SSID: WWCode

Password: password

WOMEN WHO CODE MANILA

ML and AI Study Group

Twitter: @wwcodemanila
FB: fb.com/wwcodemanila

#WWCodeManila
#YourProgrammingLanguage
#StudyGroup

Issa

Research Fellow

PCARI

New Member's Introduction

I am <name>

<your current profession>

<why did you join this study group?>

OUR MISSION

Inspiring women to excel in technology careers.

OUR VISION

A world where women are representative as technical executives, founders, VCs, board members and software engineers.

WOMEN WHO CODE MANILA

STUDY GROUP

Study groups are events where women can come together and help each other learn and understand a specific programming language, technology, or anything related to coding or engineering.

GUIDELINES

- If you have a question, just **ask**
- If you have an idea, **share it**
- **Make friends** and learn from your study groupmates
- **Do not** recruit or promote your business

SHOW & TELL

LINEAR REGRESSION IN ONE VARIABLE

(a.k.a Univariate Linear Regression)

Agenda

1. Review
2. Hypothesis Function
3. Cost Function
4. Gradient Descent Algorithm

Review

- Supervised vs. Unsupervised

Review

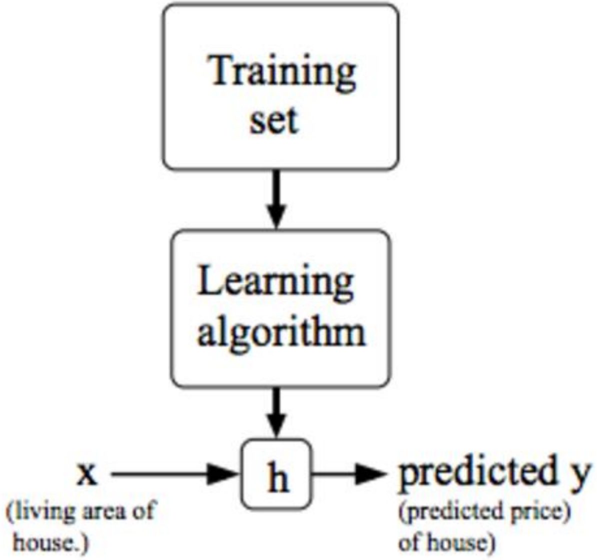
- Supervised vs. Unsupervised
- Supervised Learning:
  - **Classification** – output variables take discrete class labels
  - **Regression** – output variable takes continuous values

Review

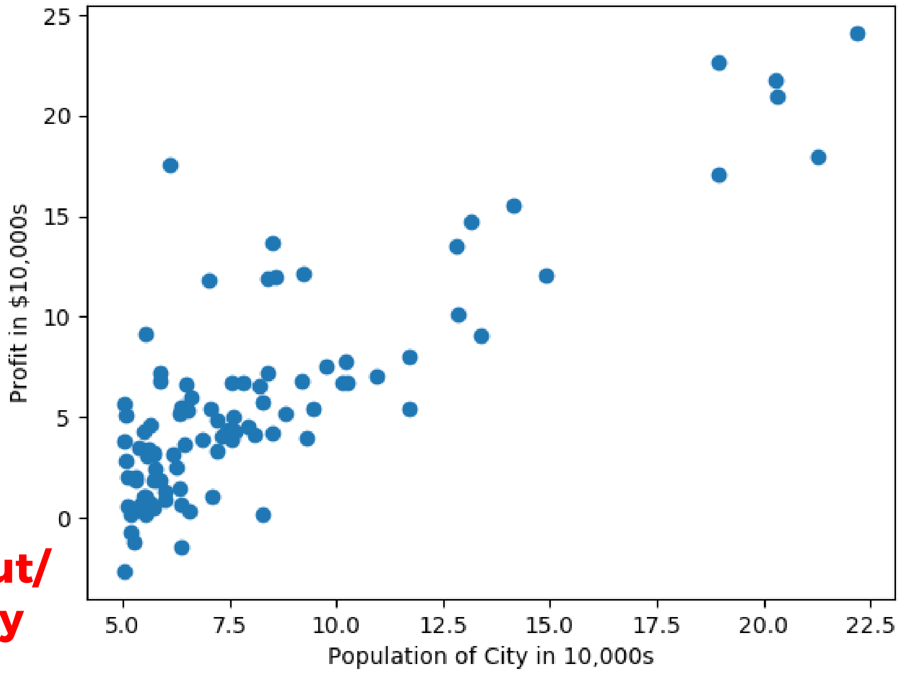
- Supervised vs. Unsupervised
- Supervised Learning:
  - **Classification** – output variables take discrete class labels
  - **Regression** – output variable takes continuous values
- **Univariate Linear Regression –** predicts a single output y from a single input value x

