MACHINE LEARNING

اَللّٰهُمَّ اُرْزُقْنِى عِلْمًا نَافِعًا وَاسِعًا عَمِيْقًا

اَللّٰهُمَّ اُرْزُقْنِی رِزْقًا وَاسِعًا حَلَالًا طَيِّبًا
مُبَارَكًا مِنْ عِنْدِكَ

WEEK 05

TYPE OF MACHINE LEARNING

- Supervised Learning.
- Unsupervised Learning.
- Reinforcement Learning.

SUPERVISED LEARNING

Living area (feet^2)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮

SUPERVISED LEARNING: REGRESSION

- When we try to predict a number from historical data, this type of supervised learning problem is called Regression Problem

LINEAR REGRESSION

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

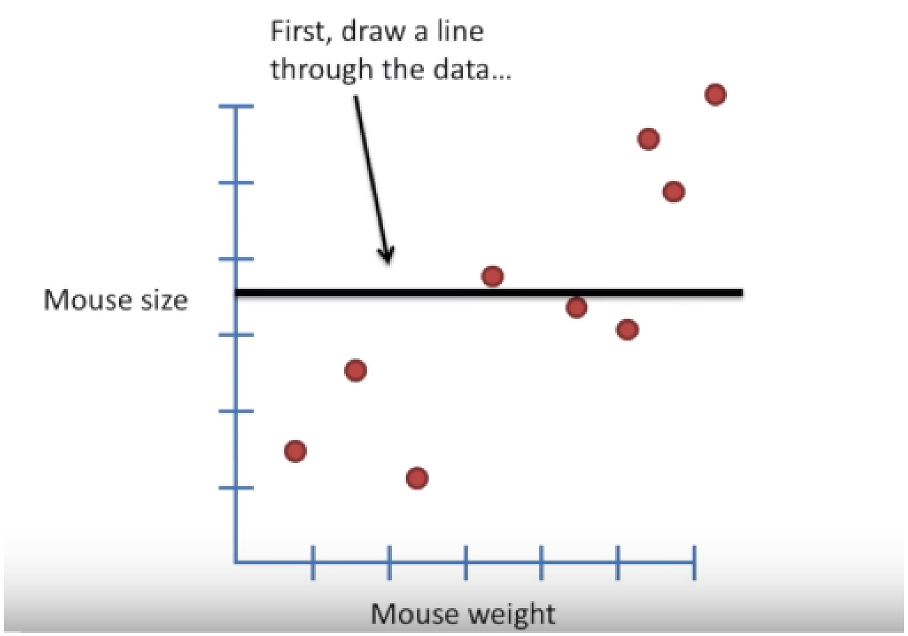
we need to find the parameter value of θ_0 and θ_1 that suits best according to the given data.

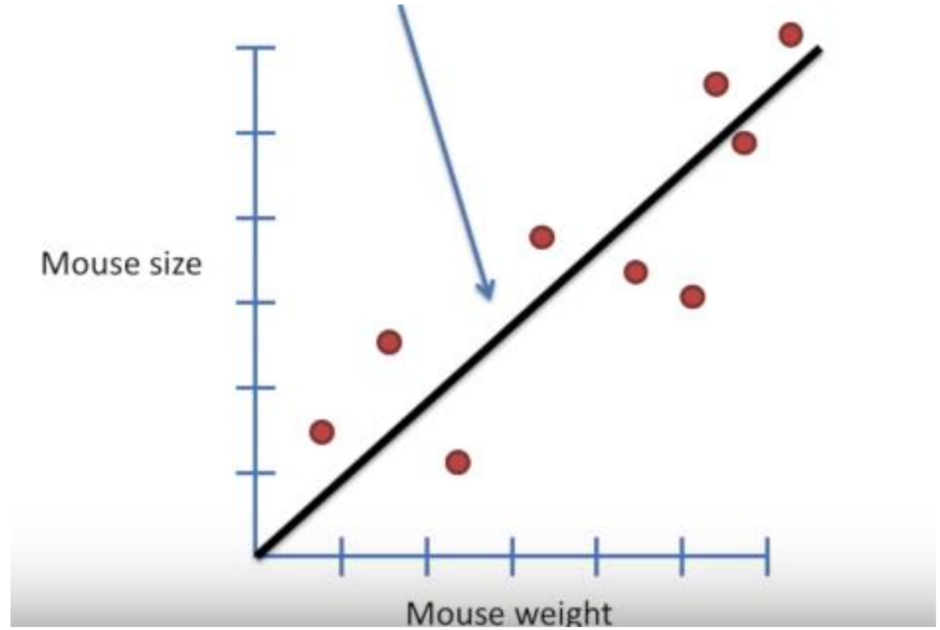
LINEAR REGRESSION

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

Therefore, sometime the univariate linear regression problem also seen as to find equation of line.

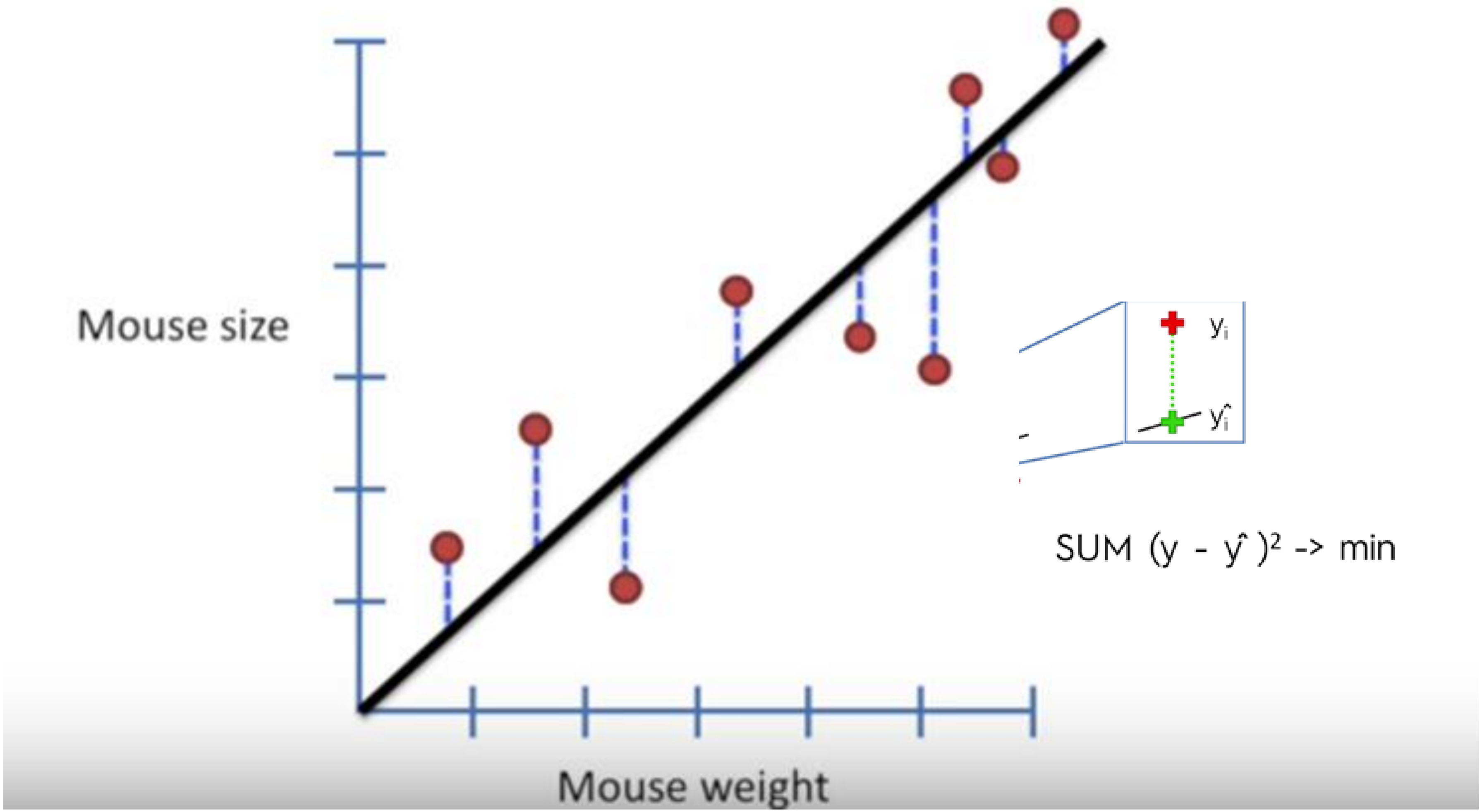
WHAT IS BEST LINE (LINEAR FUNCTION) ?





LEAST SQUARE: ONE WAY TO FIND ERROR

REVIEW



LOSS FUNCTION

- The function that find the error is called the LOSS function. This calculate the Loss of prediction for single example.

- $J(\theta) = (y^i - f(x^i))^2$

COST FUNCTION

- Over all cost for all examples, this function is called cost function.

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

OUR OBJECTIVE

- So finally, our objective is to find the value of θ_0 and θ_1 such that the value of $J(\theta)$ is minimized.

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

Remember here $f(x^i) = \theta_0 + \theta_1 x^i$

OUR OBJECTIVE

- So finally, our objective is to find the value of θ_0 and θ_1 such that the value of $J(\theta)$ is minimized.

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

REIVEW

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

Parameters: θ_0, θ_1

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

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

PROGRAMMING ASSIGNMENT 03

- Extend the previous assignments, write a program that read the data and proposed the linear regression parameters.
- (You can add range constrain on parameters between 1 to 10).
- Hint you need to change the first parameter value and keep the second parameter constant

PROGRAMMING ASSIGNMENT 05

- Repeat the Experiment for following ranges of parameters and fill the table.

Range	Number of Iterations	Time in Seconds
1 to 10		
1 to 100		
1 to 1000		
1 to 10000		

PROGRAMMING ASSIGNMENT 05

- One student Work (Reasonable Okay).

Range	Number of Iterations	Time in Seconds
1 to 10	100	0.0084
1 to 100	10,000	0.2655
1 to 1000	1,000,000	18.8086
1 to 10000	100,000,000	1372.3848

PROGRAMMING ASSIGNMENT 05

- What wrong with it ?

Range	Number of Iterations	Time in Seconds
1 to 10	100	0.0084
1 to 100	10,000	0.2655
1 to 1000	1,000,000	18.8086
1 to 10000	100,000,000	1372.3848

PROGRAMMING ASSIGNMENT 05

- What is run time of the algorithm ?

Range	Number of Iterations	Time in Seconds
1 to 10	100	0.0084
1 to 100	10,000	0.2655
1 to 1000	1,000,000	18.8086
1 to 10000	100,000,000	1372.3848

PROGRAMMING ASSIGNMENT 05

- Run time
  - (Number of possible_value)$^{\text{number of variables}}$
  - $10^{n} = 10^{2} = 100$

Range	Number of Iterations	Time in Seconds
1 to 10	100	0.0084
1 to 100	10,000	0.2655
1 to 1000	1,000,000	18.8086
1 to 10000	100,000,000	1372.3848

PROGRAMMING ASSIGNMENT 05

- We are not checking the value in points 1.1,1.2 or even 1.21,1.22 how many iteration increased in that way ?

Range	Number of Iterations	Time in Seconds
1 to 10		
1 to 100		
1 to 1000		
1 to 10000		

PROGRAMMING ASSIGNMENT 05

- We are not checking -negative values how many iteration increased in that way ?

Range	Number of Iterations	Time in Seconds
1 to 10		
1 to 100		
1 to 1000		
1 to 10000		

PROGRAMMING ASSIGNMENT 05

- If the parameters increased. Currently it is one variable what if it increases to 20 ?

Range	Number of Iterations	Time in Seconds
1 to 10		
1 to 100		
1 to 1000		
1 to 10000		

PROGRAMMING ASSIGNMENT 05

- Remember Life of Universe is 10^{16} Seconds.

Range	Number of Iterations	Time in Seconds
1 to 10		
1 to 100		
1 to 1000		
1 to 10000		

TODAY LECTURE

THE PROBLEM

- Objective is to Minimize

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

THE PROBLEM

- Objective is to Minimize

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

- But Brute Force is not a feasible solution. Need some better method.

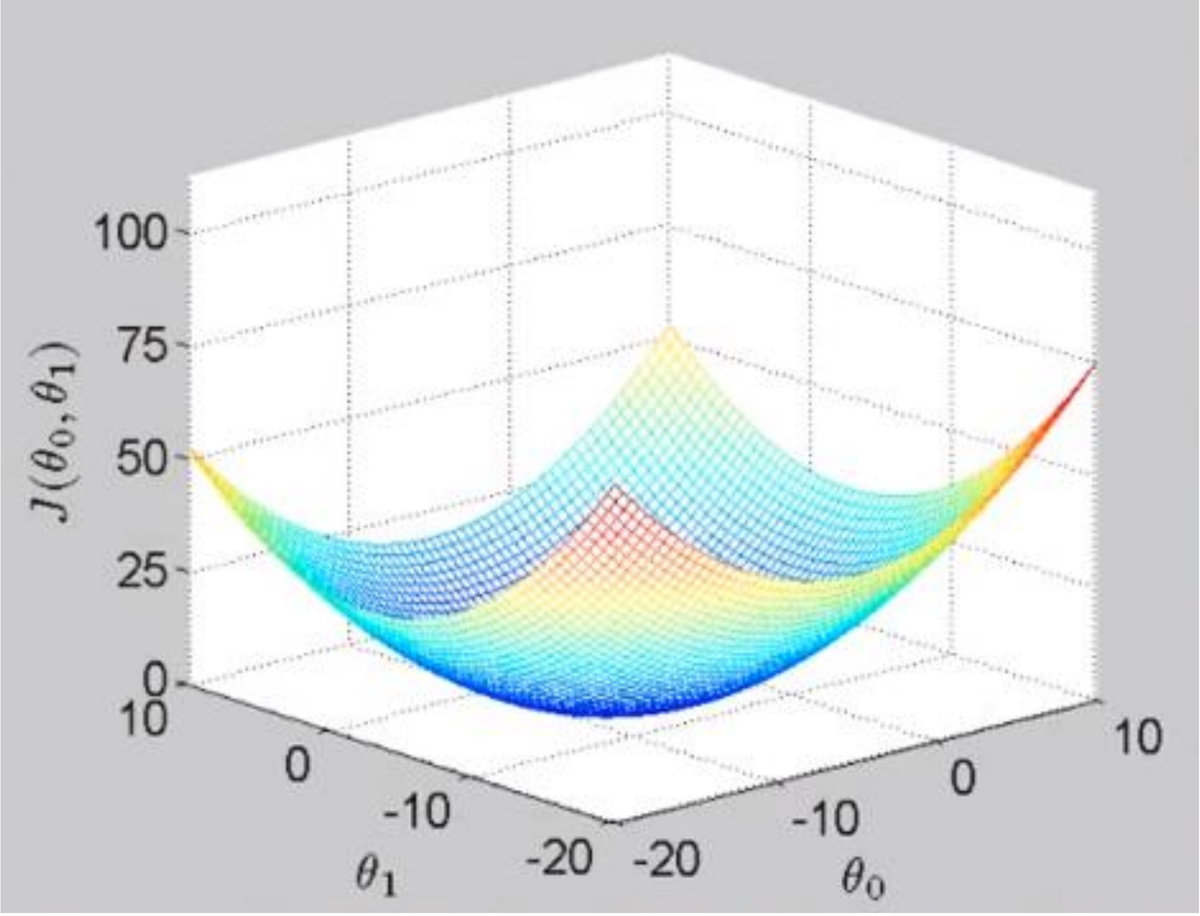
THE PROBLEM

- Let see the function Graphically.

