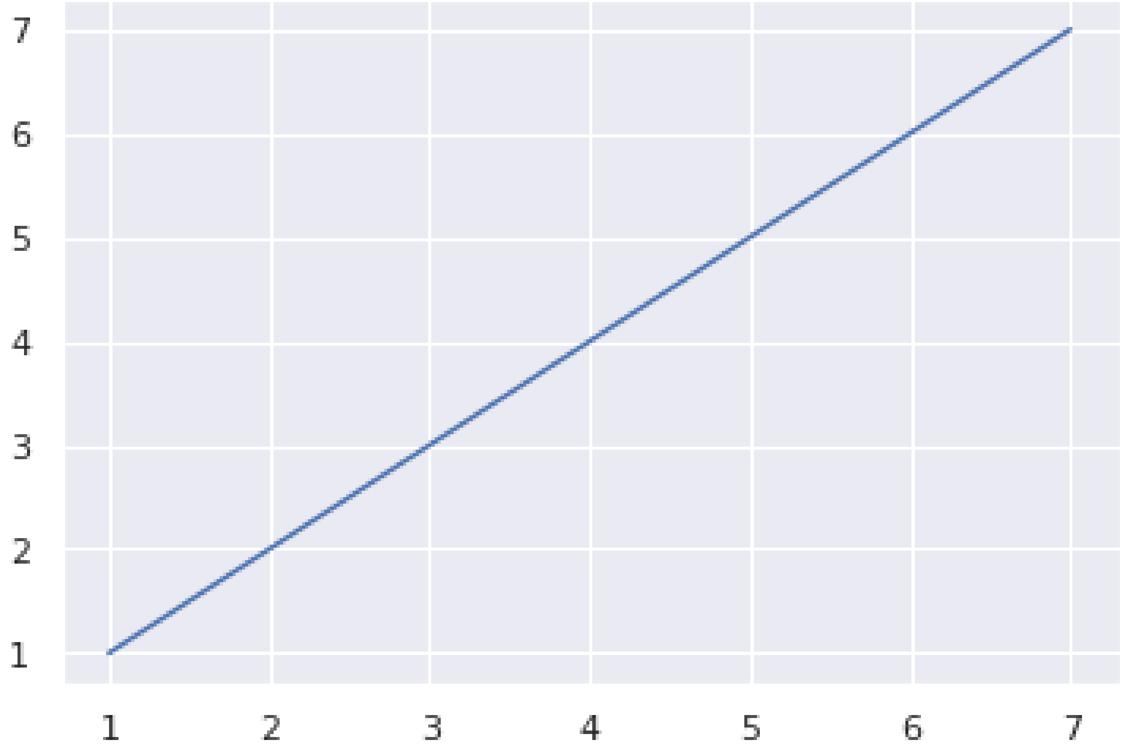
$$J(\vartheta_{0,}\vartheta_1) = \frac{1}{2m}\sum_{i=1}^{m}(\ y'^{\,(i)} - \ y^{\,(i)}\ )^2 = \frac{1}{2m}\sum_{i=1}^{m}(\ h\vartheta(x^{\,(i)}) - \ y^{\,(i)}\ )^2$$

- **Objective:**

$$\min_{\vartheta_{0,}\vartheta_1} \mathrm{J}(\vartheta_{0,}\vartheta_1)$$

Cost Function - Example

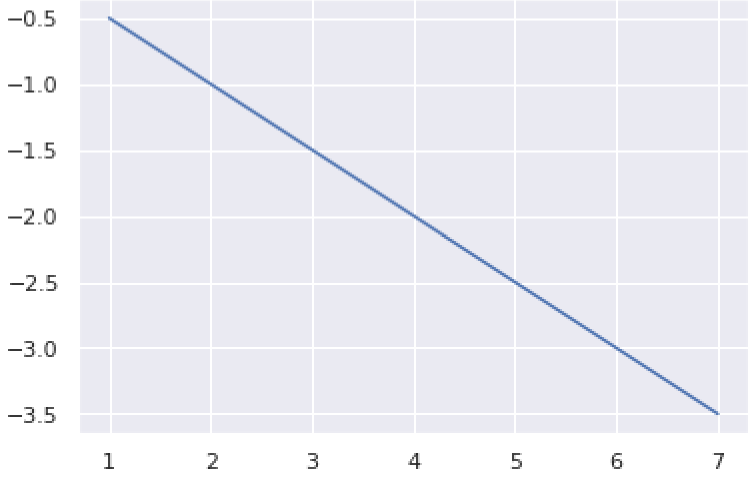
- `x = [1, 2, 3, 4, 5, 6, 7]`
- `y = [1, 2, 3, 4, 5, 6, 7]`
- $y' = h\theta(x) = \theta_0 + \theta_1 x$
- Assume $\theta_0 = 0$
- So, $y' = h\theta(x) = \theta_1 x$

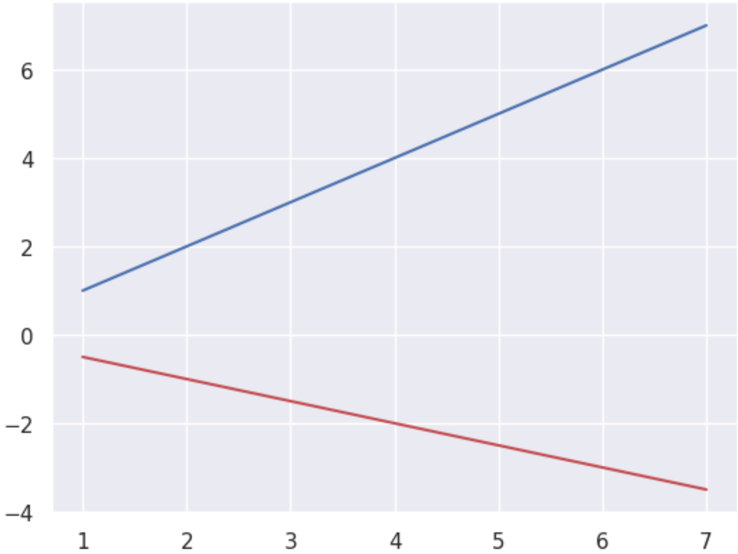


θ_1 = [-0.5, -0.25, 0.0, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.25]