Linear Regression in One Variable

e.g. **Predicting Profit of a City**



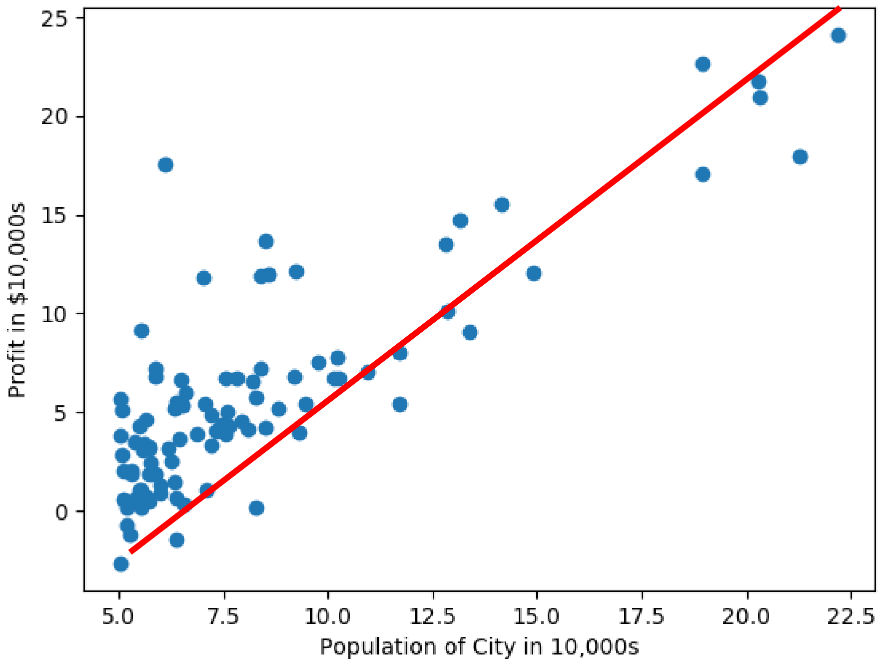
Output/ Class y



Input feature x

Linear Regression

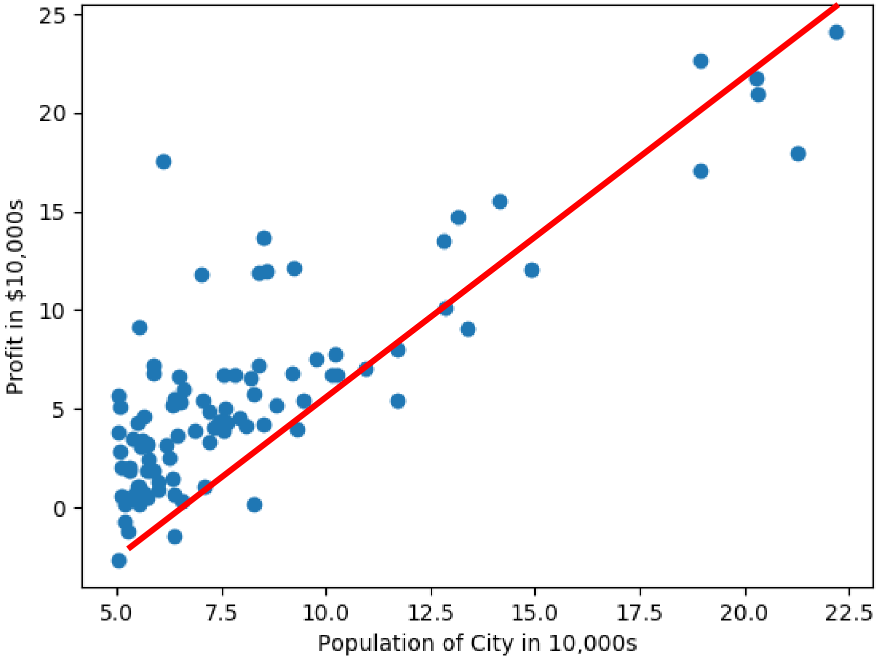
We want to **fit** a line through the data.



Linear Regression

We want to **fit** a line through the data.

And make it so that the line **generalizes** the data well.

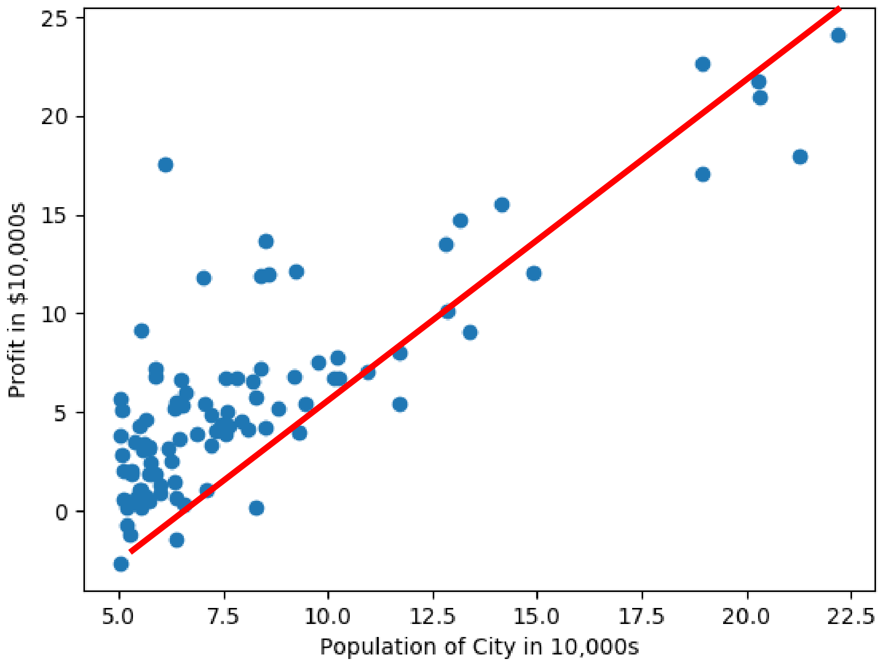


Linear Regression

We want to **fit** a line through the data.