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

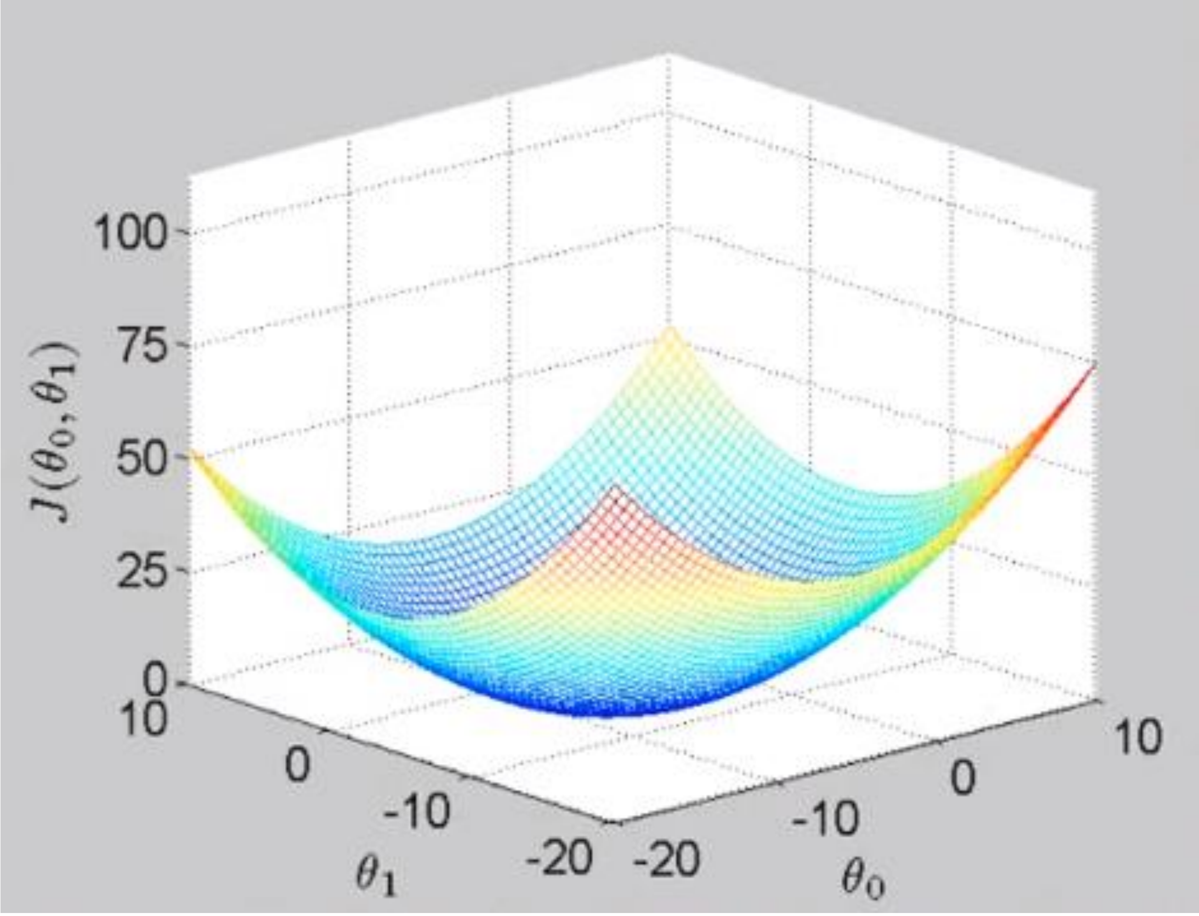
THE PROBLEM

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



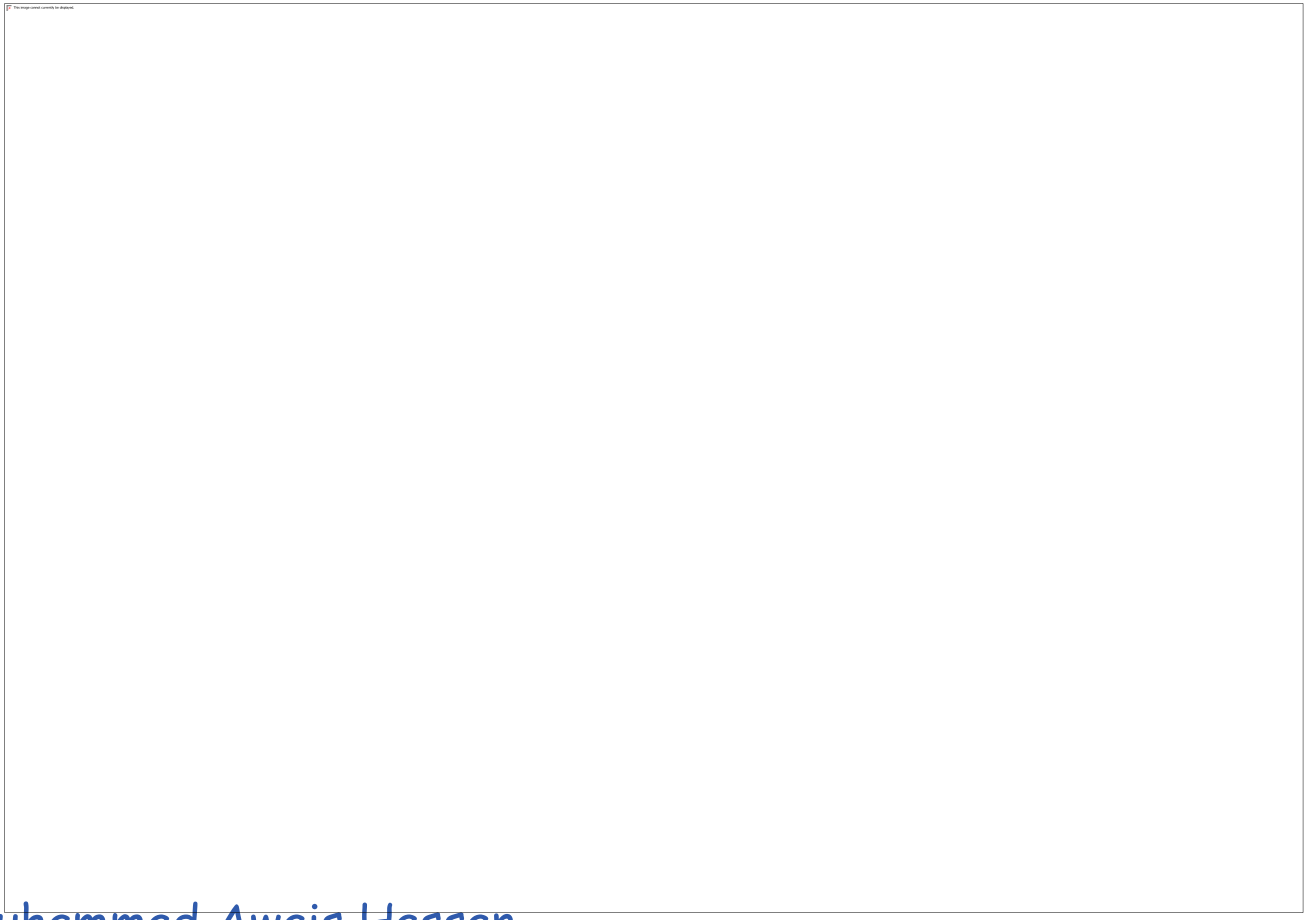
THE PROBLEM

Can we proposed an algorithm that find the value of θ_0 and θ_0 were $J(\theta)$ is minimum



SHAPE OF SURFACE DEPENDS ON DATA

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



GOOD NEWS

- Yes, We can write an algorithm for that.

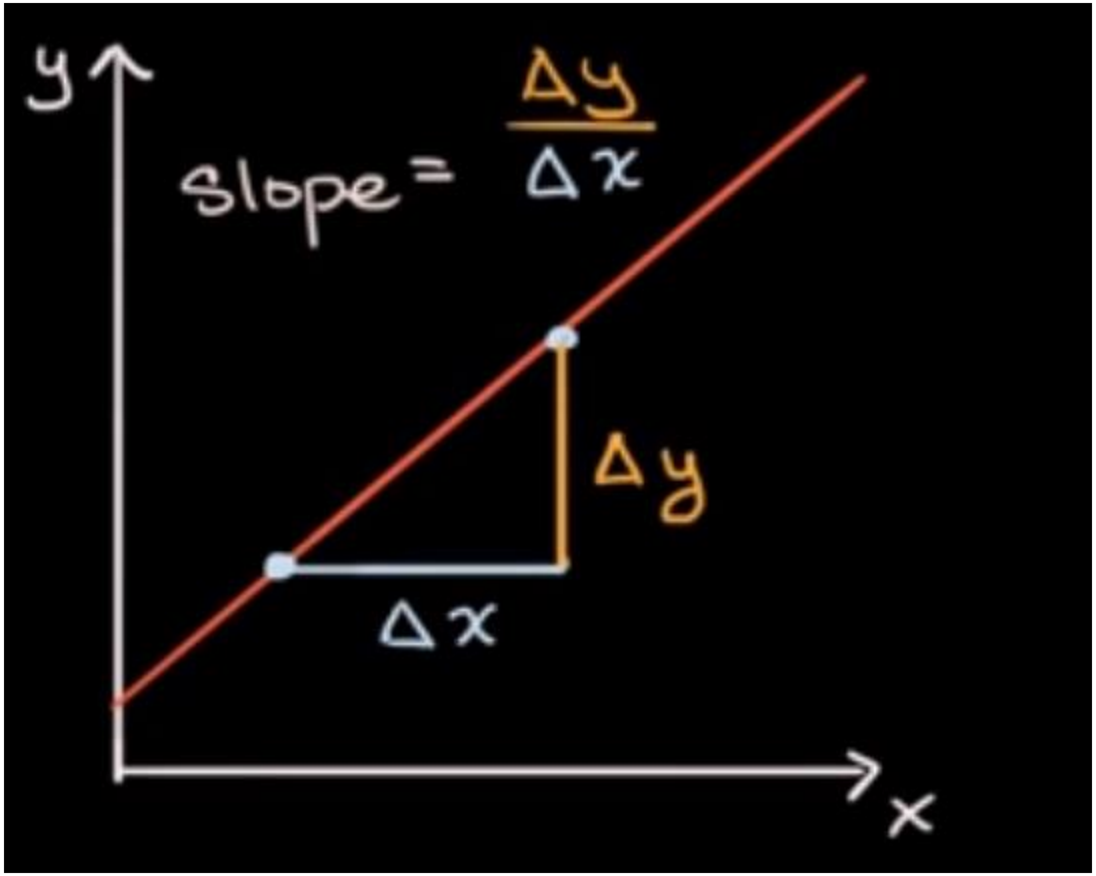
BAD NEWS

- But, we need to review math basics before doing that.