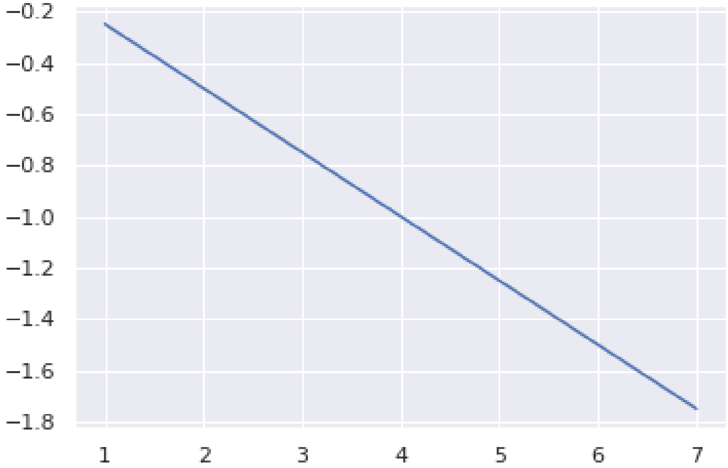
Cost Function - Example

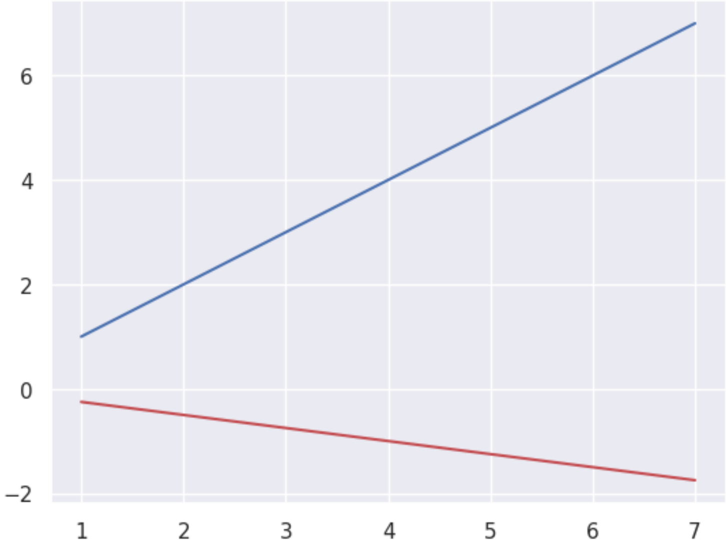
- $\vartheta_1 = -0.5, J(\vartheta_1) = 45.0$





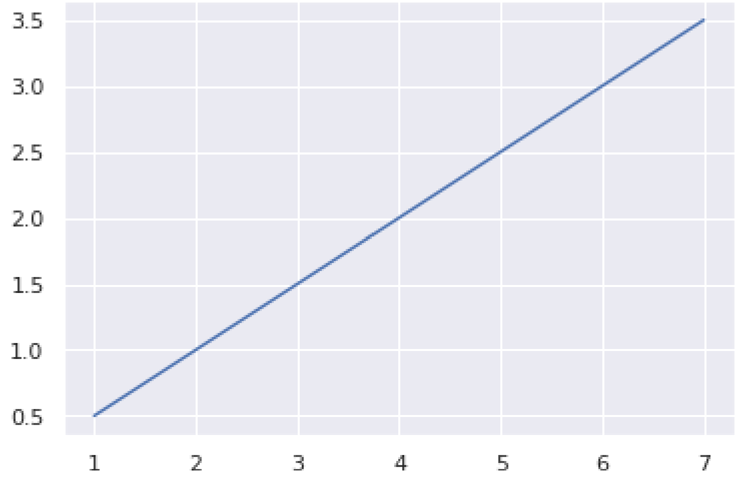
- $\vartheta_1 = -0.25, J(\vartheta_1) = 31.25$