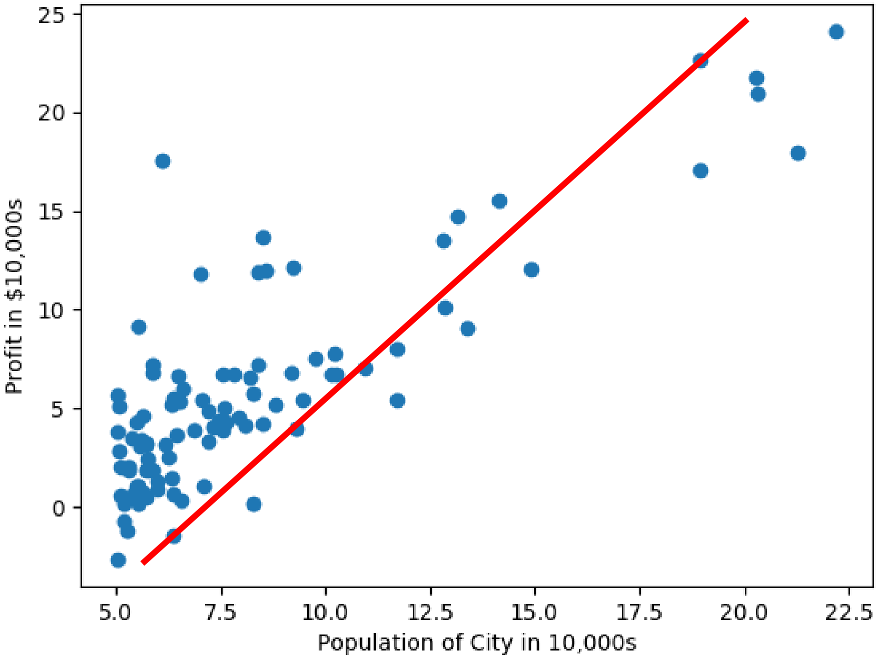
And make it so that the line **generalizes** the data well.

e.g. If Imus City has a population of about 21k, then its predicted profit is...?



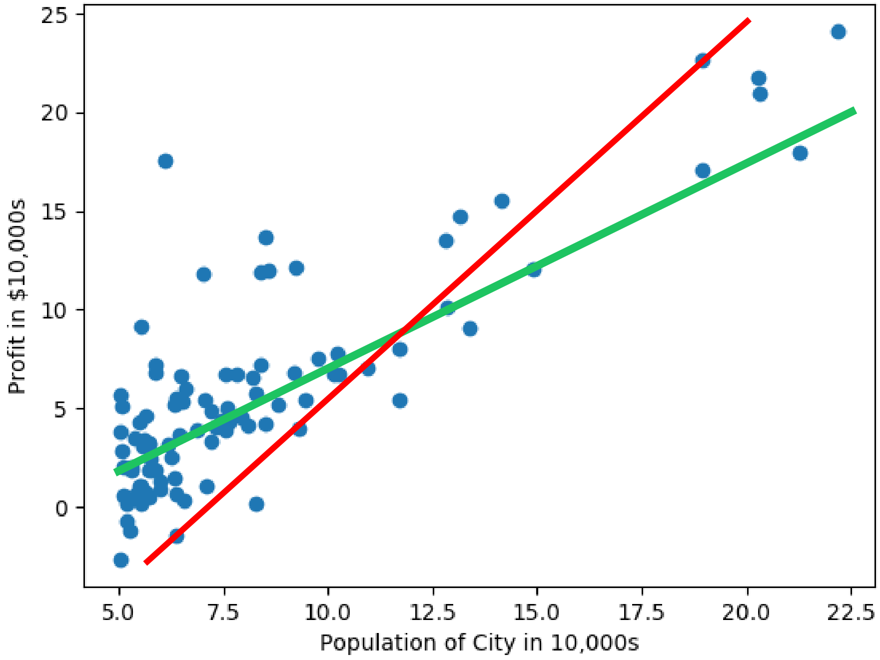
Linear Regression

How do we come up with the “right” line?



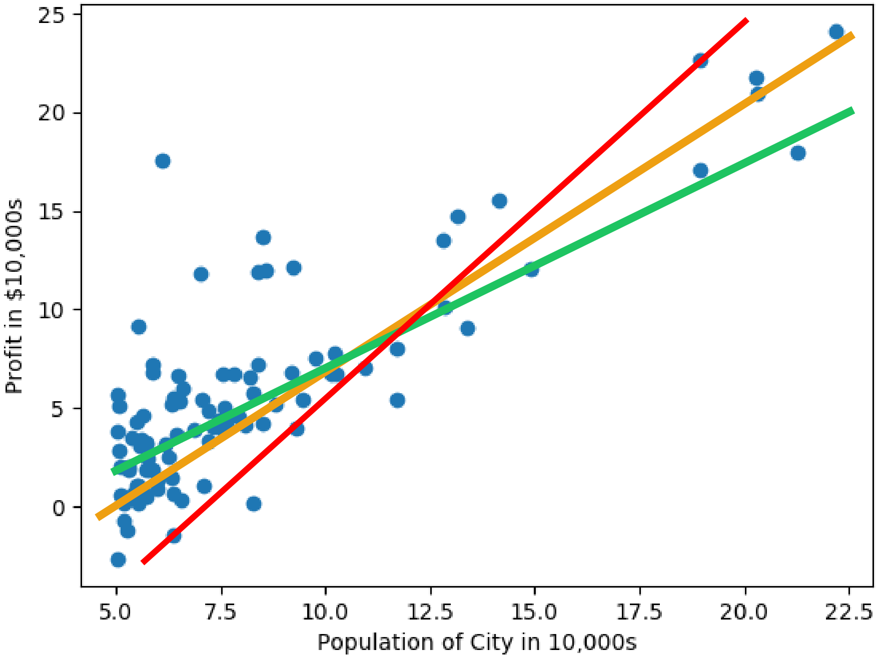
Linear Regression

How do we come up with the “right” line?