MATH REVIEW

SLOPE OF LINE

SLOPE OF LINE

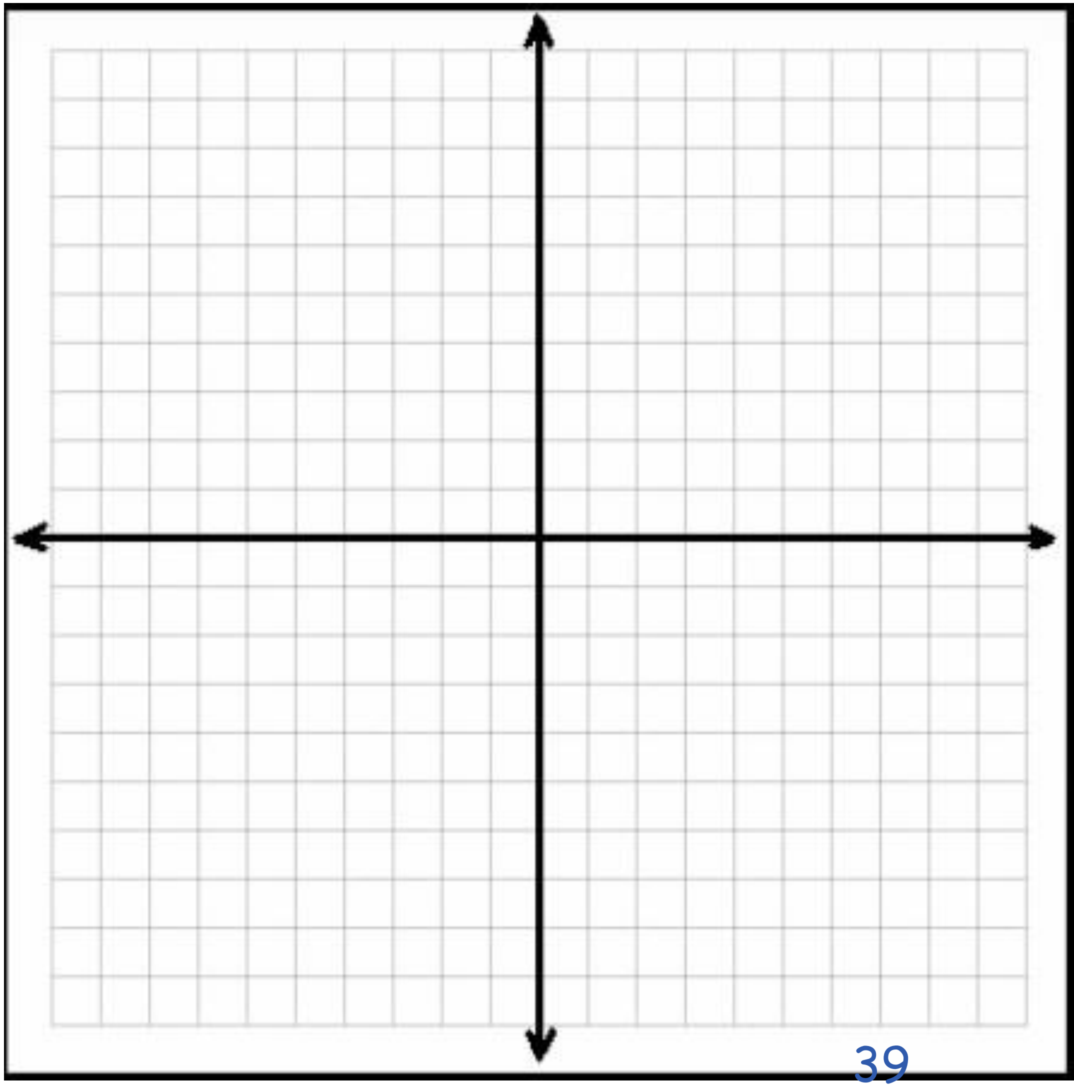


SLOPE INTERCEPT FORM

$$y = mx + c$$

Draw Line for following Equation

$$y = \frac{5}{3}x - 2$$



SLOPE INTERCEPT FORM

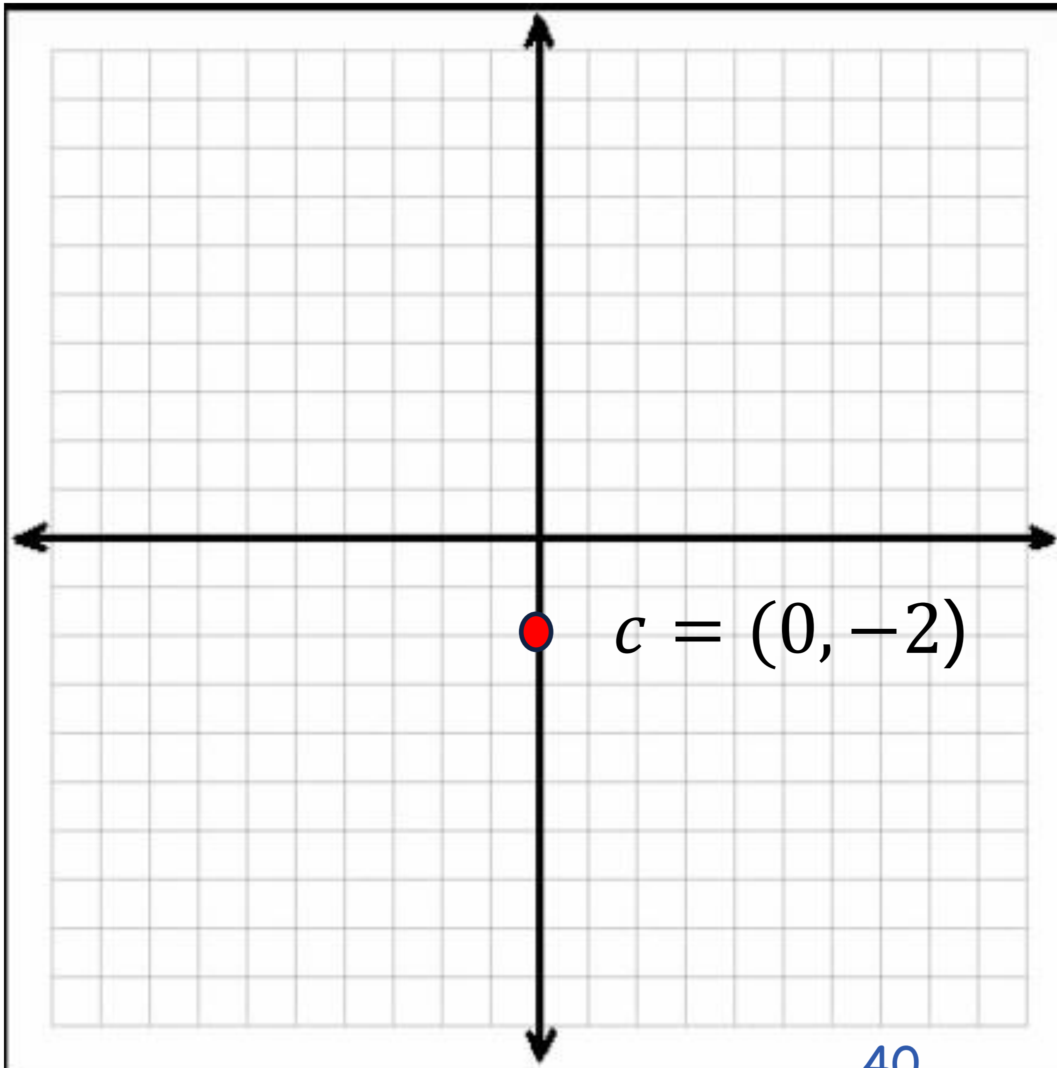
$$y = mx + c$$

$$m = \frac{\Delta Y}{\Delta X}$$

$$y = \frac{5}{3}x - 2$$

$$c = -2$$

$$m = \frac{\Delta Y}{\Delta X} = \frac{5}{3}$$



SLOPE INTERCEPT FORM

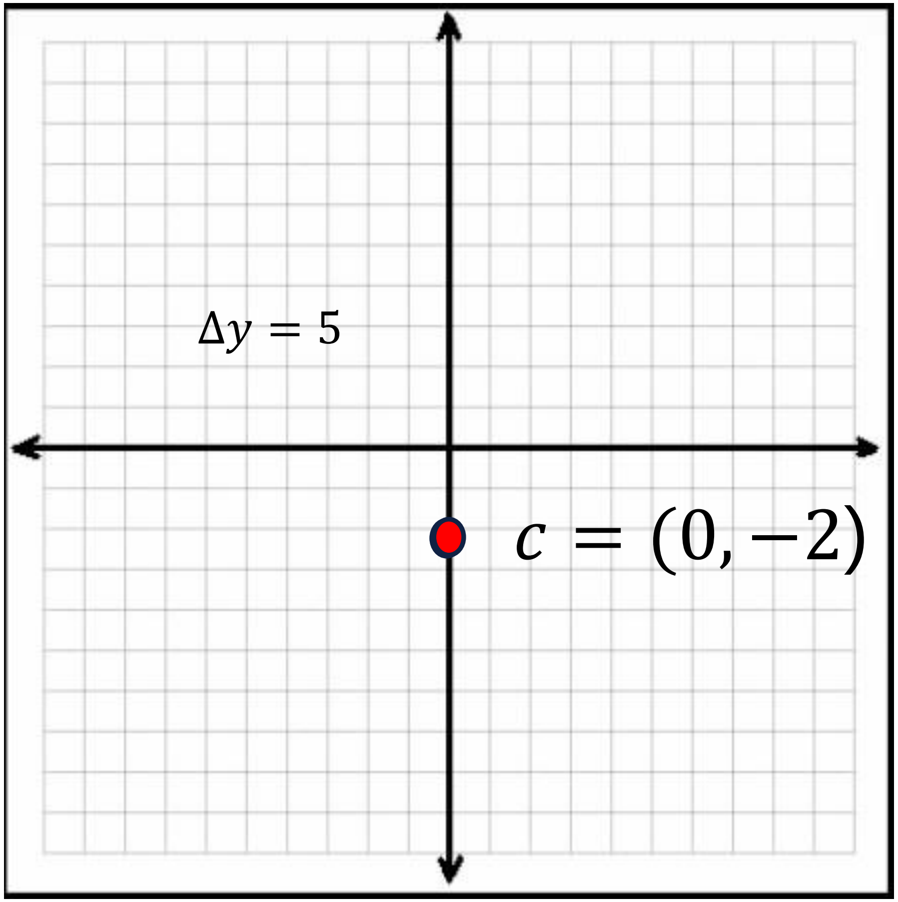
$$y = mx + c$$

$$m = \frac{\Delta Y}{\Delta X}$$

$$y = \frac{5}{3}x - 2$$

$$c = -2$$

$$m = \frac{\Delta Y}{\Delta X} = \frac{5}{3}$$



SLOPE INTERCEPT FORM

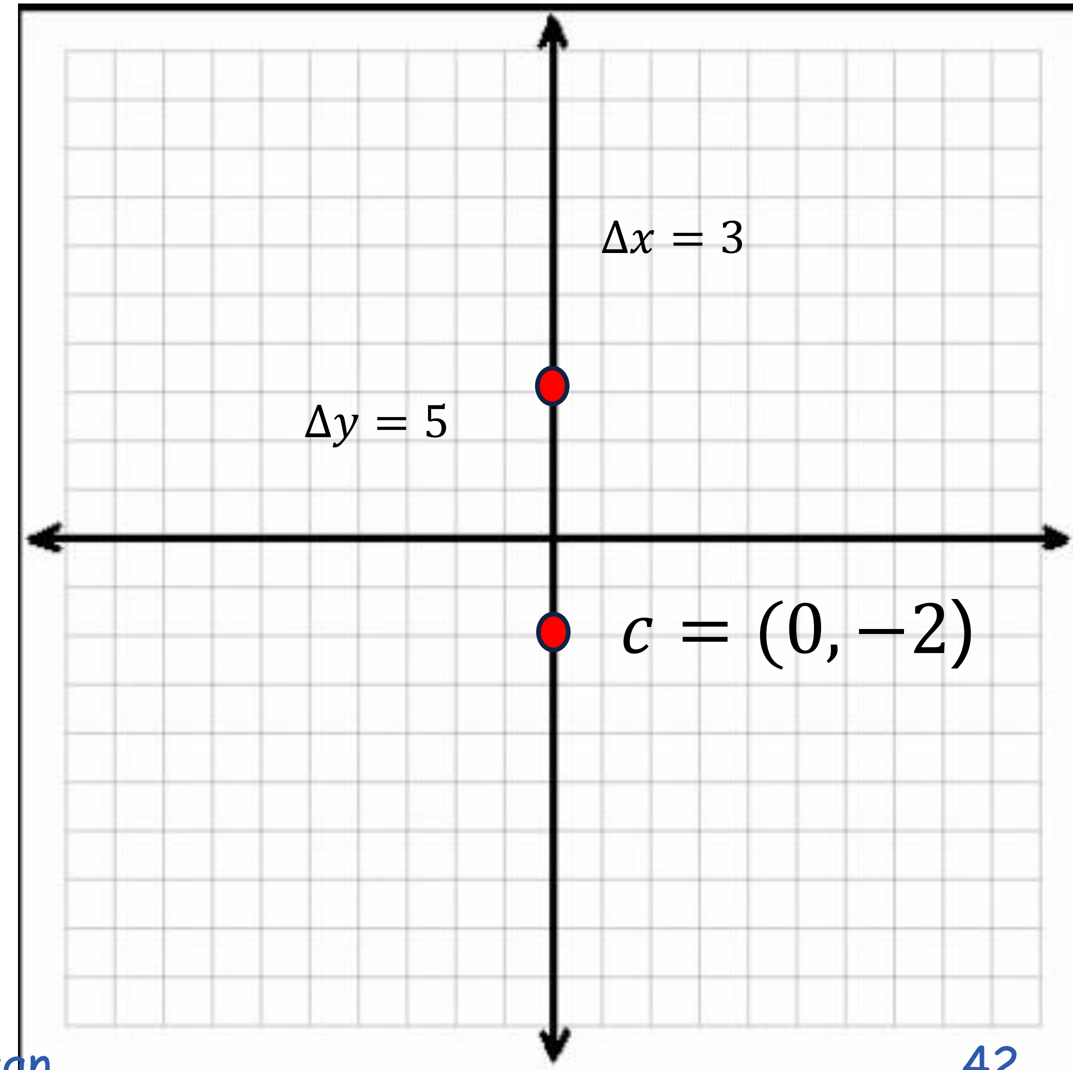
$$y = mx + c$$

$$m = \frac{\Delta Y}{\Delta X}$$

$$y = \frac{5}{3}x - 2$$

$$c = -2$$

$$m = \frac{\Delta Y}{\Delta X} = \frac{5}{3}$$



SLOPE INTERCEPT FORM

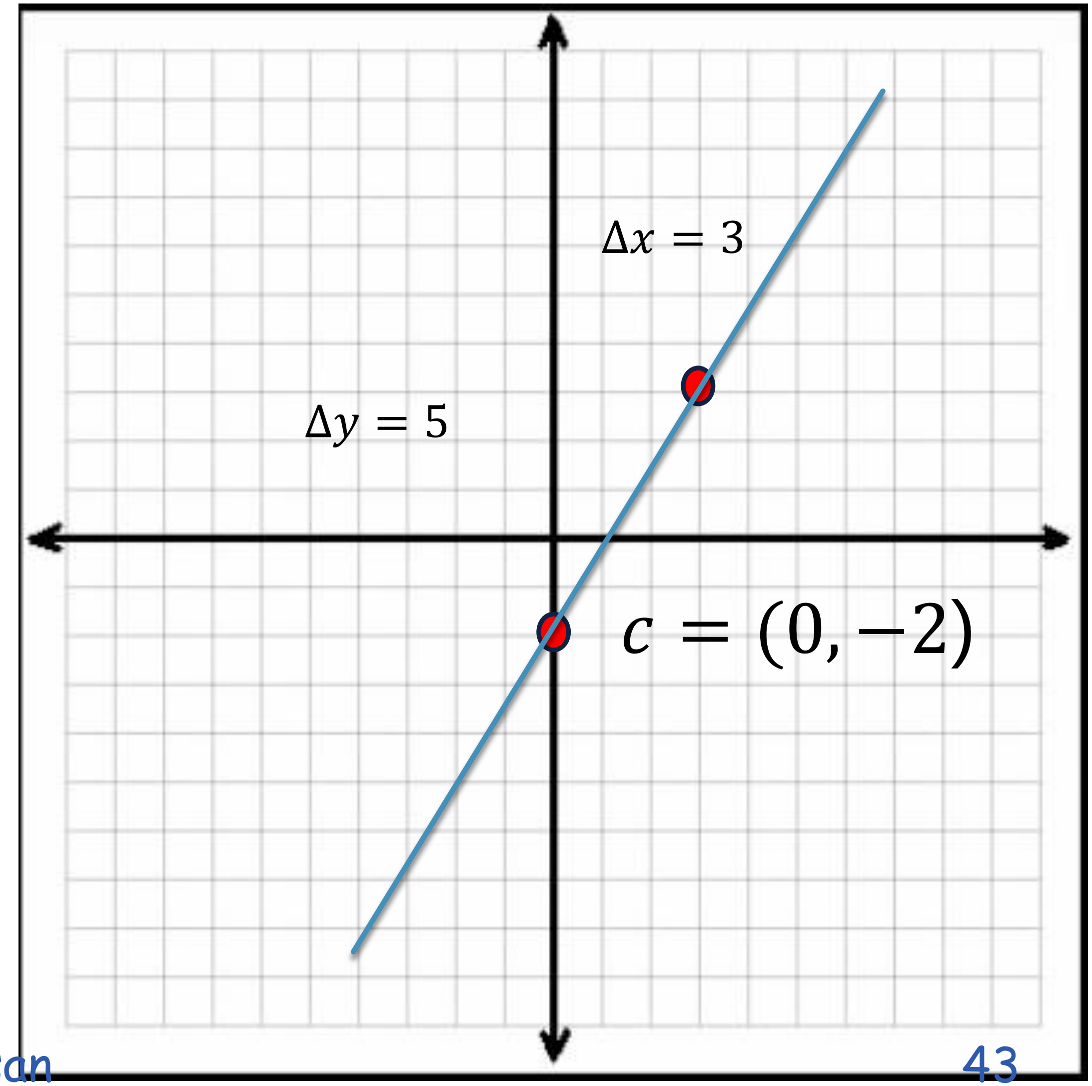