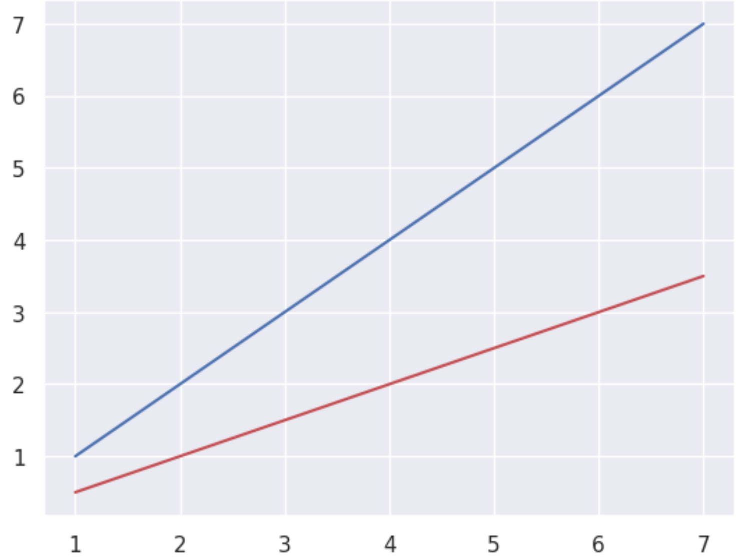




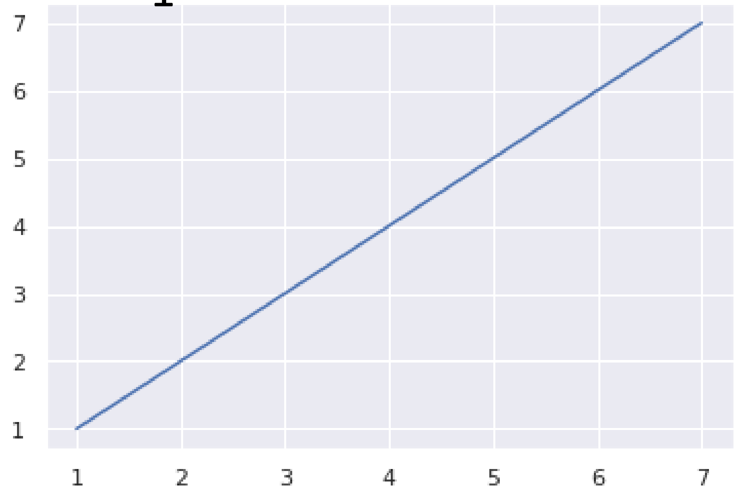
Cost Function - Example

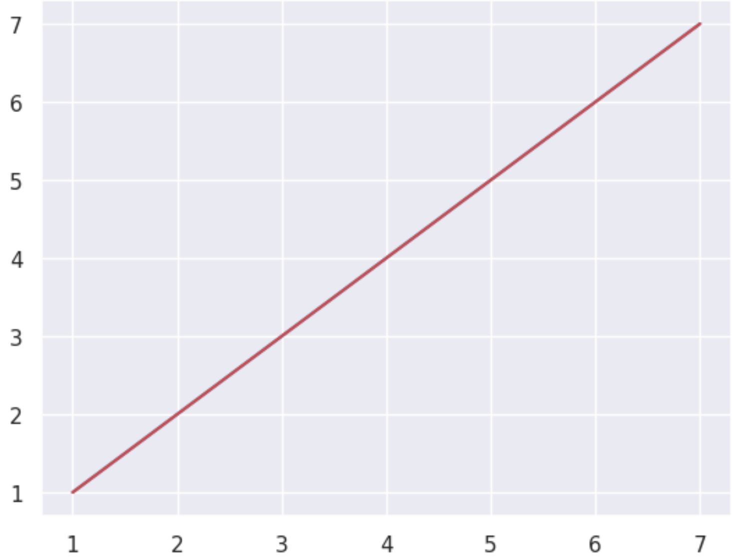
- ϑ_1 = 0.5, J(ϑ_1)= 5.0





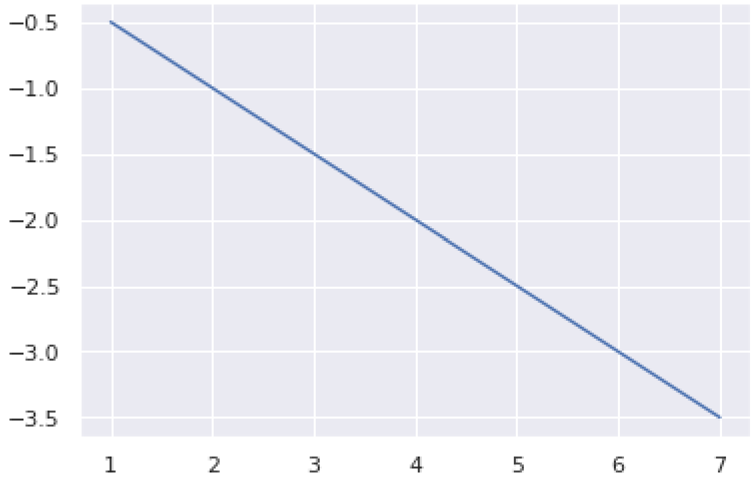
- ϑ_1 = 1.0, J(ϑ_1)= 0.0



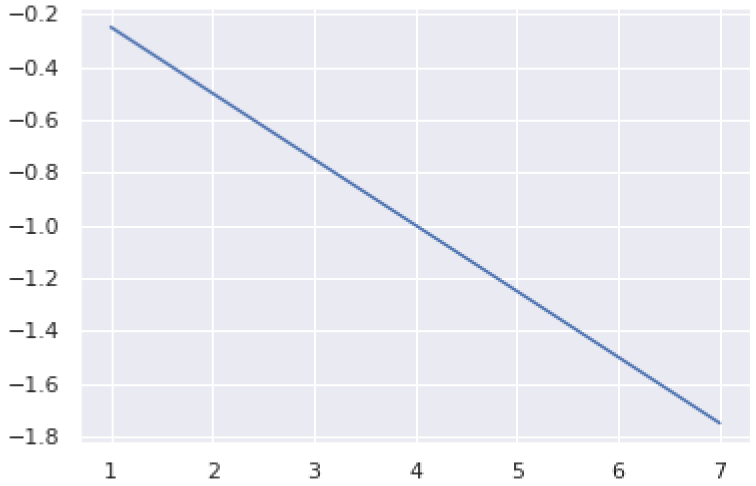


Cost Function - Example

- ϑ_1 = 1.5, J(ϑ_1)= 5.0



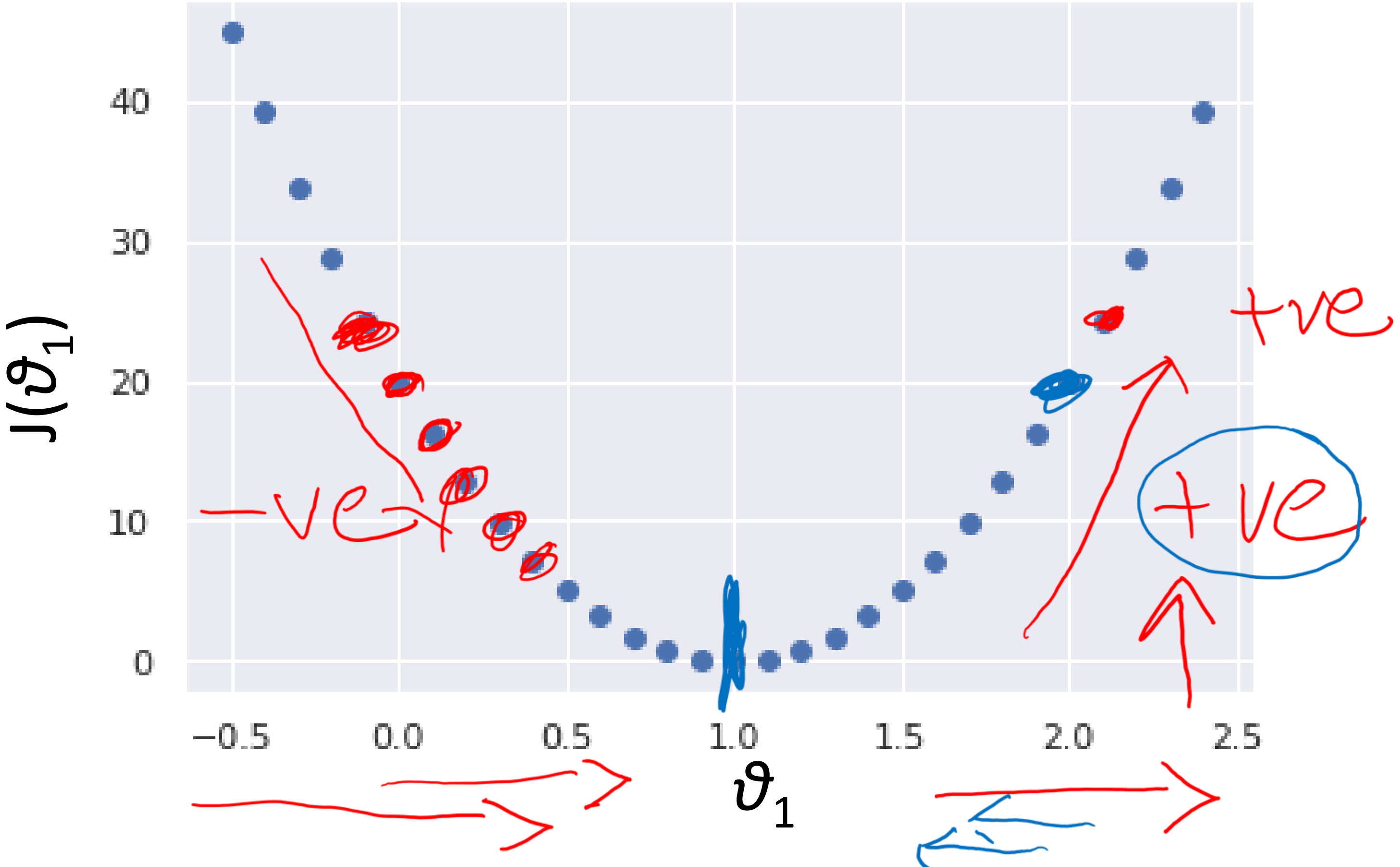
- ϑ_1 = 2.0, J(ϑ_1)= 20.0



Graph of J(ϑ_1)