Linear Regression

How do we come up with the “right” line?



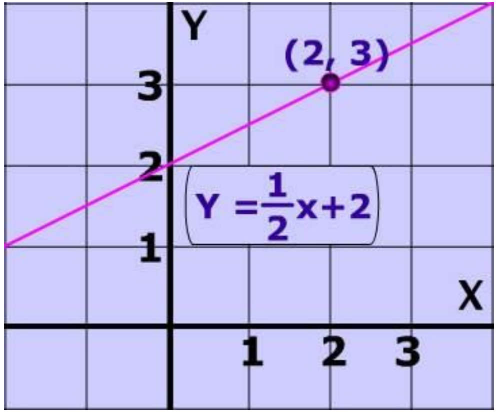
Equation of a Line

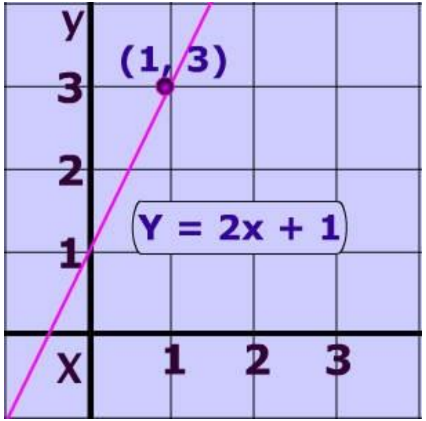
- Recall the equation of a line?

Equation of a Line

- Recall the equation of a line?

$$y = \ mx + b$$





Exercise:

Draw the lines:

$$y = {}^{3}/_{2}\, x + 1$$
$$y = -2x - 5$$

Equation of a Line

- Recall the equation of a line?

$$y = mx + b$$

- More generally,

$$y = \theta_1 x + \theta_0$$

where θ_0, θ_1 are called “weights”.
θ_0 has a special name called the “bias”

Hypothesis Function

- We need to find the right values for θ_0 and θ_1.
- How do we get the right values?

Hypothesis Function

- We need to find the right values for θ_0 and θ_1.
- How do we get the right values?
- We start with a guess (or *hypothesis*) using random weights

Hypothesis Function

- We need to find the right values for θ_0 and θ_1.
- How do we get the right values?
- We start with a guess (or *hypothesis*) using random weights
- We define our **hypothesis function** as

$$h(x) = \theta_1 x + \theta_0$$

where θ_0 and θ_1 are set to some random values (e.g. $\theta_0 = 0, \theta_1 = 0$)

Hypothesis Function

- Let x be the input features and y be the true output values.
- Suppose we come up with two hypotheses:

Input x	Output y	$h_\theta(x)$ $\theta_0 = 2,\ \theta_1 = 3$	$h_\theta(x)$ $\theta_0 = 4,\ \theta_1 = 3$
0	4	?	?
1	8	?	?
2	9	?	?
3	13	?	?

Hypothesis Function

- Let x be the input features and y be the true output values.
- Suppose we come up with two hypotheses:

Input x	Output y	$h_\theta(x)$ $\theta_0 = 2, \theta_1 = 3$	$h_\theta(x)$ $\theta_0 = 4, \theta_1 = 3$
0	4	2	?
1	8	5	?
2	9	8	?
3	13	11	?

How good is our hypothesis?

- We need a way to measure how close our hypothesis $h_\theta(x)$ is to the true output values y.
- i.e. we want a function that measures how good or bad our hypothesis function performs

How good is our hypothesis?

- We need a way to measure how close our hypothesis $h_\theta(x)$ is to the true output values y.
- i.e. we want a function that measures how good or bad our hypothesis function performs
- We will call such a function the **Cost Function,** which will measure the average error of the $h_\theta(x)$.

Cost Function

$$\text{Cost } J(\theta_0, \theta_1) = \underbrace{\frac{1}{2m}}_{\text{average over } m \text{ examples}} \underbrace{\sum_{i=1}^{m}}_{\text{sum over all } m \text{ examples}} \underbrace{(h(x_i) - y_i)}_{\text{error - we want this to be close to zero}}{}^2$$

"Mean Squared Error (MSE)"

error - we want this to be close to zero

average over m examples

sum over all m examples

Note: The error is squared and the mean is halved as a convenience for computing the gradient descent later.

Minimizing the Cost Function

- We want to minimize the cost function, or the mean squared error (MSE).

Minimizing the Cost Function

- We want to minimize the cost function, or the mean squared error (MSE).