$$y = mx + c$$

$$m = \frac{\Delta Y}{\Delta X}$$

$$y = \frac{5}{3}x - 2$$

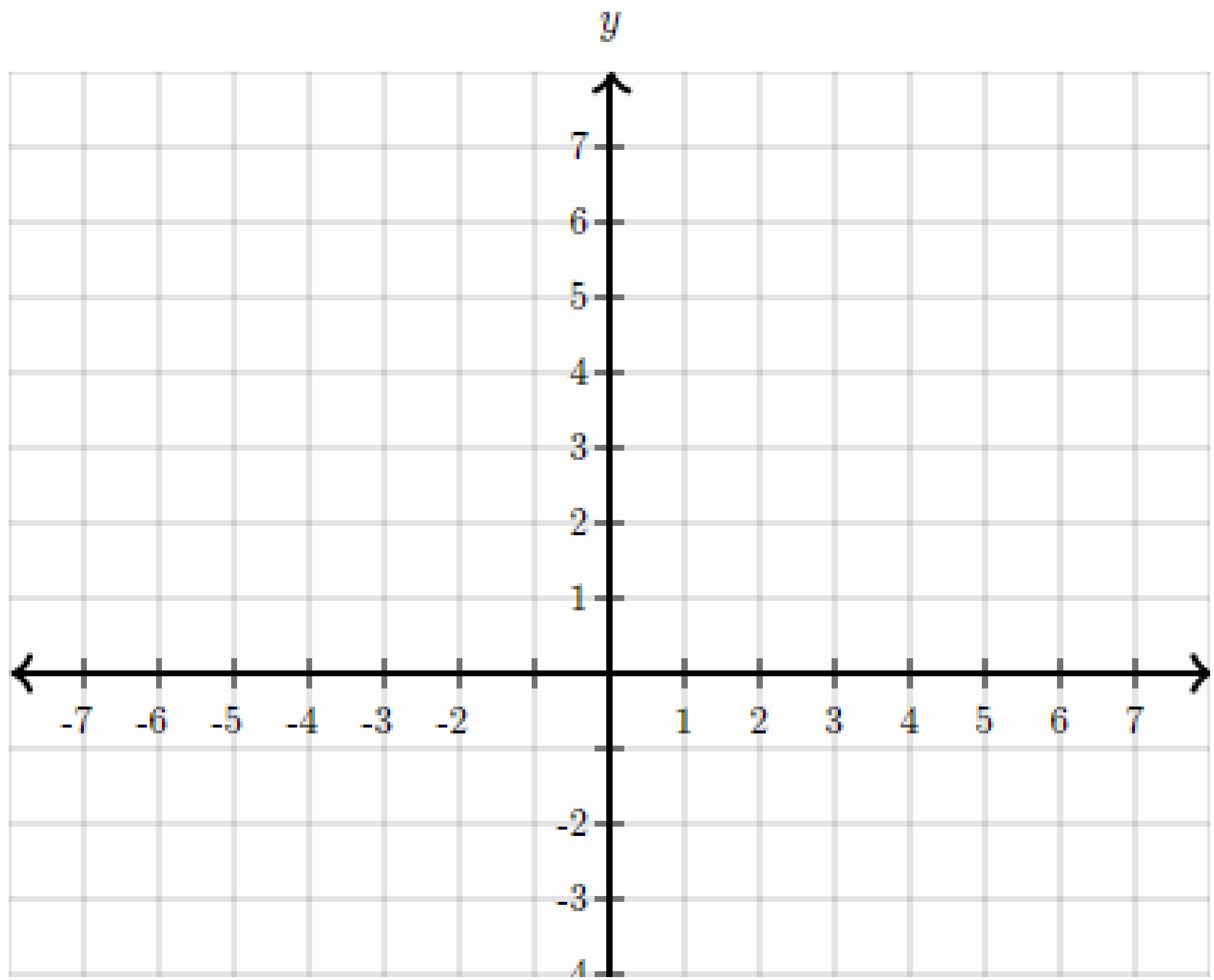
$$c = -2$$

$$m = \frac{\Delta Y}{\Delta X} = \frac{5}{3}$$

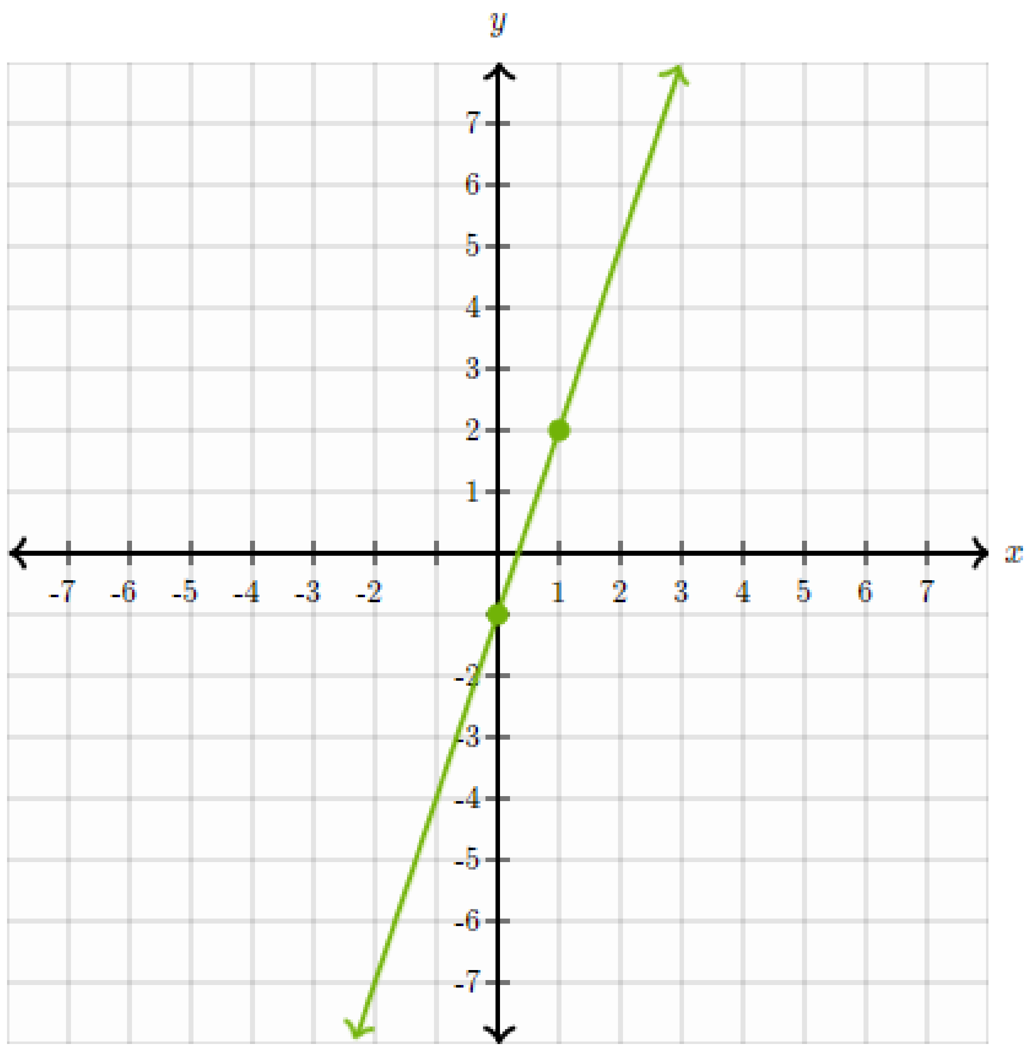


PRACTICE: DRAW LINE

$$y = 3x - 1$$

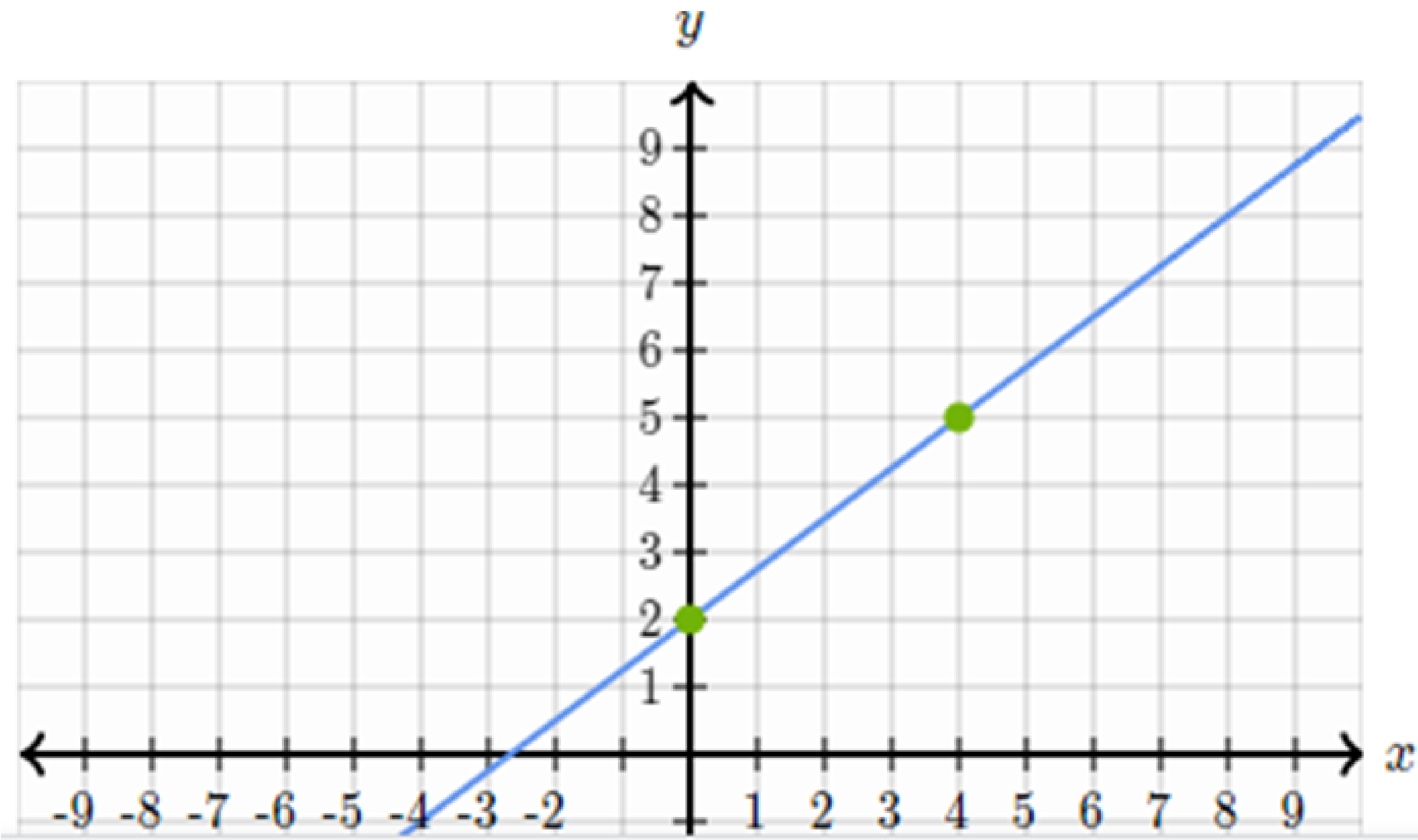


PRACTICE: DRAW LINE Y=3X-1



ONE MORE EXAMPLE

Graph $y = \frac{3}{4}x + 2.$



RATE OF CHANGE

- For Linear Functions, Slope represents the rate of change of function and it is constant.

RATE OF CHANGE OF NON LINEAR FUNCTIONS

AVERAGE RATE OF CHANGE

AVERAGE RATE OF CHANGE

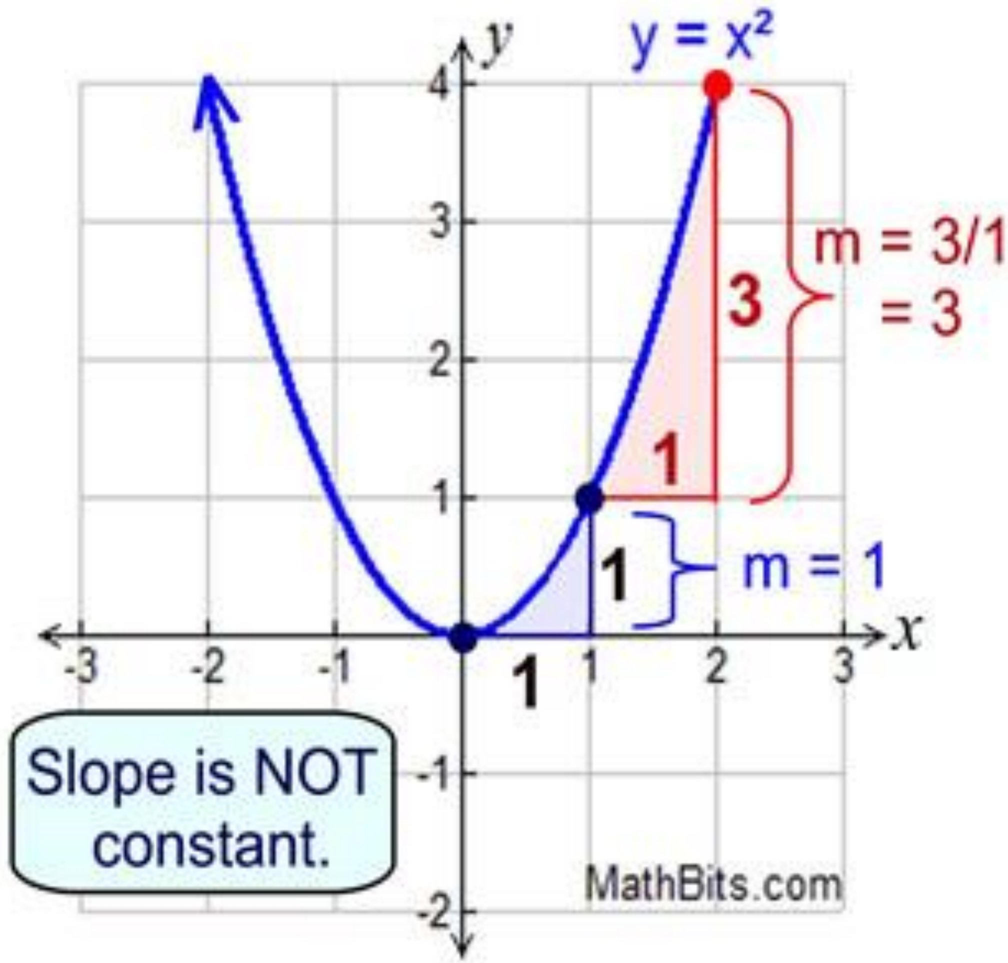
- The process of computing the "average rate of change", however, remains the same as was used with straight lines:

$$\text{slope} = \frac{y_2 - y_1}{x_2 - x_1}$$

- two points are chosen, and is computed.