- θ_1 = [-0.5, -0.25, 0.0, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.25]
- $J(\theta_1)$ =[45.0, 31.25, 20.0, 11.25, 5.0, 1.25, 0.0, 1.25, 5.0, 11.25, 20.0, 31.25]

$\vartheta_1 = (-ve)$



Recap of the Last Lecture

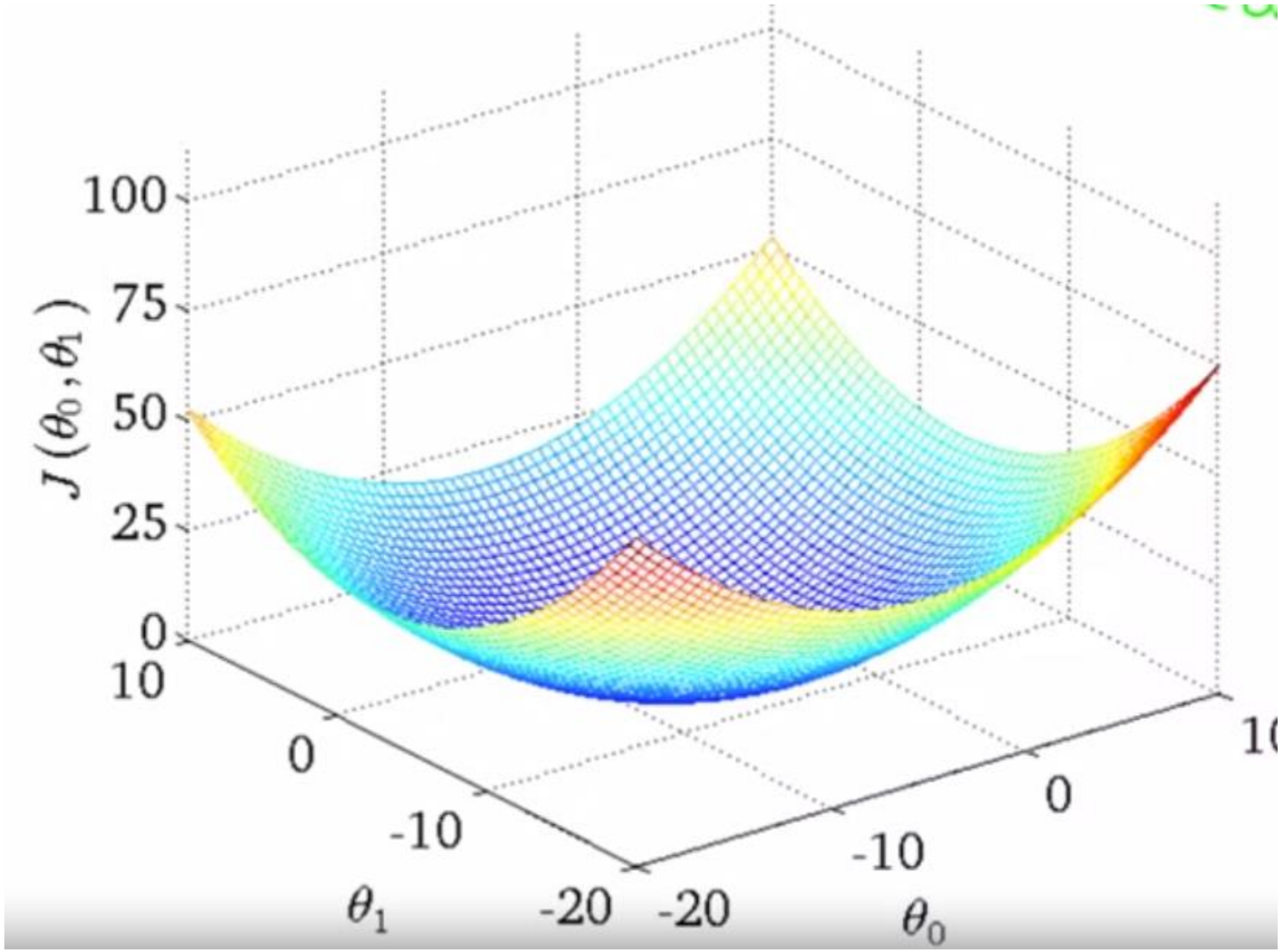
- Univariate Linear Regression

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

- 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(ϑ_{0}, ϑ_1)