- In other words, we want to find the θ_0 and θ_1 that minimizes $J(\theta_0, \theta_1)$

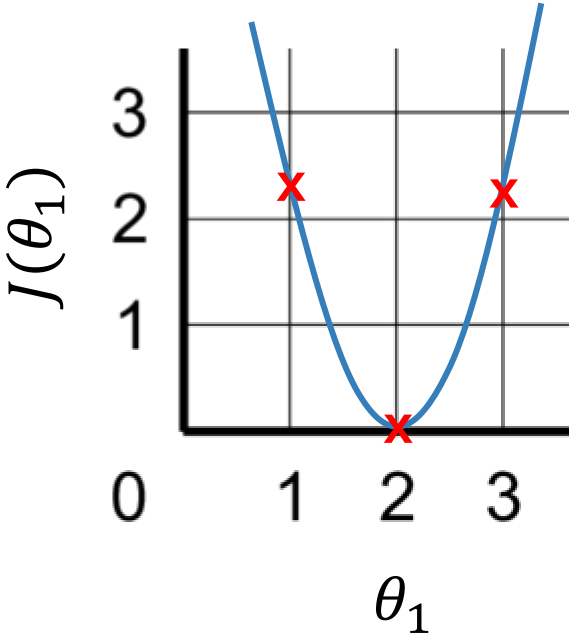
Minimizing the Cost Function

Input x	Output y
0	0
1	2
2	4
3	6

Suppose $\theta_0 = 0$.

1. Plot $h_\theta(x)$ when:
   - $\theta_1 = 1$
   - $\theta_1 = 2$
   - $\theta_1 = 3$
2. Solve for:
   - $J(\theta_1 = 1)$
   - $J(\theta_1 = 2)$
   - $J(\theta_1 = 3)$

$J(\theta)$ vs θ plot

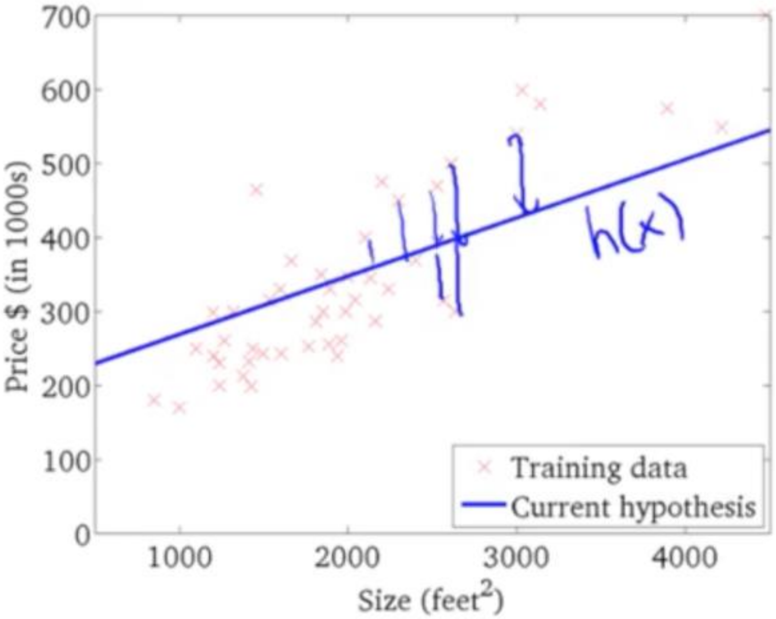


- A parabola makes sense, since J is a quadratic function.
- The value of θ_1 that minimizes J is 2.

Minimizing J

$h_\theta(x)$

(for fixed θ_0, θ_1, this is a function of x)



In general,

the values of the weights θ_1, θ_0 that minimize $J(\theta_1, \theta_0)$

$\Downarrow$

minimizes the distance between $h(x)$ and every data point

Gradient Descent Algorithm

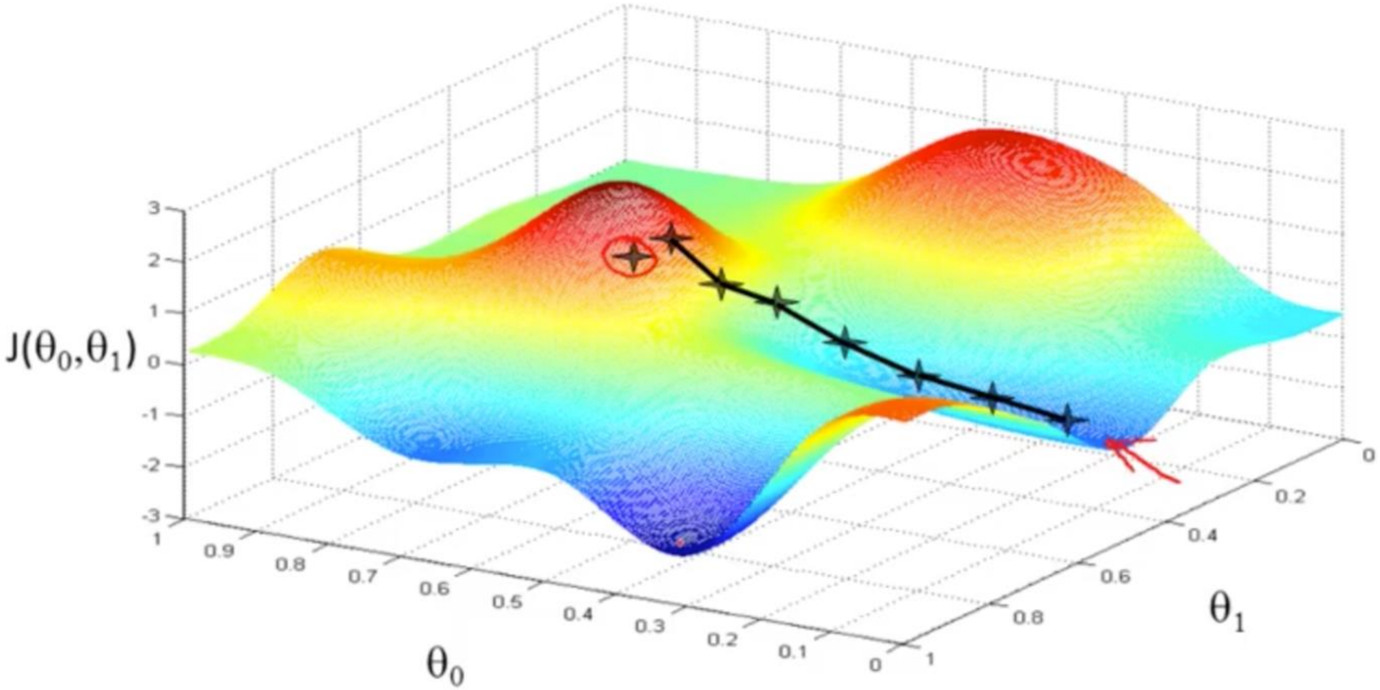
- Have some function $J(\theta_0, \theta_1)$