AVERAGE RATE OF CHANGE

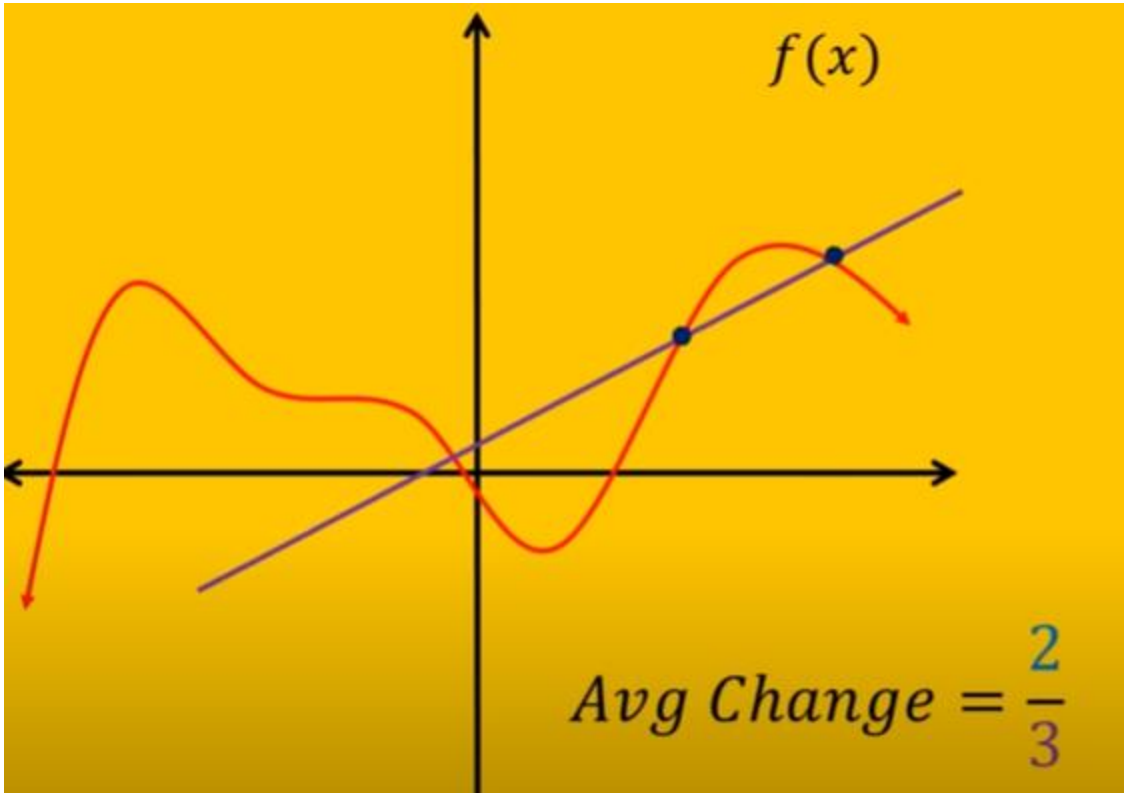
$$\text{slope} = \frac{y_2 - y_1}{x_2 - x_1}$$



Dr. Muhammad Awais Hassan
Department of Computer Science UET, Lahore

AVERAGE RATE OF CHANGE

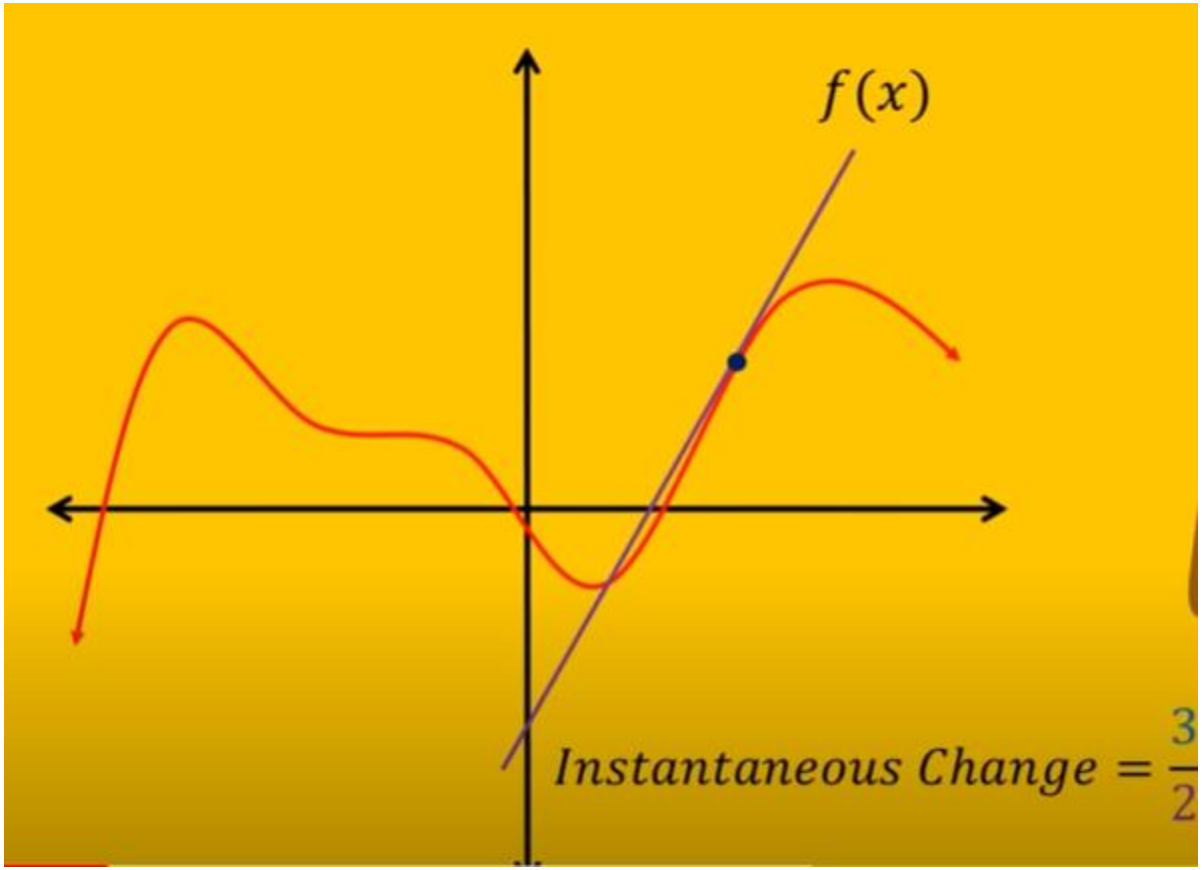
- Average Rate of Change always calculated between two points



INSTANTANEOUS RATE OF CHANGE

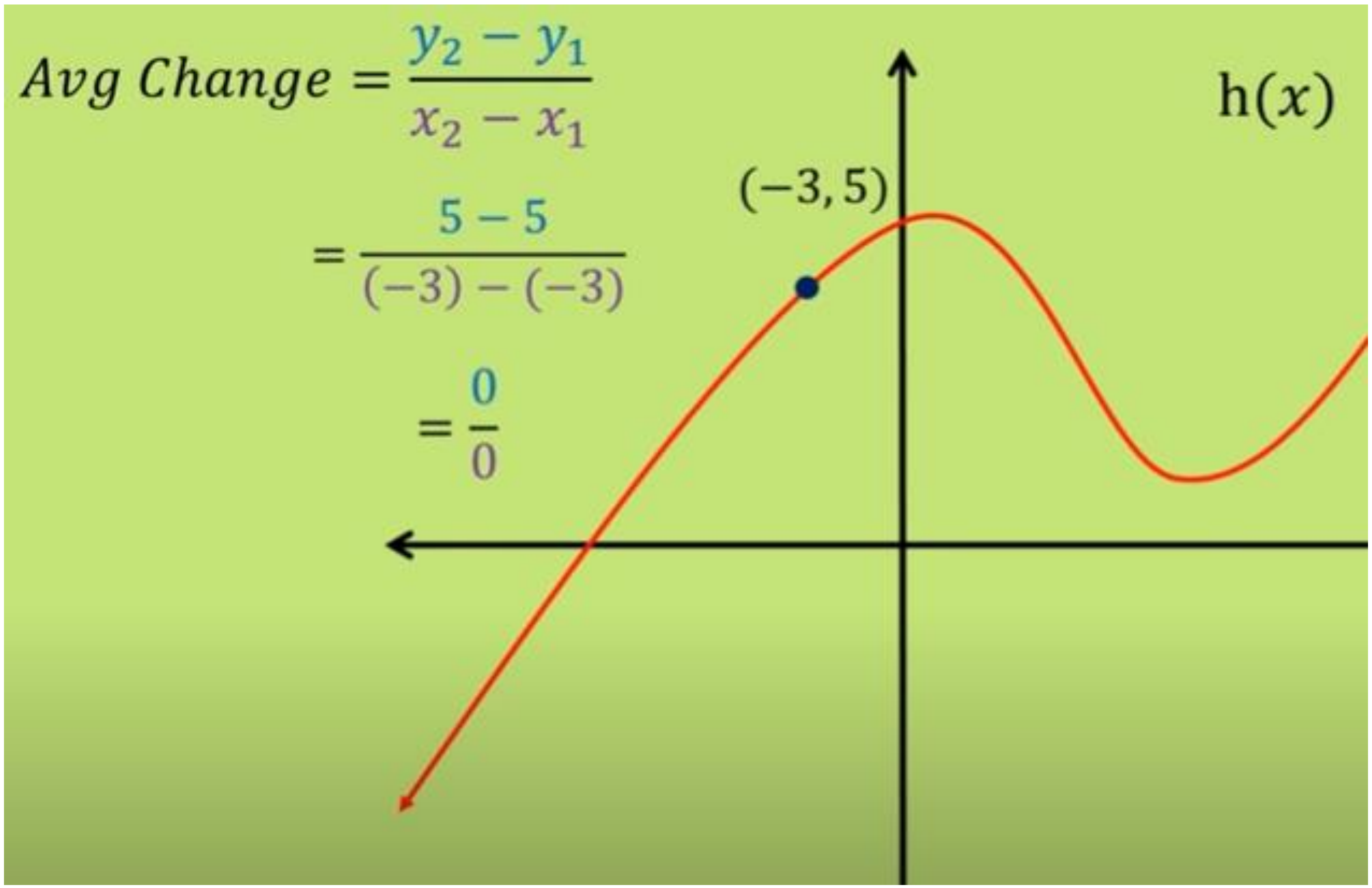
INSTANTANEOUS RATE OF CHANGE

- Instantaneous rate of change is calculated at single point



INSTANTANEOUS RATE OF CHANGE

- But we can not use the average rate formula

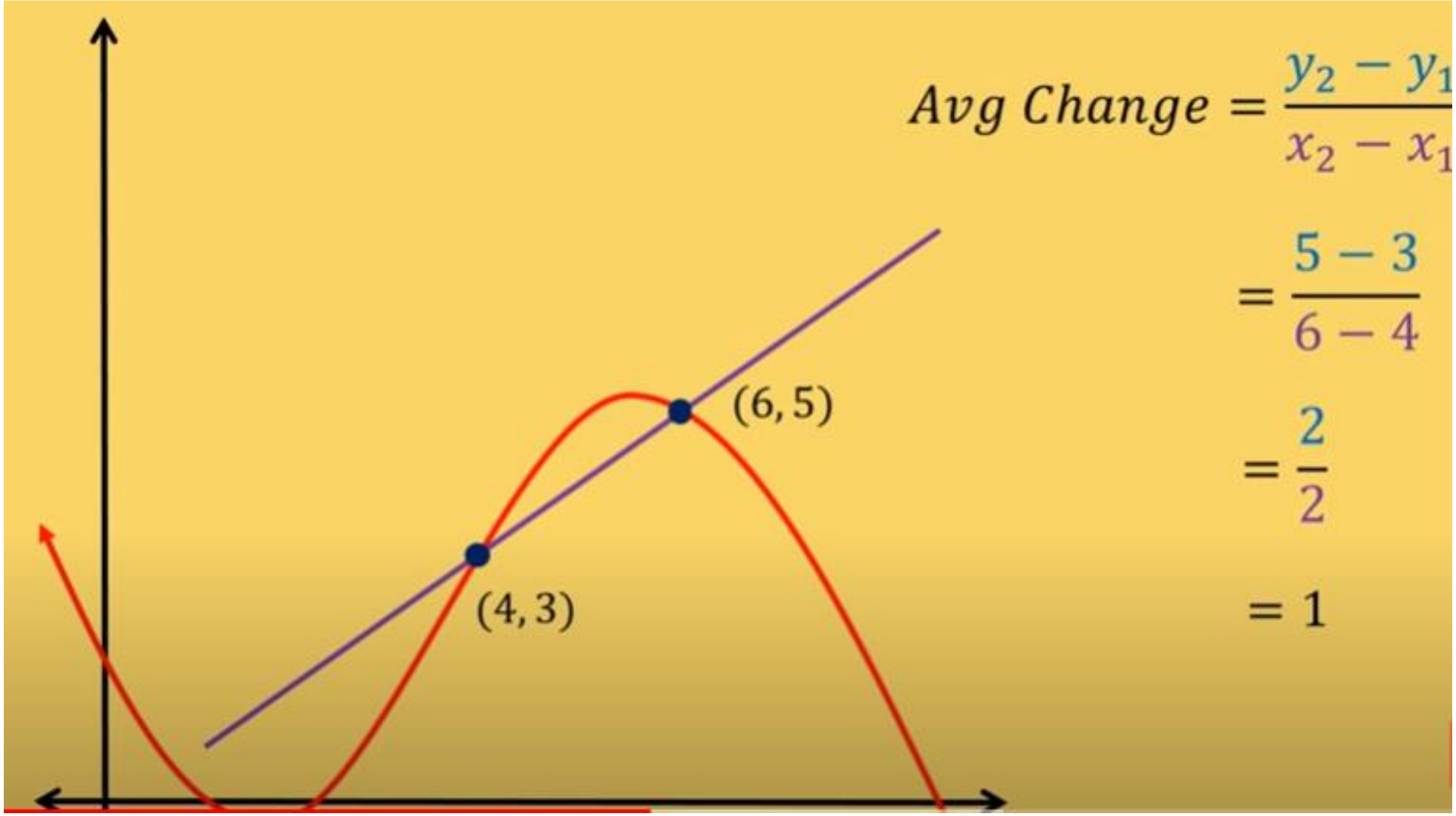


INSTANTANEOUS RATE OF CHANGE

However we use the average rate intelligently to calculate the instantaneous rate of change