Some Concepts

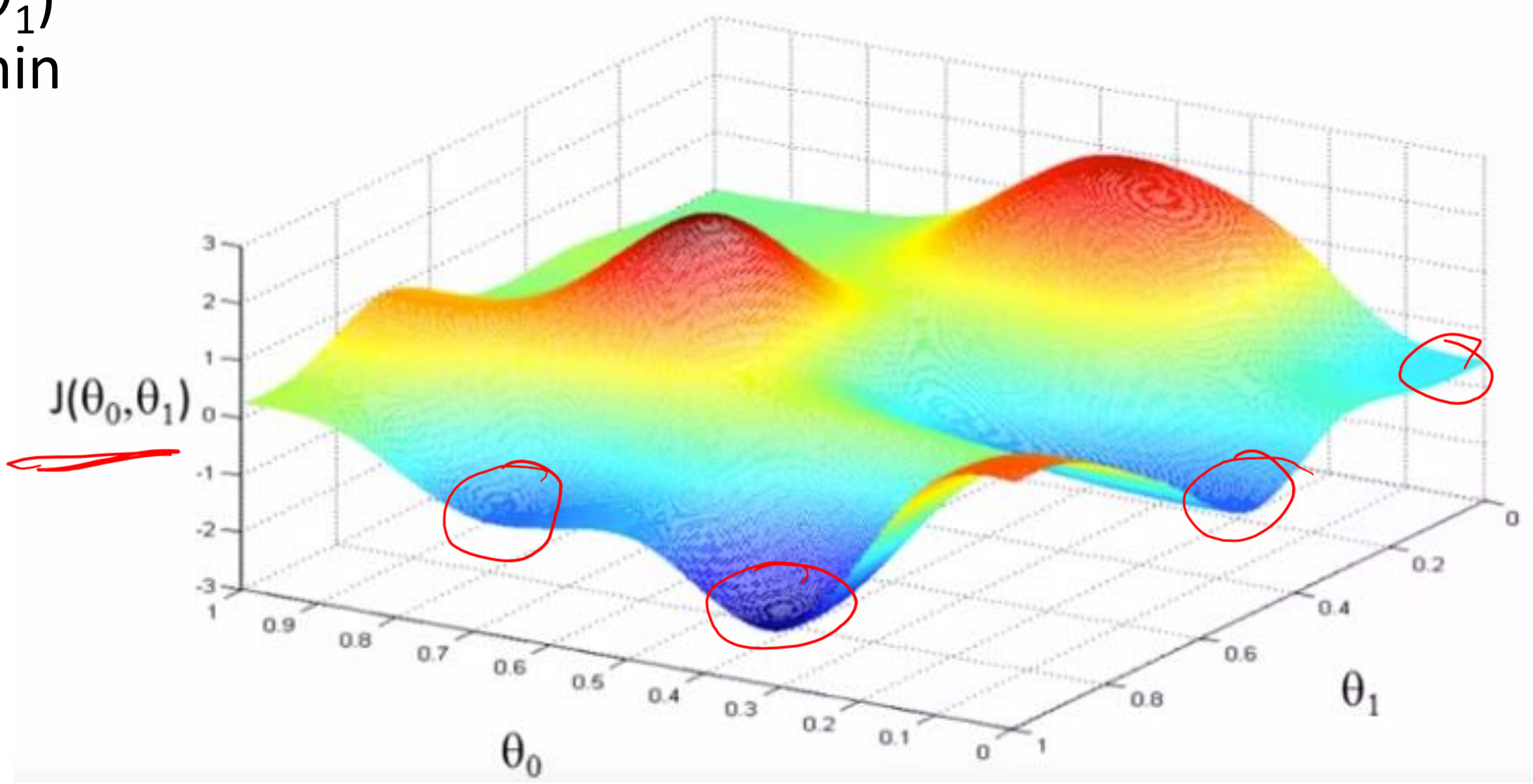
- Global Minimum

- Local Minimum

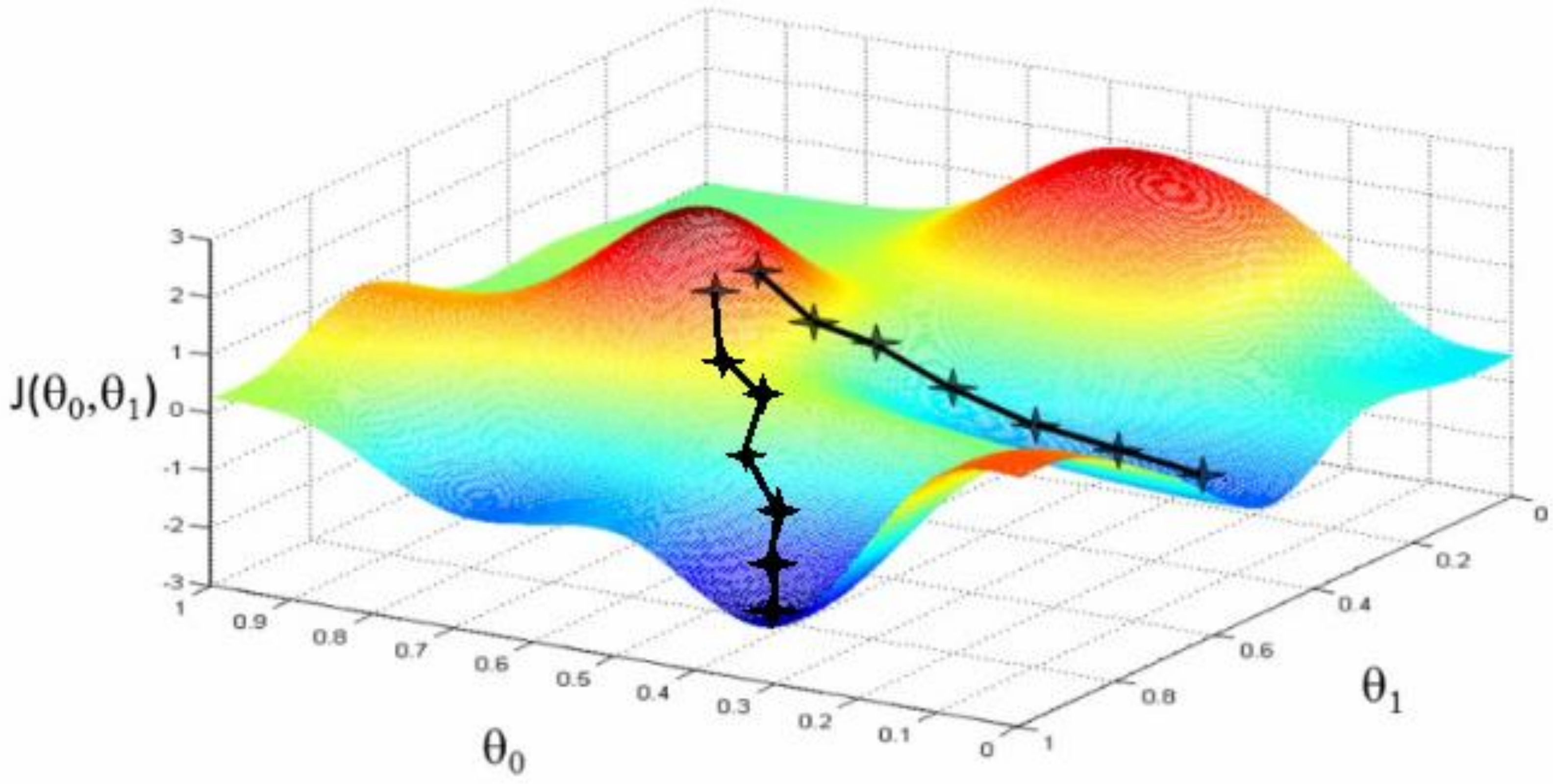
- Convex Functions

Gradient Descent Algorithm

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



Solving Minimization Problem



Derivatives

$f(x) = 4x$

$f(x) = x^3$

$f(x) = (x + 2)^4$

$f(x,y) = (3x + 2y + 2)^2$

Gradient

- What is gradient?
  - The gradient of a function at a point is a vector that points in the direction of the greatest rate of increase of the function at that point.
  - It represents the steepest ascent in a multivariable function and is calculated using partial derivatives with respect to each variable.
- Mathematical Representation:
  - For a function $f(x, y)$, the gradient is denoted as $\nabla f(x, y)$, and it points in the direction where $f(x, y)$ increases most rapidly.
  - The magnitude of the gradient indicates how fast the function is changing in that direction.
- Applications in Optimization:
  - In machine learning, the gradient is used in gradient-based optimization methods like Gradient Descent to minimize loss functions by moving in the direction opposite to the gradient (steepest descent).

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
- $temp0 = \vartheta_0 - \alpha \frac{\partial}{\partial \vartheta_0} J(\vartheta_0, \vartheta_1)$
- $temp1 = \vartheta_1 - \alpha \frac{\partial}{\partial \vartheta_1} J(\vartheta_0, \vartheta_1)$
- $\vartheta_0 :=$ temp0
- $\vartheta_1 :=$ temp1

Linear Regression with Gradient Descent

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

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

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

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

- Repeat until **converge**

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

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

- Simultaneous update

Recap of Univariate Linear Regression

Size of Plot x_1	Price y
5	15
10	25
7	35
…	…
4	5

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