- Want to find $\min\limits_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

Gradient Descent Algorithm

- Have some function $J(\theta_0, \theta_1)$
- Want to find $\min\limits_{\theta_0, \theta_1} J(\theta_0, \theta_1)$
- Outline of Gradient Descent Algorithm
    - Start with some θ_0, θ_1
    - Keep changing θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$ until we hopefully end up at the minimum.

Gradient Descent Algorithm



Gradient Descent Algorithm

- We want to take steps down the cost function in the direction of the steepest descent
  - The **direction** is given by the **derivative of the cost function**:

$$\frac{\partial}{\partial \theta_j} J(\theta_j) \qquad (\text{for } j = 0 \text{ and } j = 1)$$

  - The **size** of of each step is given by some **learning rate** α.

Gradient Descent Algorithm

repeat until convergence:

$$\theta_j := \theta_j - \underbrace{\alpha}_{\text{Learning rate}} \underbrace{\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)}_{\text{Derivative of the cost function}} \quad (\text{for } j = 0 \text{ and } j = 1)$$

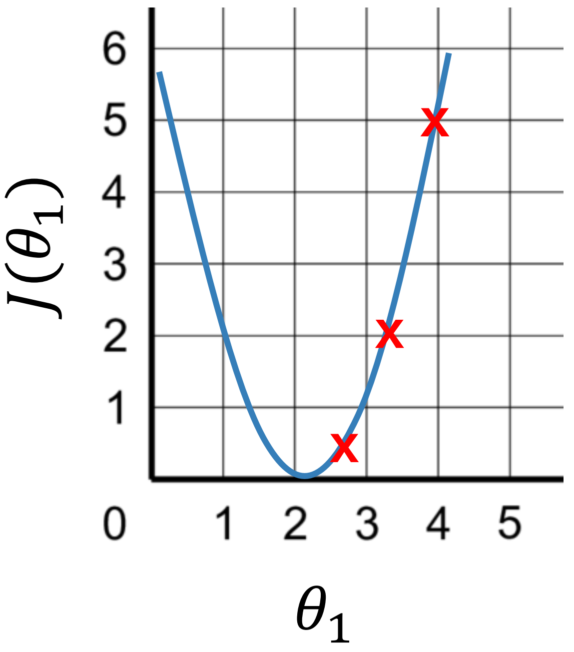
Learning rate
(gives the step size of the descent)

Derivative of the cost function
(gives the direction of the descent)

- a:= b (this means assignment)
- a = b (truth assertion)

Let’s talk about derivatives for a sec.

Derivatives

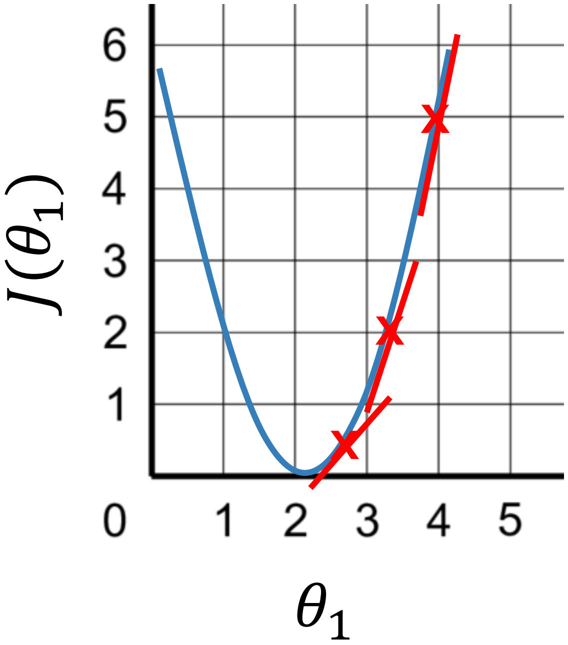


To get to the minimum value, we take the **derivative** of our cost function.