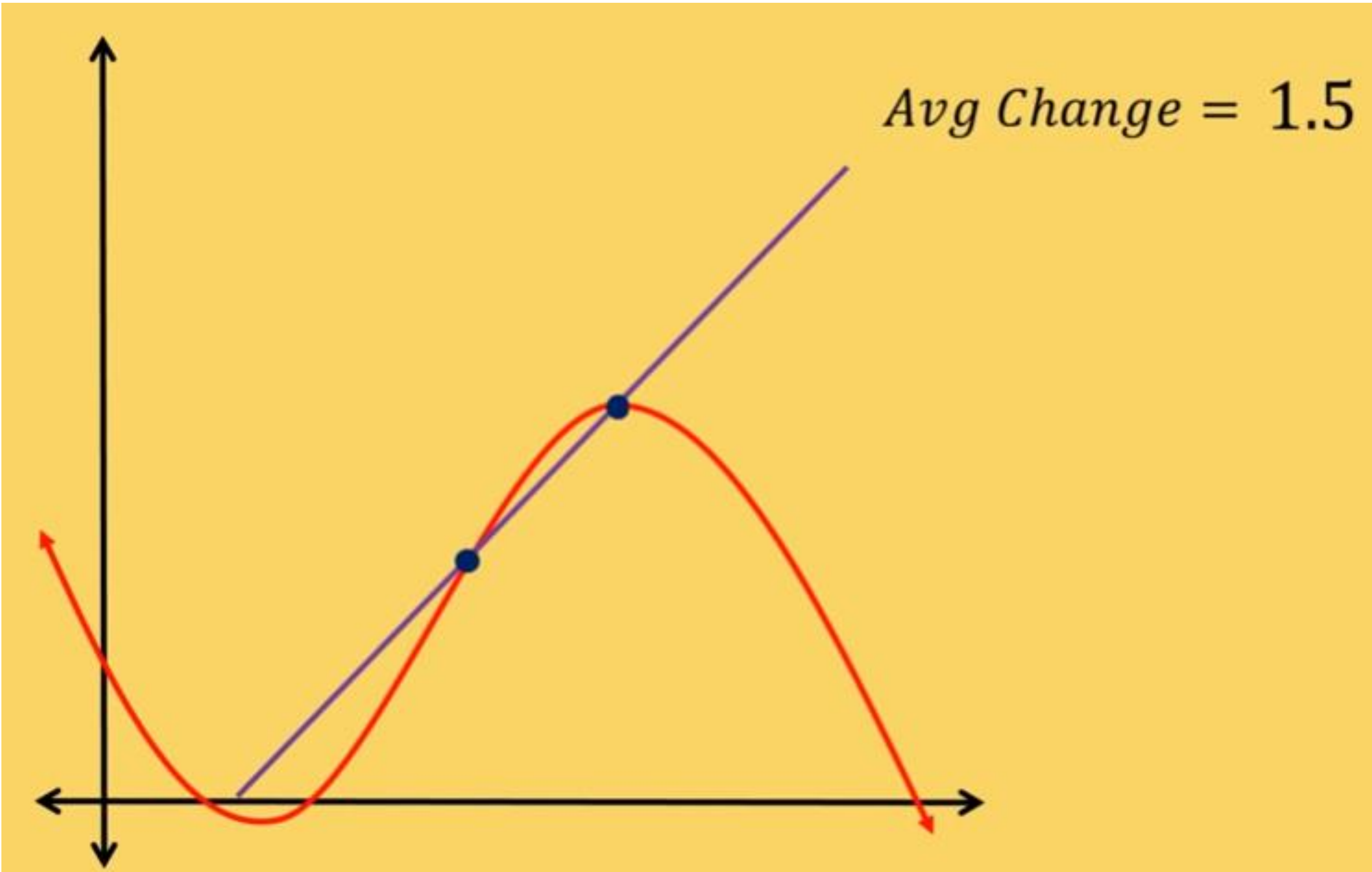
INSTANTANEOUS RATE OF CHANGE

First, we find the average rate between two points.



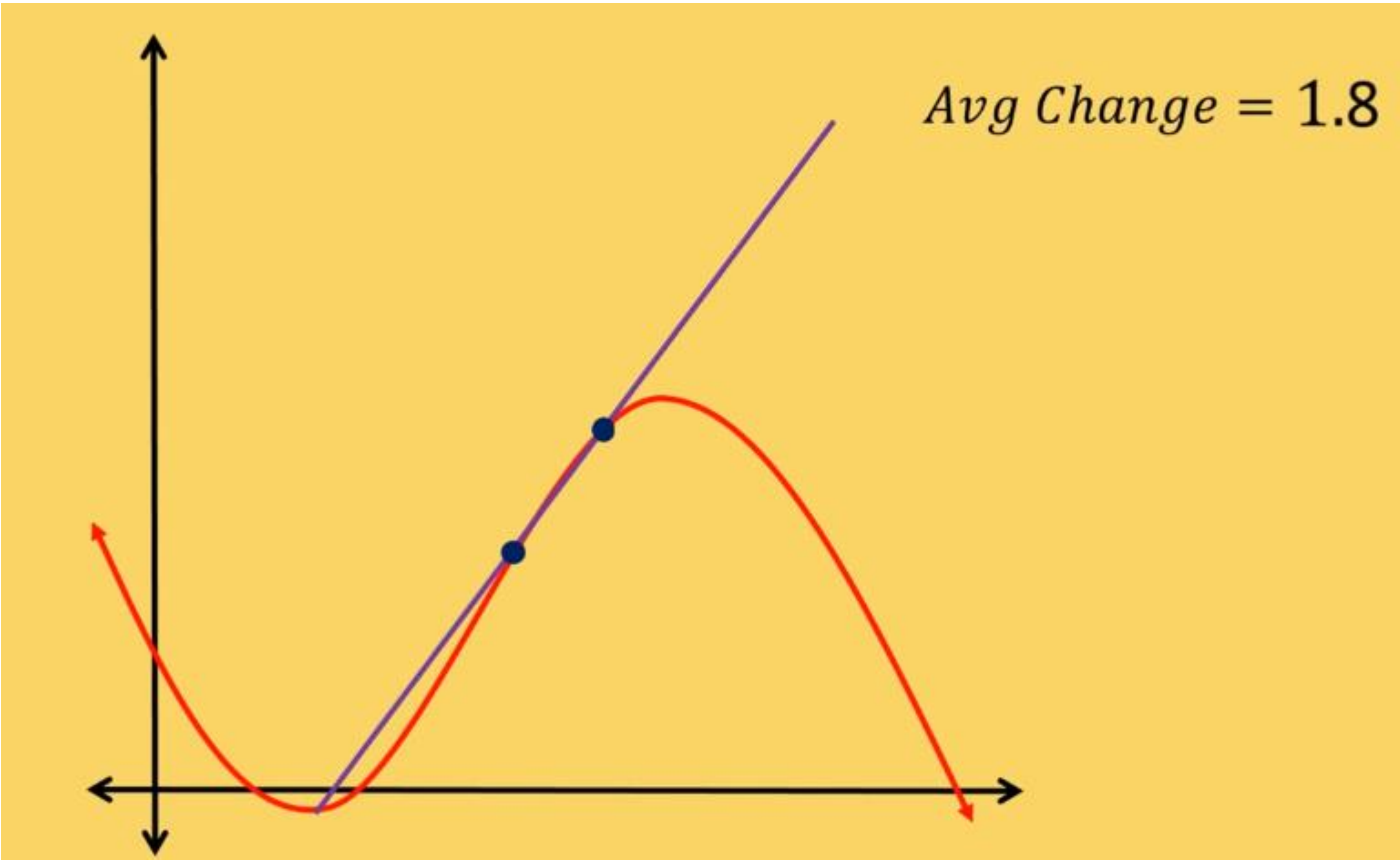
INSTANTANEOUS RATE OF CHANGE

If we move second point **more closer**, we shall get bit different average change



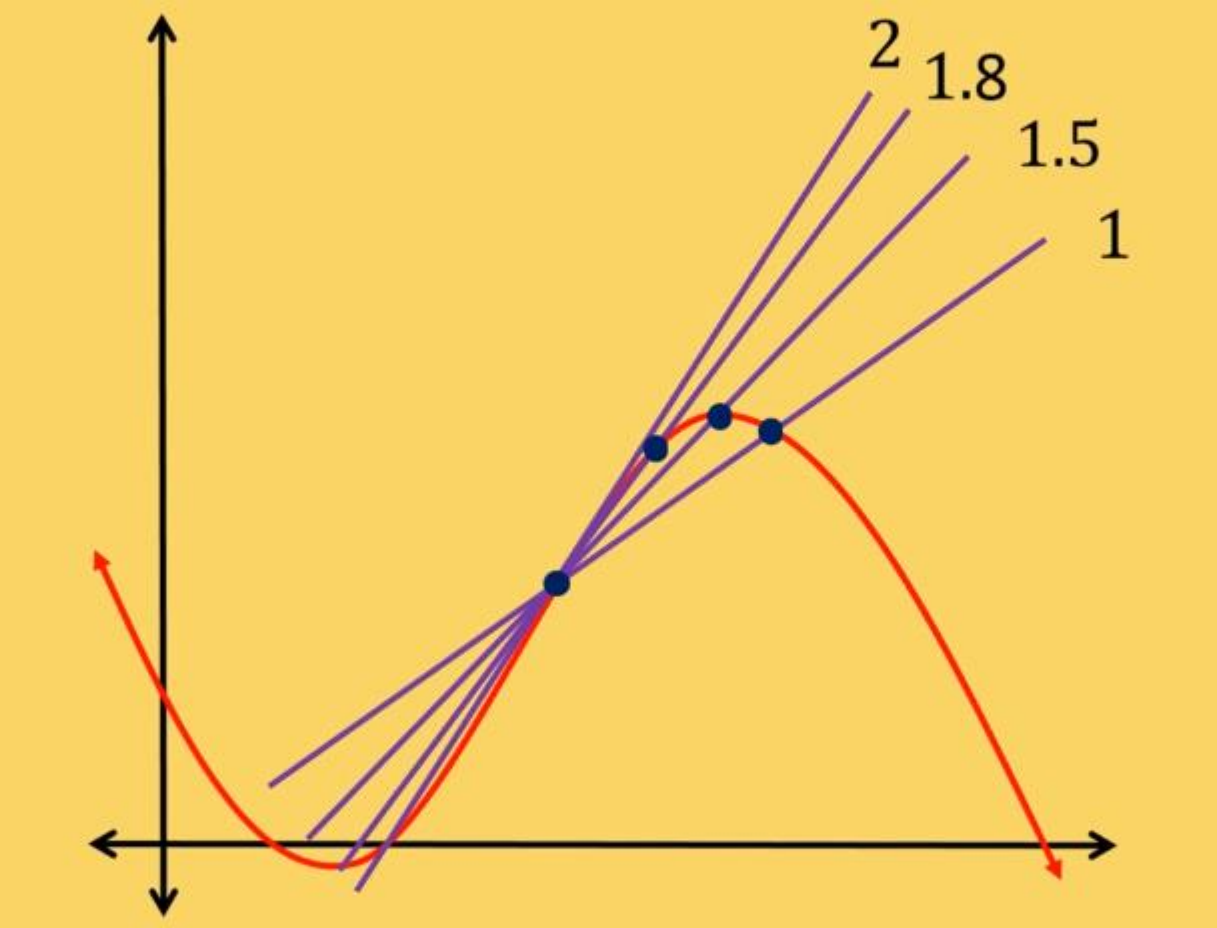
INSTANTANEOUS RATE OF CHANGE

If we move second point **further more closer**, we shall get bit different average change



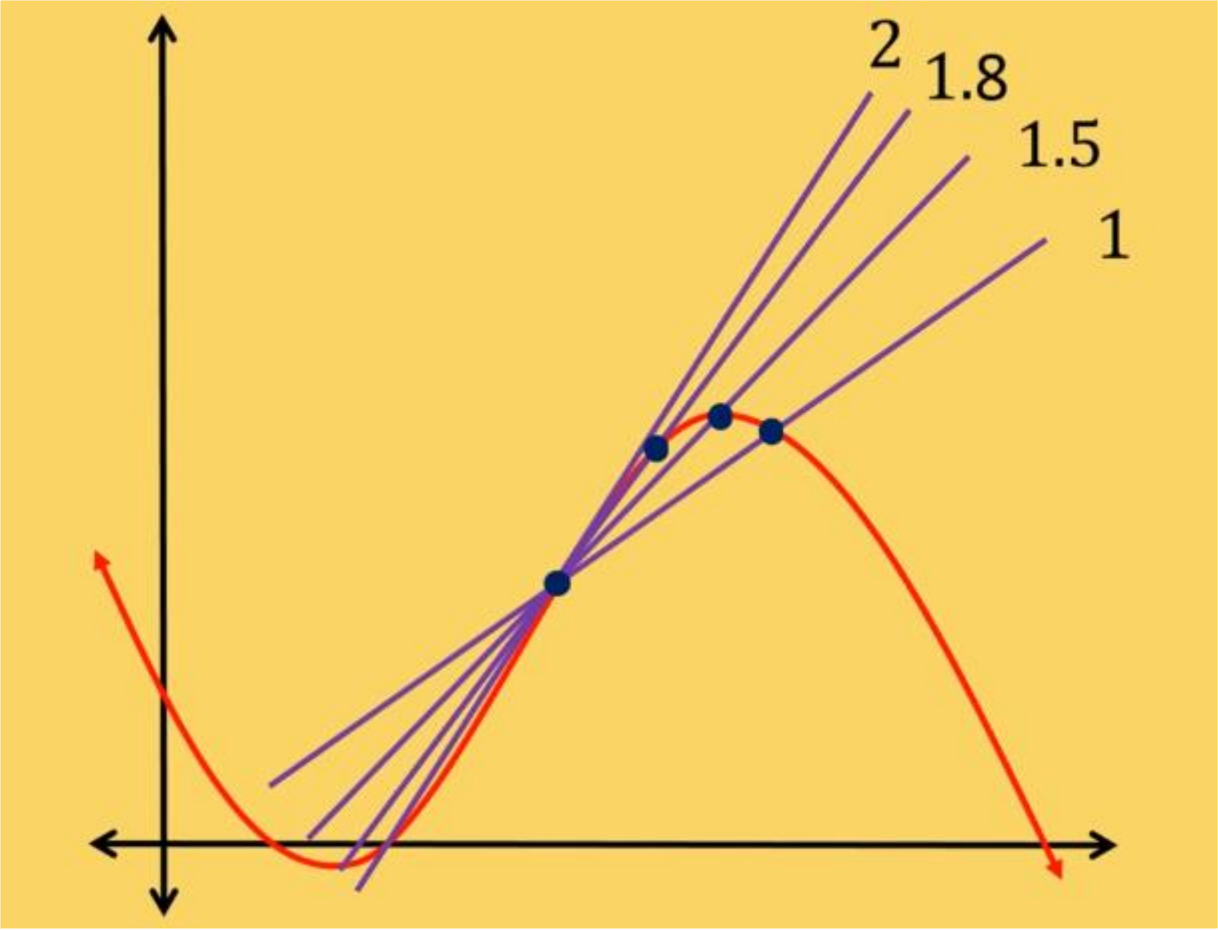
INSTANTANEOUS RATE OF CHANGE

As we are moving second point closer and closer the average values is reaching toward the 2