Linear Regression with Multiple Features

mean
max
min

mean
max
min

μ mean
min
max

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
10	0.9	0	2.5	25
7	1.5	1	1.9	35
…	…	…	…	…
4	0.5	0	10	5

$x^{(1)}$

$x^{(m)}$

5 - mean

max - min

4 - mean

max - min

2 - μ

max - min

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_j *is jth feature*

n is the number of features

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

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

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

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**

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

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

- **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

- 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) {

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

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

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

}

mean

min

max

Feature Scaling

- Gradient Descent:

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

- 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

Normal Equations

- Method to solve for θ analytically
- $y' = h_{\vartheta}(x) = \vartheta_0 + \vartheta_1 x$
- $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_{\theta}(x^{(i)}) - y^{(i)})^2$
- $\frac{d}{d(\vartheta)} J(\vartheta) = 0$

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:

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

Objective:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

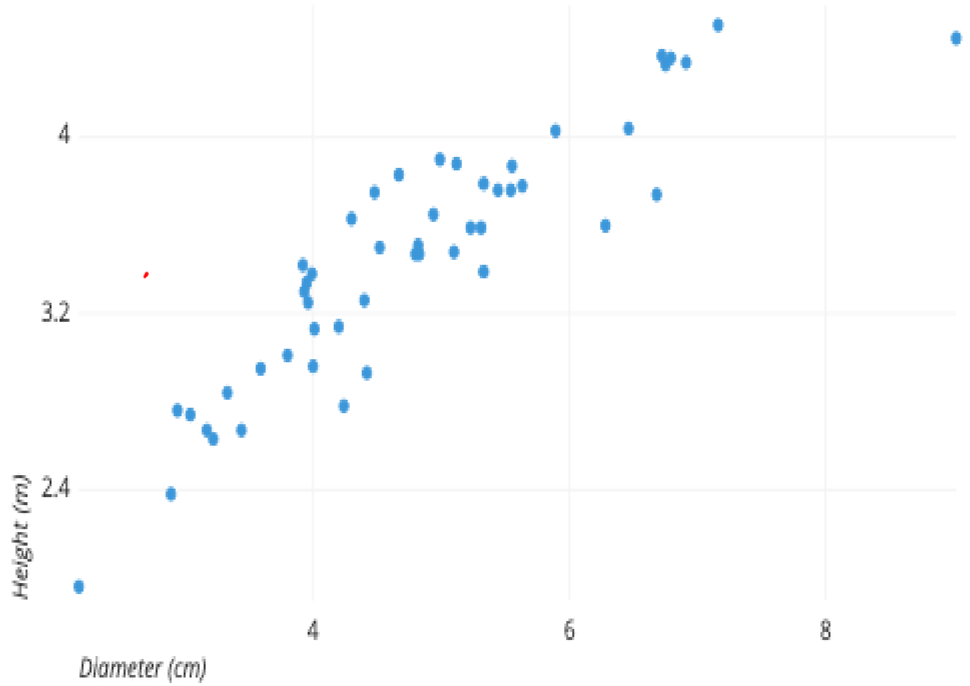
A X B

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

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

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

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



Comparison of Gradient Descent and Normal Equation

Gradient Descent

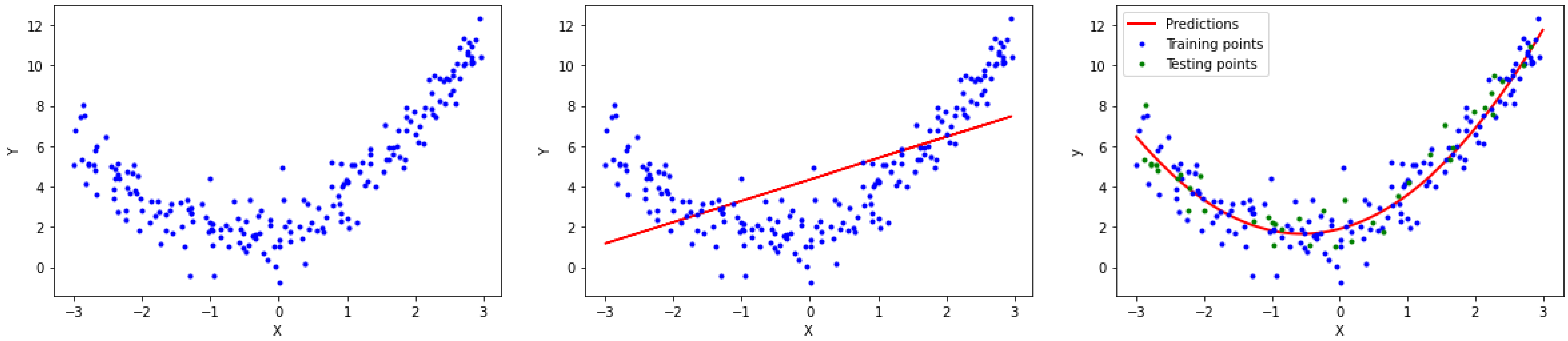
- Need to choose learning rate
- Need many iterations
- Works well for even large number of features