$$\frac{\partial}{\partial \theta_1} J(\theta_1)$$

Derivative – the slope of the tangent line at a point

Derivatives

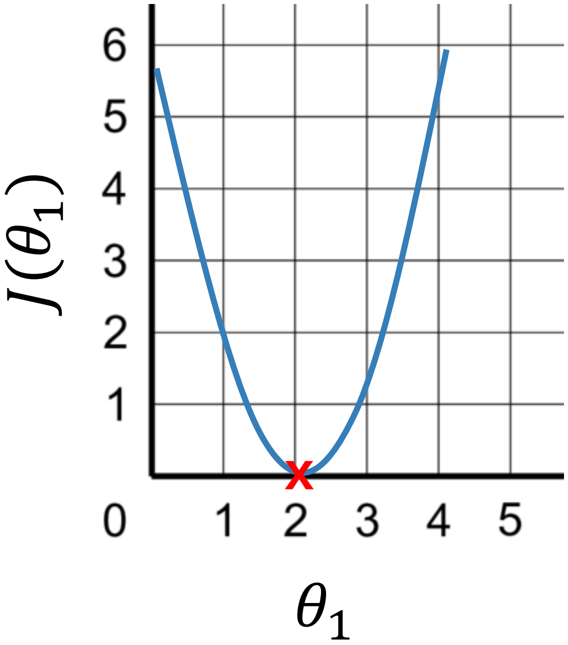


To get to the minimum value, we take the **derivative** of our cost function.

$$\frac{\partial}{\partial \theta_1} J(\theta_1)$$

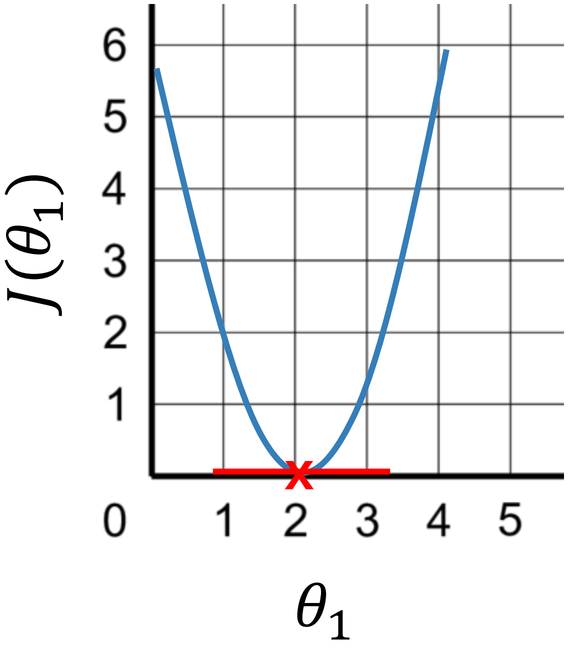
Derivative – the slope of the tangent line at a point

Derivatives



What is the slope at the minimum of J?