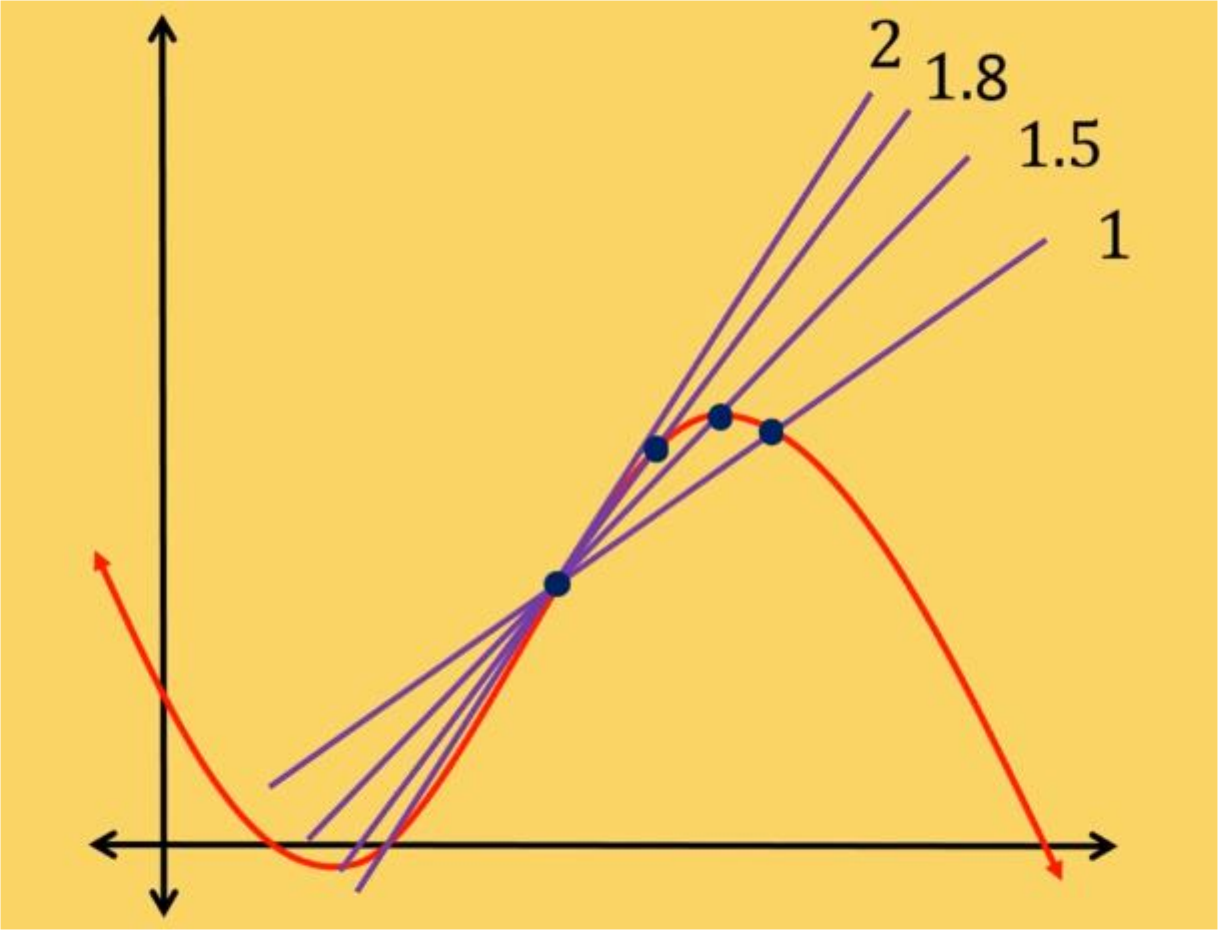
INSTANTANEOUS RATE OF CHANGE

So we can say to find change at single point we have to gradually move toward that point



INSTANTANEOUS RATE OF CHANGE

So we can say to find change at single point we have to gradually move toward that point



INSTANTANEOUS RATE OF CHANGE

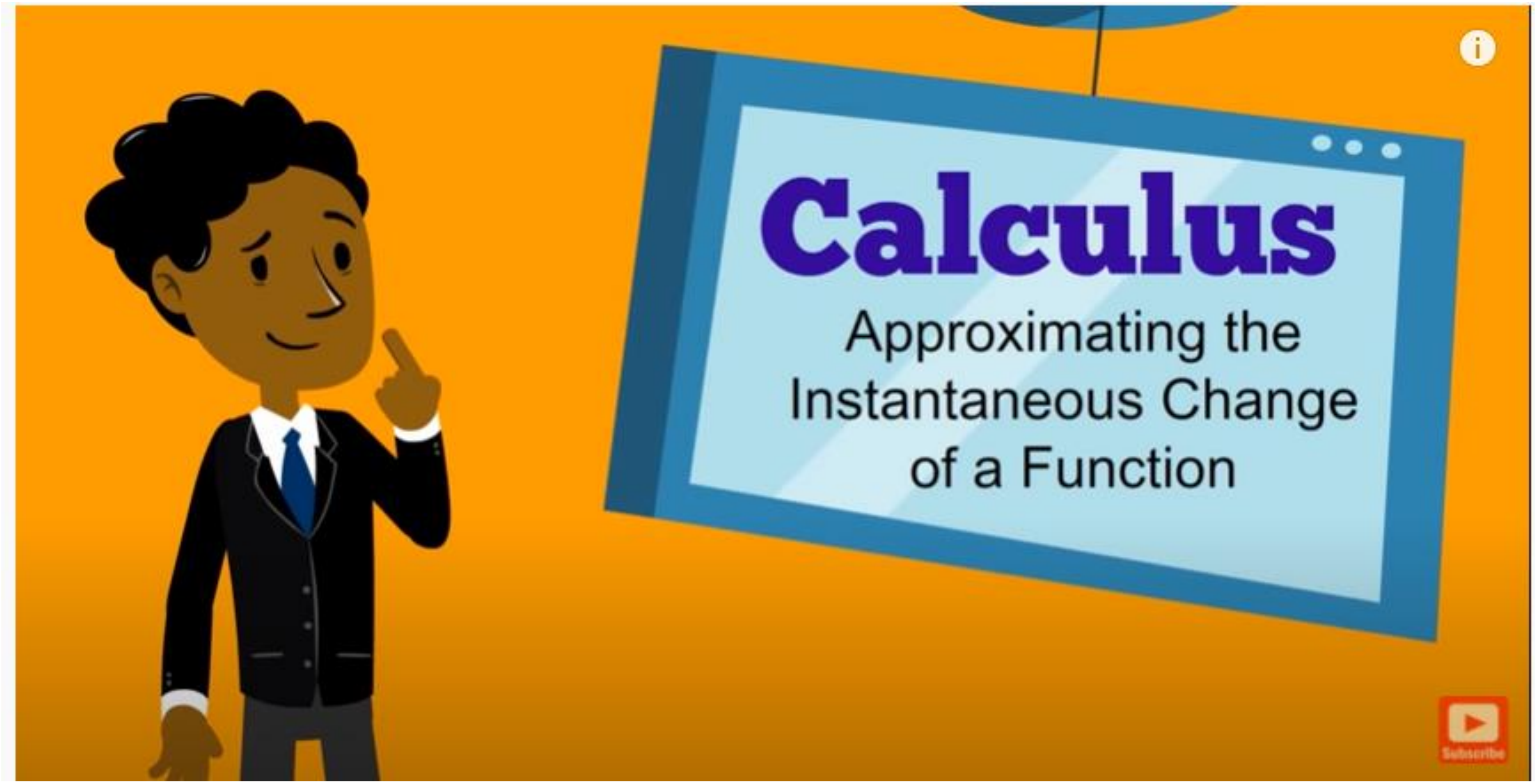
- Let we function $y = f(x) = x^2$, what is rate of change in y at $x=2$?

RATE OF CHANGE

- Let a function $y = f(x) = x^2$, what is rate of change in y at $x=2$?

REFERENCES

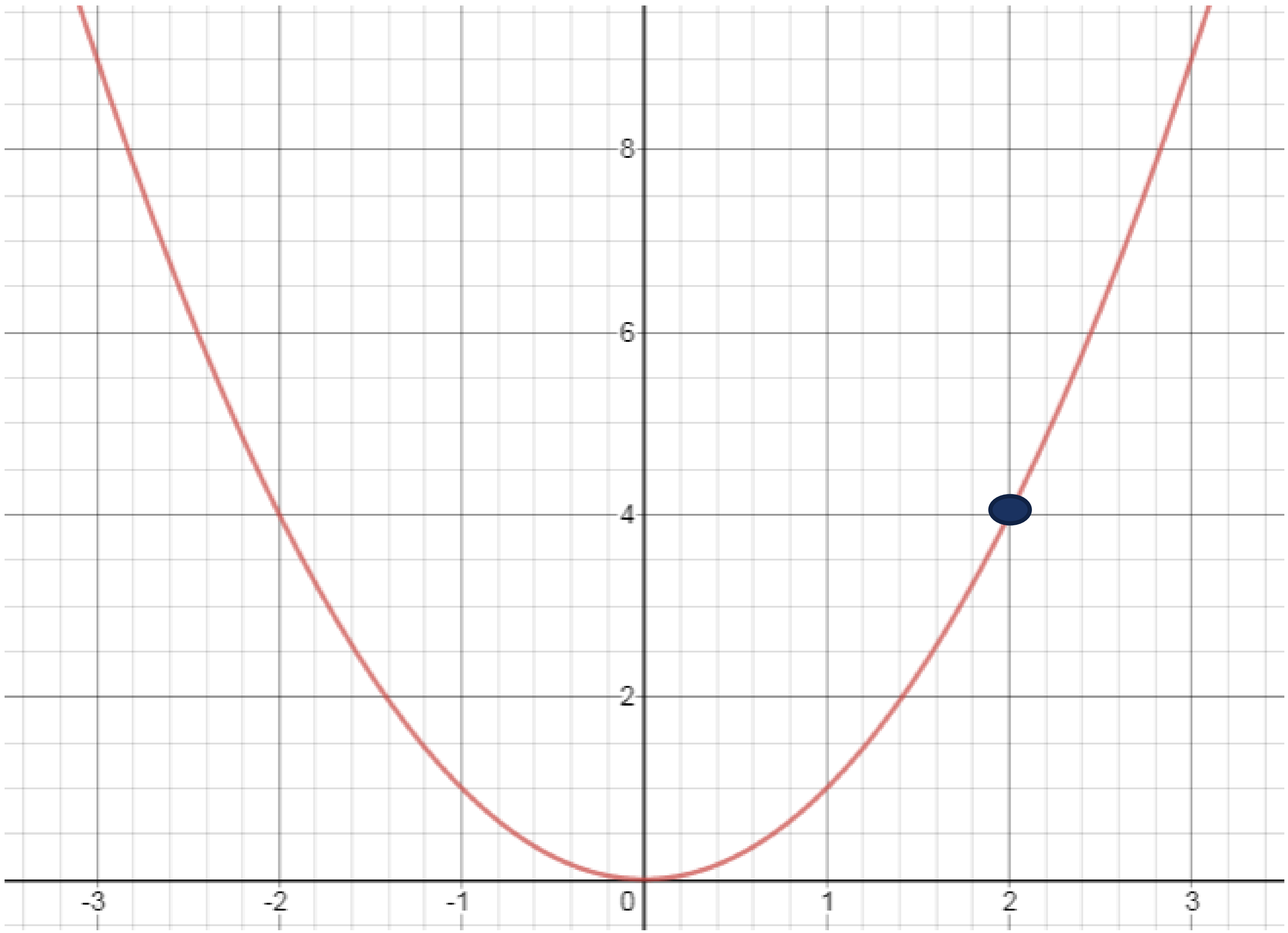
Some Content taken from following source: https://youtu.be/jTvTthjtIrk