Normal Equation

- No Need to choose learning rate
- No iterations required
- Works slow if number of features is very large $O(n^3)$

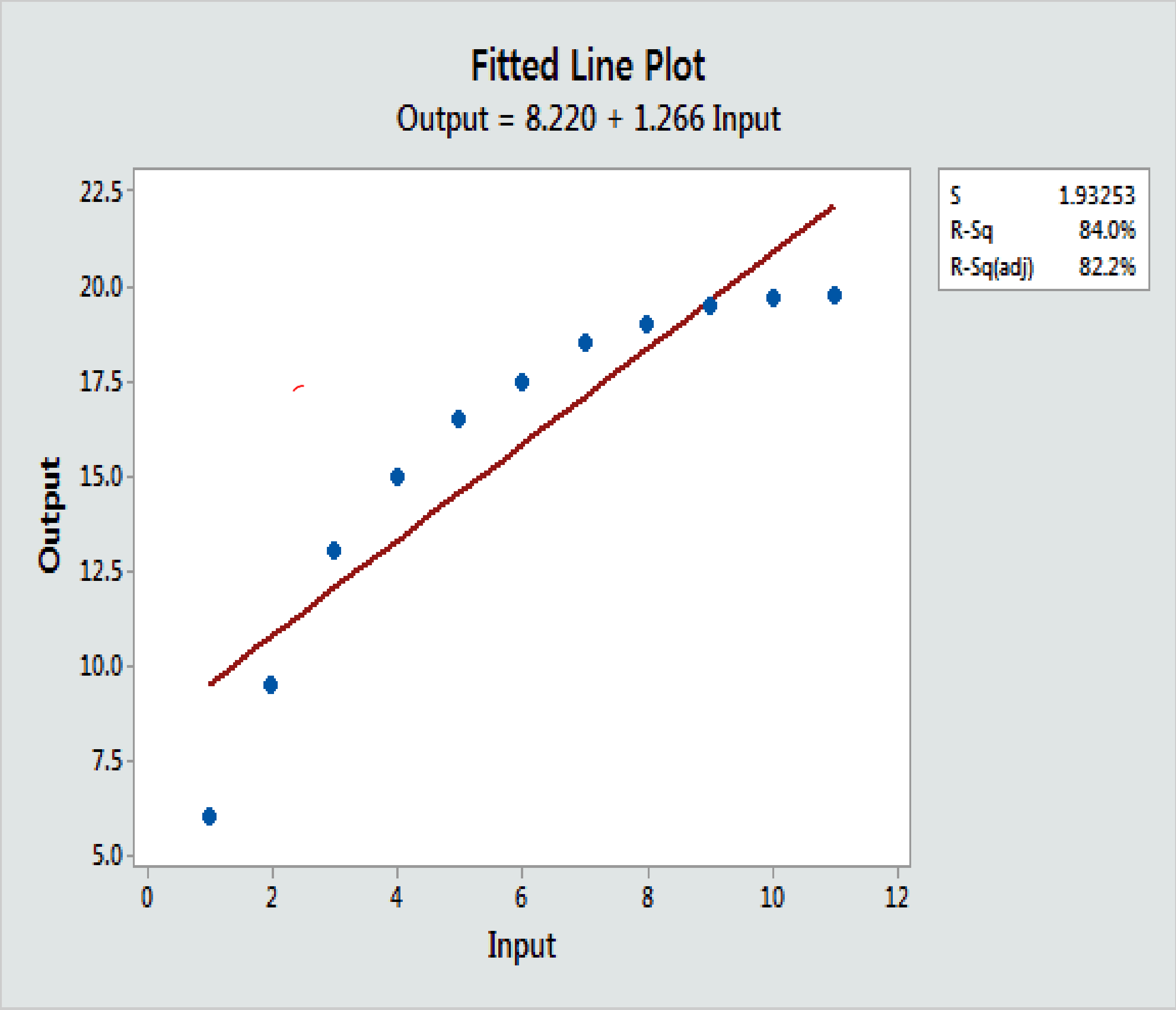
Polynomial Regression

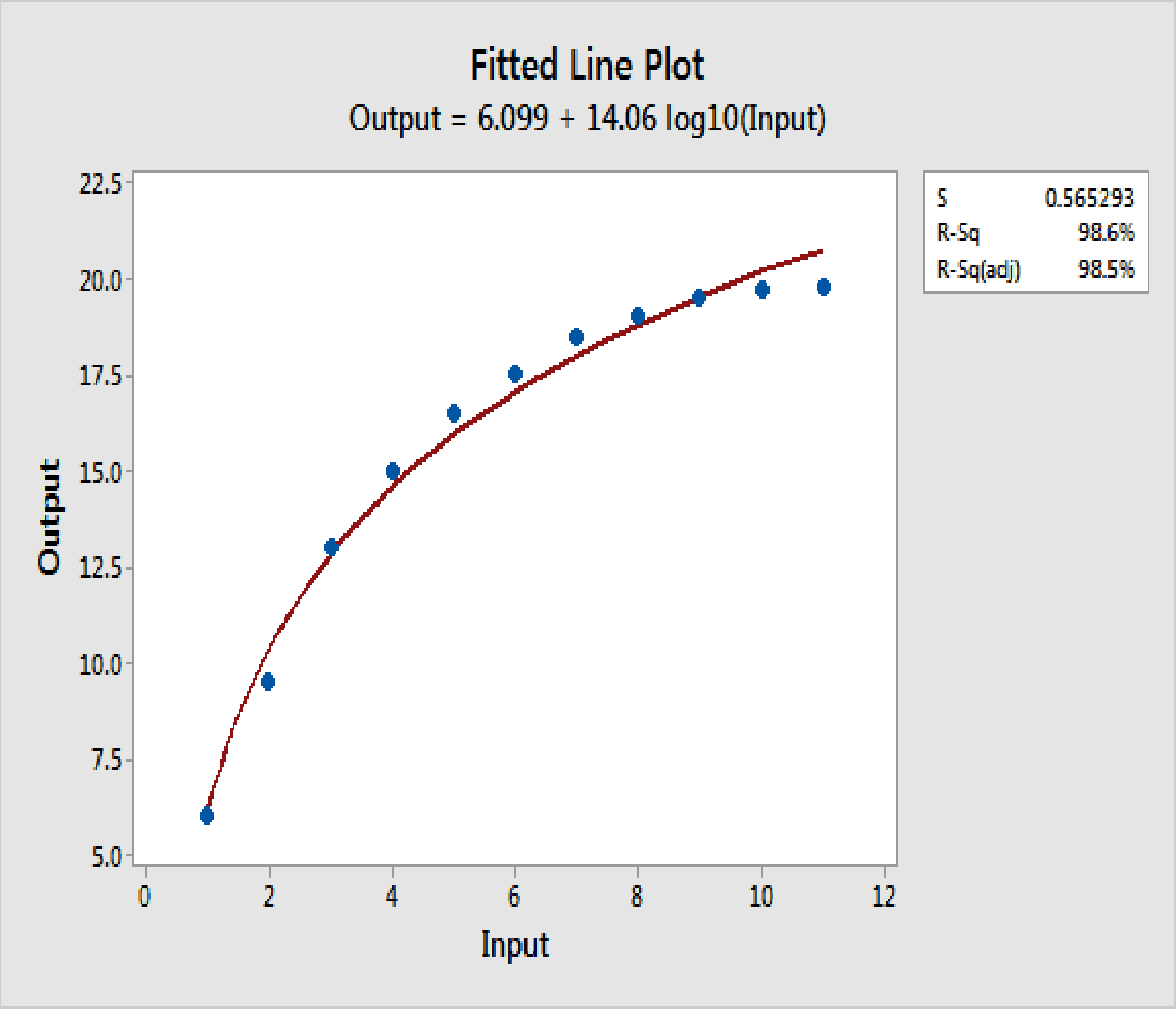
- If the relationship between data is not linear:



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

Quadratic Model





Normal Equations

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

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

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

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

Linear Algebra review (optional)

Matrices and vectors

Courtesy: Andrew Ng

Matrix: Rectangular array of numbers:

Dimension of matrix: number of rows x number of columns

Matrix Elements (entries of matrix)

$$A = \begin{bmatrix} 1402 & 191 \\ 1371 & 821 \\ 949 & 1437 \\ 147 & 1448 \end{bmatrix}$$

$A_{ij} =$ "i, j entry" in the i^{th} row, j^{th} column.

Vector: An n x 1 matrix.

$$y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

$y_i = i^{th}$ element

1-indexed vs 0-indexed:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \qquad y = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Linear Algebra review (optional)

Addition and scalar multiplication

Matrix Addition

$$\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0.5 \\ 2 & 5 \\ 0 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0.5 \\ 2 & 5 \end{bmatrix} =$$