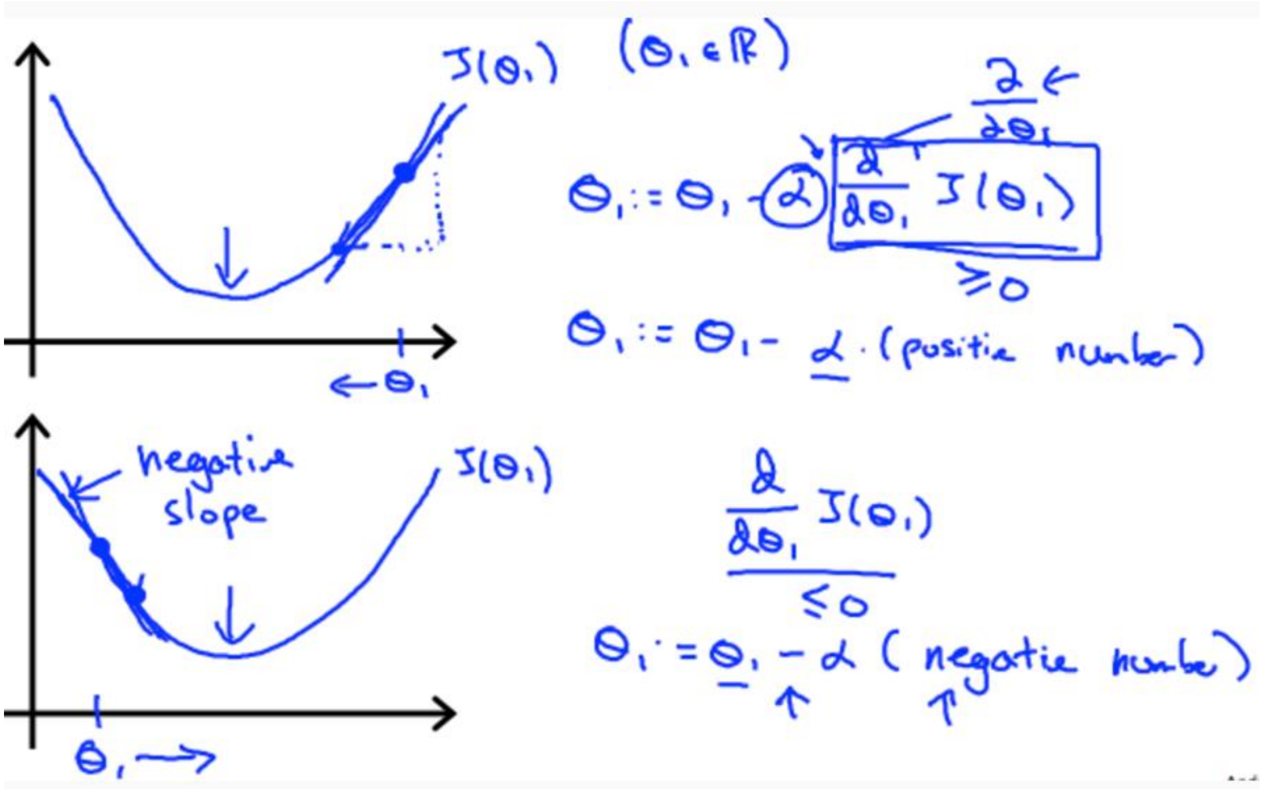
Derivatives



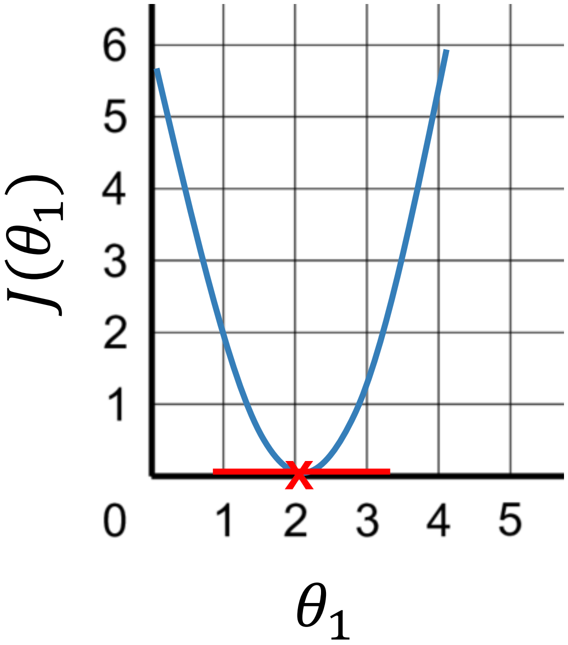
What is the slope at the minimum of J?

The slope given by $\frac{\partial}{\partial \theta_1} J(\theta_1)$ will always be zero!

Derivative Intuition



Covergence



We say that our algorithm has converged when

$$\frac{\partial}{\partial \theta_1} J(\theta_1) = 0$$

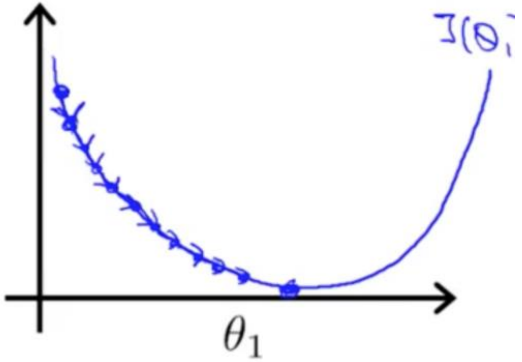
$$\theta_1 := \theta_1 - \alpha(0)$$

θ_1 ceases to update to a new value.

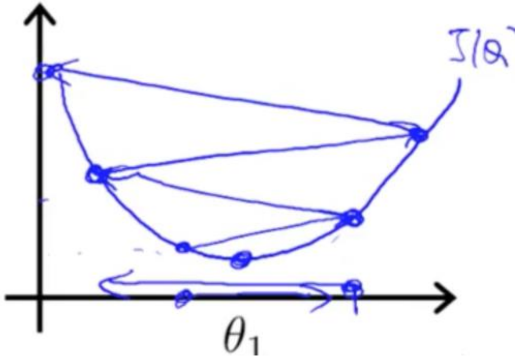
Learning Rate Intuition

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

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.



Gradient Descent Algorithm

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad (\text{for } j = 0 \text{ and } j = 1)$$

}

Correct: Simultaneous update

$$\text{temp0} := \theta_0 - \alpha \tfrac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$
$$\text{temp1} := \theta_1 - \alpha \tfrac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$
$$\theta_0 := \text{temp0}$$
$$\theta_1 := \text{temp1}$$

Gradient Descent Algorithm

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad (\text{for } j = 0 \text{ and } j = 1)$$

}

Correct: Simultaneous update

$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$
$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$
$\theta_0 := \text{temp0}$
$\theta_1 := \text{temp1}$

Incorrect:

$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$
$\theta_0 := \text{temp0}$
$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$
$\theta_1 := \text{temp1}$

Gradient Descent Algorithm

repeat until convergence: {

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} ((h_\theta(x_i) - y_i) x_i)$$

}

Partner/Group/Individual Exercise

Partner/Group/Individual Presentation

References

T.I.L.

SHARE IT!
In front!

On Twitter: @wwcodemanila
Or FB: fb.com/wwcodemanila

Don’t forget to tag WWCodeManila so we can retweet or share it.

Feedback Form

https://goo.gl/YzSqcS

Please don’t rate the event on meetup.
Not helpful. It is best to just tell your concerns via the feedback form. We are a building a community not a Yelp restaurant.

THANK YOU :)