Scalar Multiplication

$$3 \times \begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 4 & 0 \\ 6 & 3 \end{bmatrix} / 4 =$$

Combination of Operands

$$3 \times \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix} - \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} /3$$

Linear Algebra review (optional)

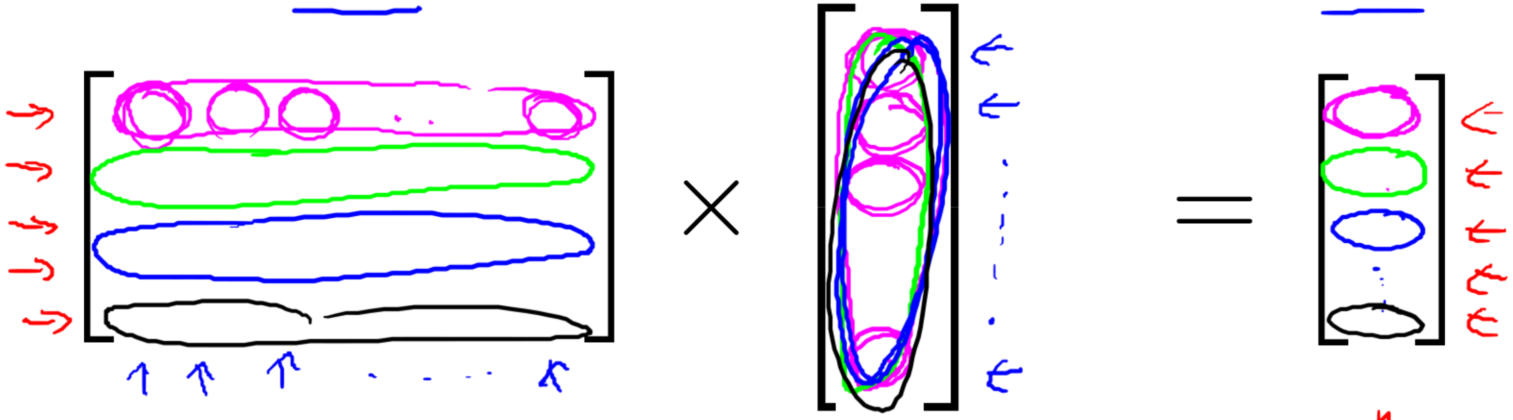
Matrix-vector multiplication

Example

$$\begin{bmatrix} 1 & 3 \\ 4 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix} =$$

Details:

$$A \times x = y$$



m x n matrix (m rows, n columns)	n x 1 matrix (n-dimensional vector)	m-dimensional vector

To get y_i, multiply A's i^{th} row with elements of vector x, and add them up.

Example

$$\begin{bmatrix} 1 & 2 & 1 & 5 \\ 0 & 3 & 0 & 4 \\ -1 & -2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \boxed{14} \\ \boxed{13} \\ \boxed{-7} \end{bmatrix} = \begin{bmatrix} 14 \\ 13 \\ -7 \end{bmatrix}$$

$\boxed{3} \times 4$ $\quad$ 4×1 $\quad$ 3×1

$$1 \times 1 + 2 \times 3 + 1 \times 2 + 5 \times 1 = 14$$

$$0 \times 1 + 3 \times 3 + 0 \times 2 + 4 \times 1 = 13$$

$$-1 \times 1 + (-2) \times 3 + 0 \times 2 + 0 \times 1 = -7$$

Reading - Homework

- Machine Learning
  - Resource R1
  - Book B1: 2.7, 2.8, 3.1, 4.1, 4.2, 4.3
  - Book B2: Chapter 4 (Page 113 to 143)
  - Book B3: 2.6