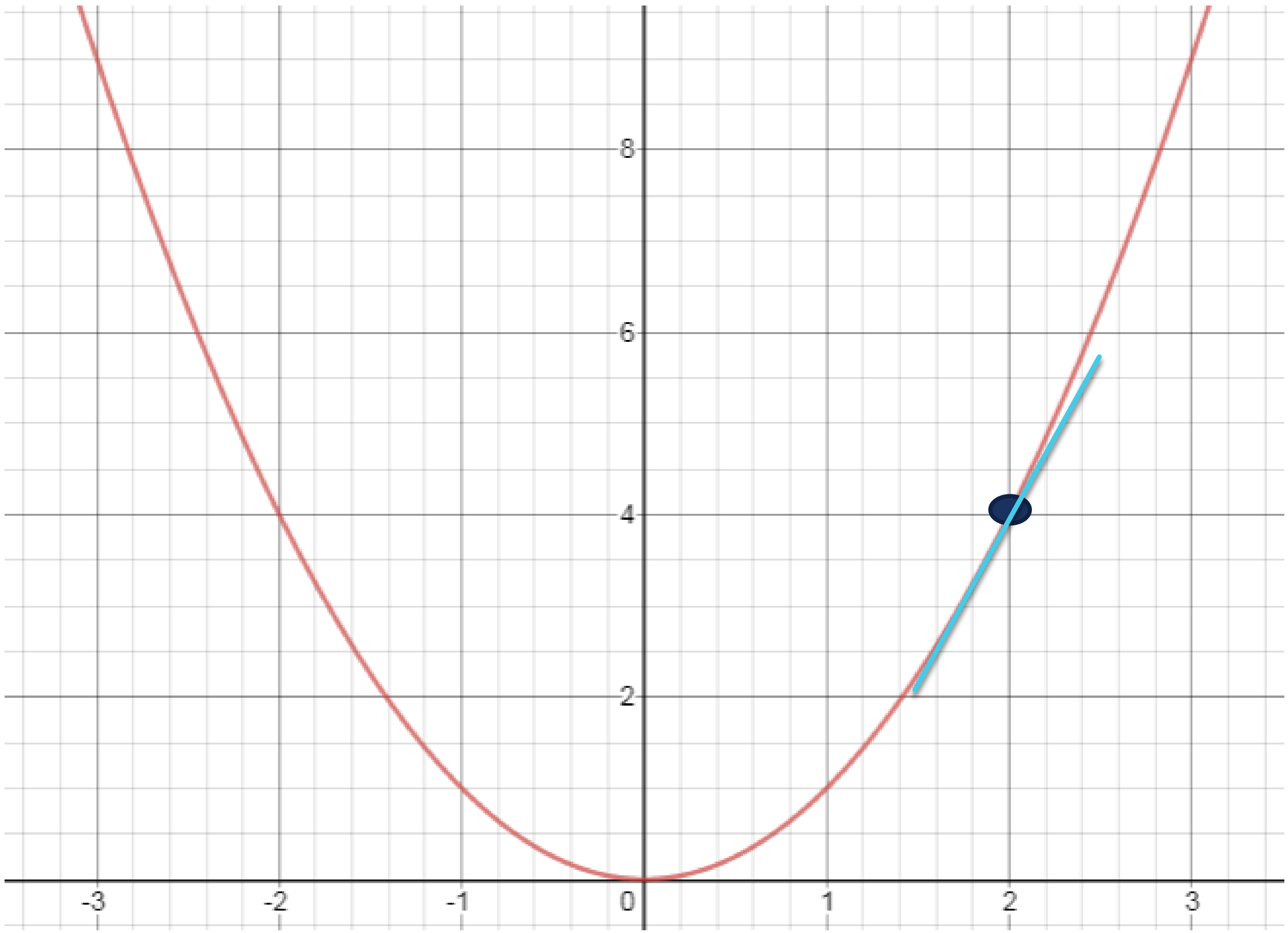
A SIMPLE METHOD

INSTANTANEOUS RATE OF CHANGE

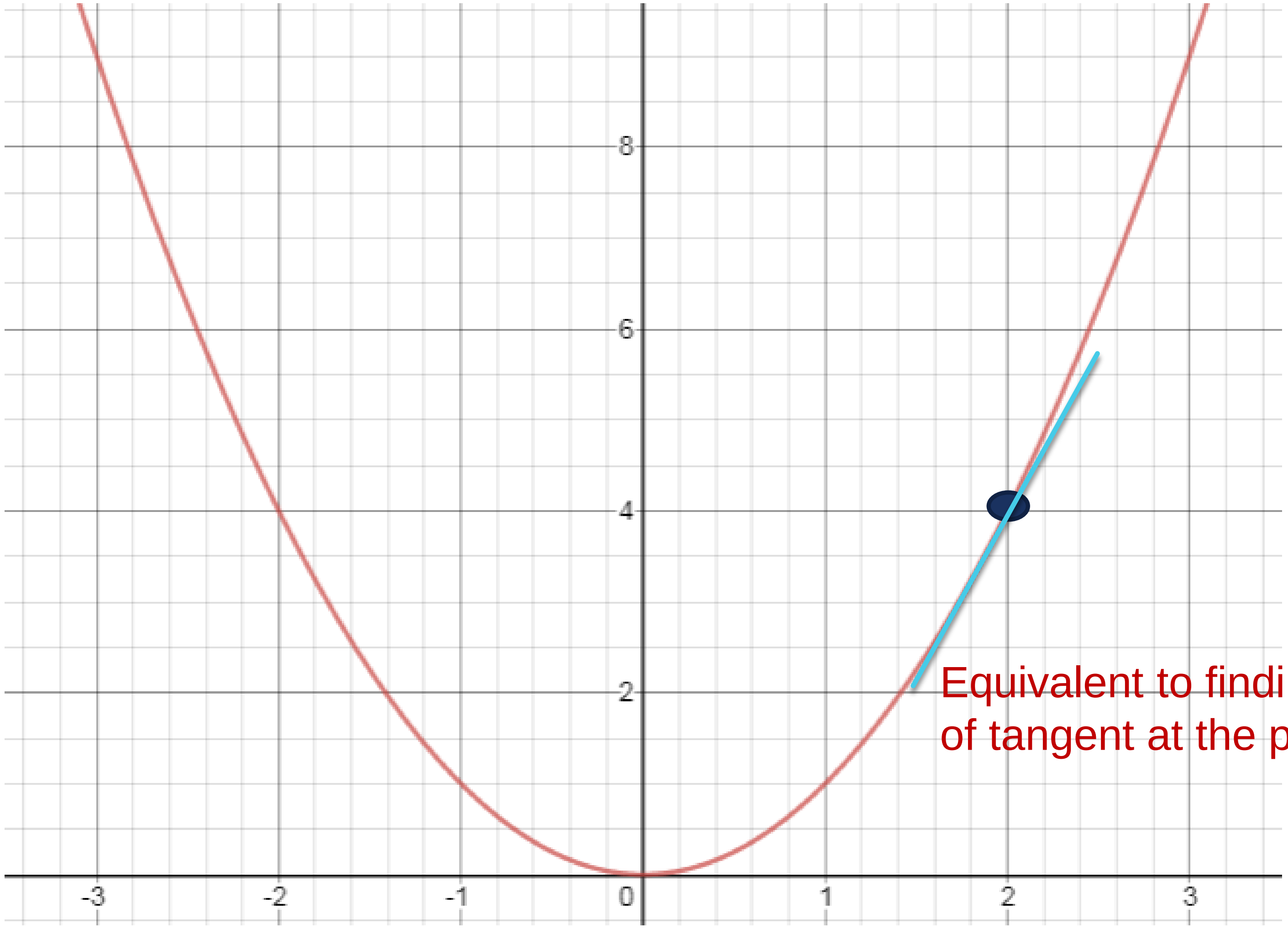
RATE OF CHANGE IN Y AT X=2



RATE OF CHANGE IN Y AT X=2

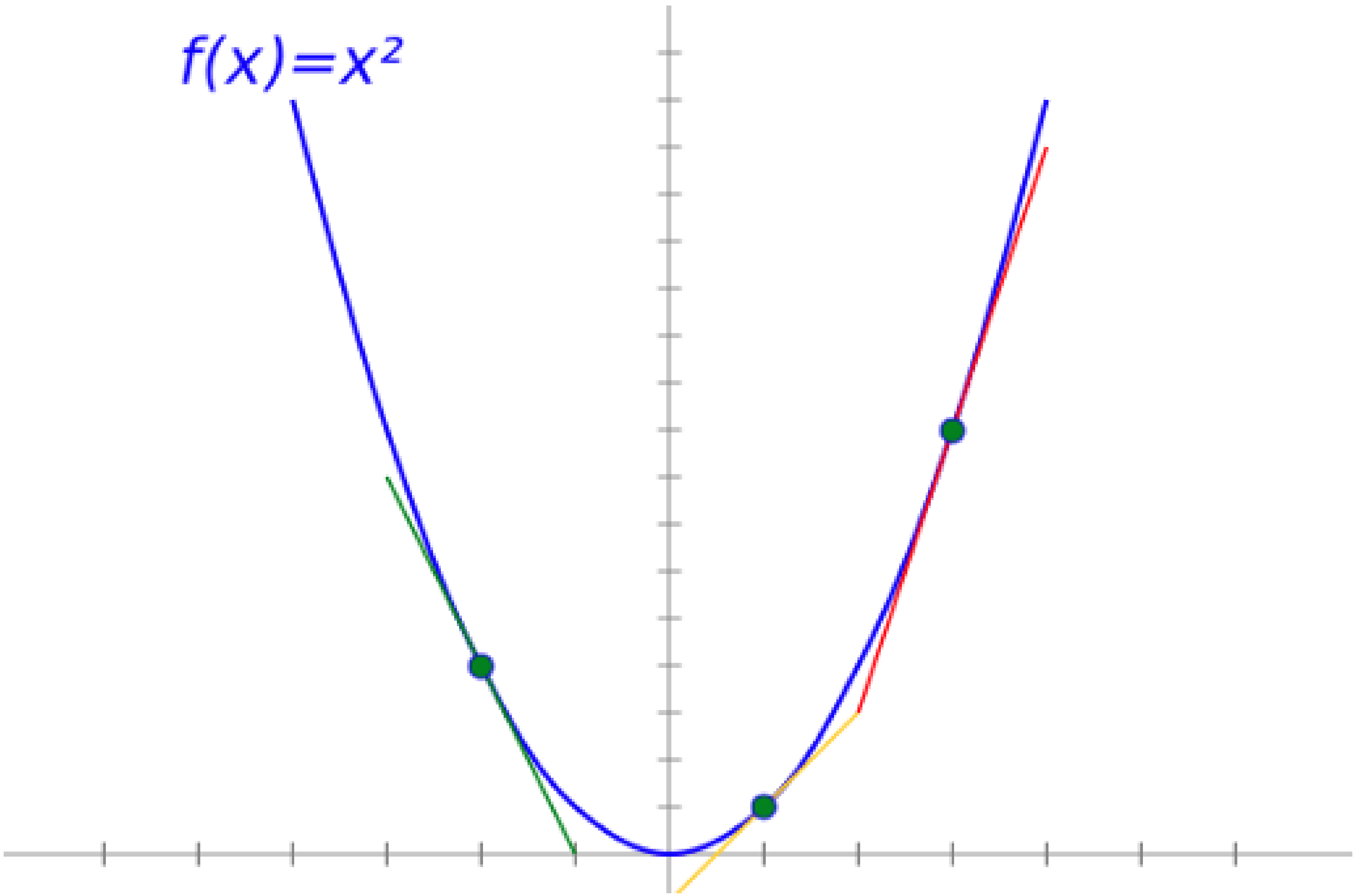


RATE OF CHANGE IN Y AT X=2



Equivalent to finding the slope of tangent at the point

FIND THE RATE OF CHANGE IN Y AT DIFFERENT POINTS



THAT IS ALSO CALLED DERIVATIVE

The slope or instantaneous rate of change at any point is also called as derivative of function at that point.

THAT IS ALSO CALLED DERIVATIVE

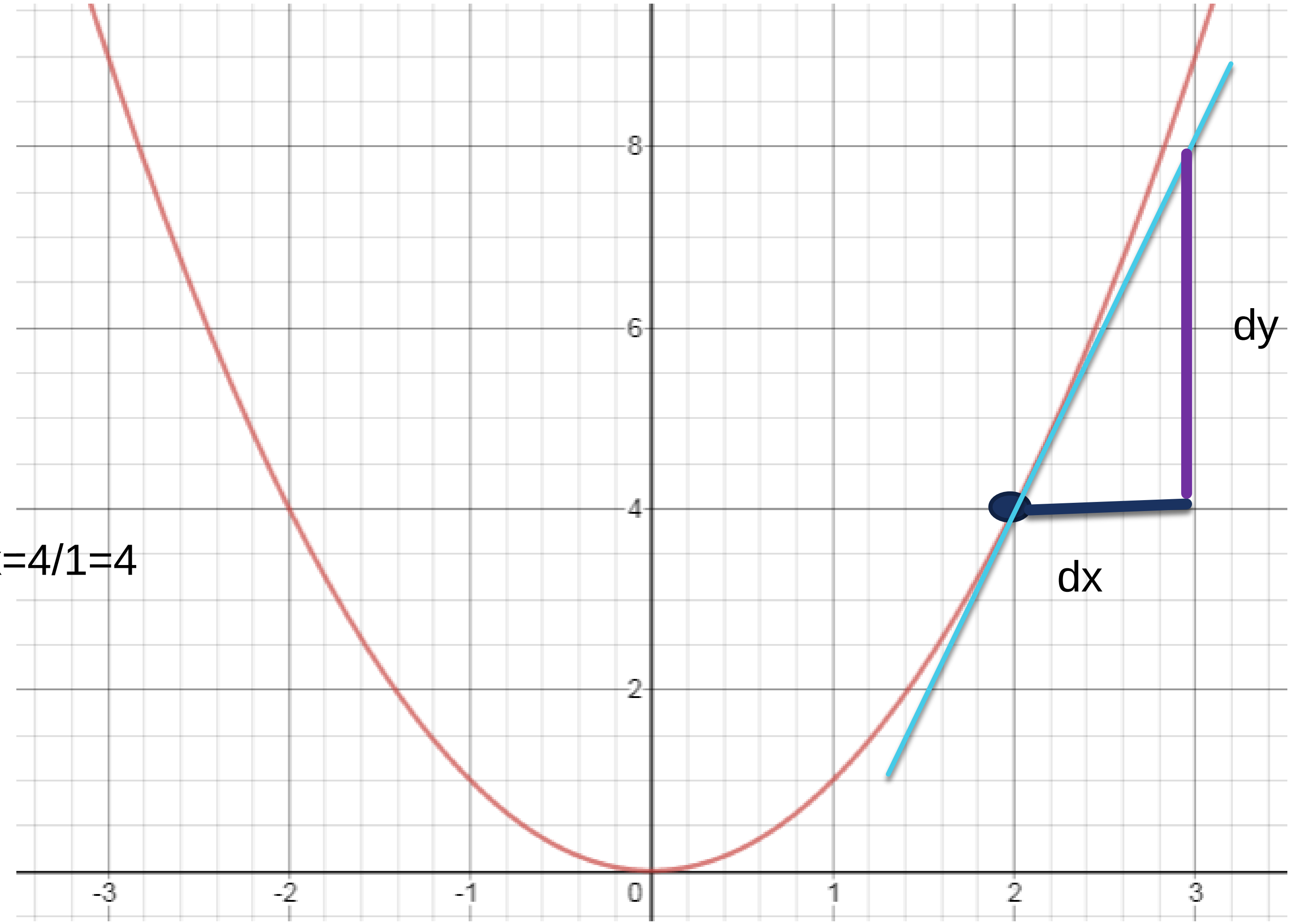
The derivative of f(x), is another function f'(x) that computes the slope of f(x) at a given point x.

THAT IS ALSO CALLED DERIVATIVE

In this situation, for $x = 2$, the slope of $f(x) = x^2$ is $2x$ or $2*2 = 4$.

RATE OF CHANGE IN Y AT X=2

dy=4
dx=1
dy/dx=4/1=4



DERIVATIVE

- In simpler terms, the derivative is instantaneous rate of change or the slope of a function at a given point.

PARTIAL DERIVATIVES

PARTIAL DERIVATIVE

For a multivariable function,

$f(x, y) = x^2y$, computing partial derivatives looks something like this:

$$\frac{\partial f}{\partial x} = \underbrace{\frac{\partial}{\partial x} x^2 y}_{\substack{\text{Treat } y \text{ as constant;} \\ \text{take derivative.}}} = 2xy$$

$$\frac{\partial f}{\partial y} = \underbrace{\frac{\partial}{\partial y} x^2 y}_{\substack{\text{Treat } x \text{ as constant;} \\ \text{take derivative.}}} = x^2 \cdot 1$$

WHY PARTIAL DERIVATIVES.

When the input of a function is made up of multiple variables, we want to see how the function changes as we *change one of those variables* while holding all the others constant

PARTIAL DERIVATIVES.

Can be thought of as "a tiny change in the function's output"

Used instead of "d" in usual $\frac{df}{dx}$ notation to emphasize that this is a partial derivative.

Multivariable function

$$\frac{\partial f}{\partial x}$$

Indicates which input variable is changed slightly.

Can be thought of as "a tiny change in x"

GRADIENT

SLIDES FROM KHAN ACADEMY TO EXPLAIN GRADIENTS

GRADIENT

- The gradient stores all the partial derivative information of a multivariable function.

GRADIENT

- But it's more than a mere storage mechanism, it has several wonderful interpretations and many, many uses.

WHAT IS THE GRADIENT

- The gradient of a scalar-valued multivariable function $f(x, y, \dots)$, denoted ∇f, packages all its partial derivative information into a vector:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \vdots \end{bmatrix}$$

In particular, this means ∇f is a vector-valued function.

GRADIENT VECTOR

Scalar-valued multivariable function

$$\nabla f(\underbrace{x_0, y_0, \dots}_{\substack{\nabla f \text{ takes the same} \\ \text{type of inputs as } f}}) = \underbrace{\begin{bmatrix} \frac{\partial f}{\partial x}(x_0, y_0, \dots) \\ \\ \frac{\partial f}{\partial y}(x_0, y_0, \dots) \\ \\ \vdots \end{bmatrix}}_{\substack{\nabla f \text{ outputs a vector with} \\ \text{all possible partial derivatives of } f.}}$$

Notation for gradient, called “nabla”.

EXAMPLE

If $f(x, y) = x^2 - xy$, which of the following represents ∇f?

Choose 1 answer:

(A) $\begin{bmatrix} 2x - x \\ x^2 - y \end{bmatrix}$

(B) $\begin{bmatrix} 2x - y \\ -x \end{bmatrix}$

EXAMPLE

If $f(x, y) = x^2 - xy$, which of the following represents ∇f?

Choose 1 answer:

(A) $\begin{bmatrix} 2x - x \\ x^2 - y \end{bmatrix}$

CORRECT (SELECTED)

$\begin{bmatrix} 2x - y \\ -x \end{bmatrix}$

INTERPRETATION

- If you imagine standing at a point $f(x_0, y_0,\ldots)$ in the input space of f, the vector $\nabla f(x_0,y_0,\ldots)$ tells you which direction you should travel to increase the value of f most rapidly.

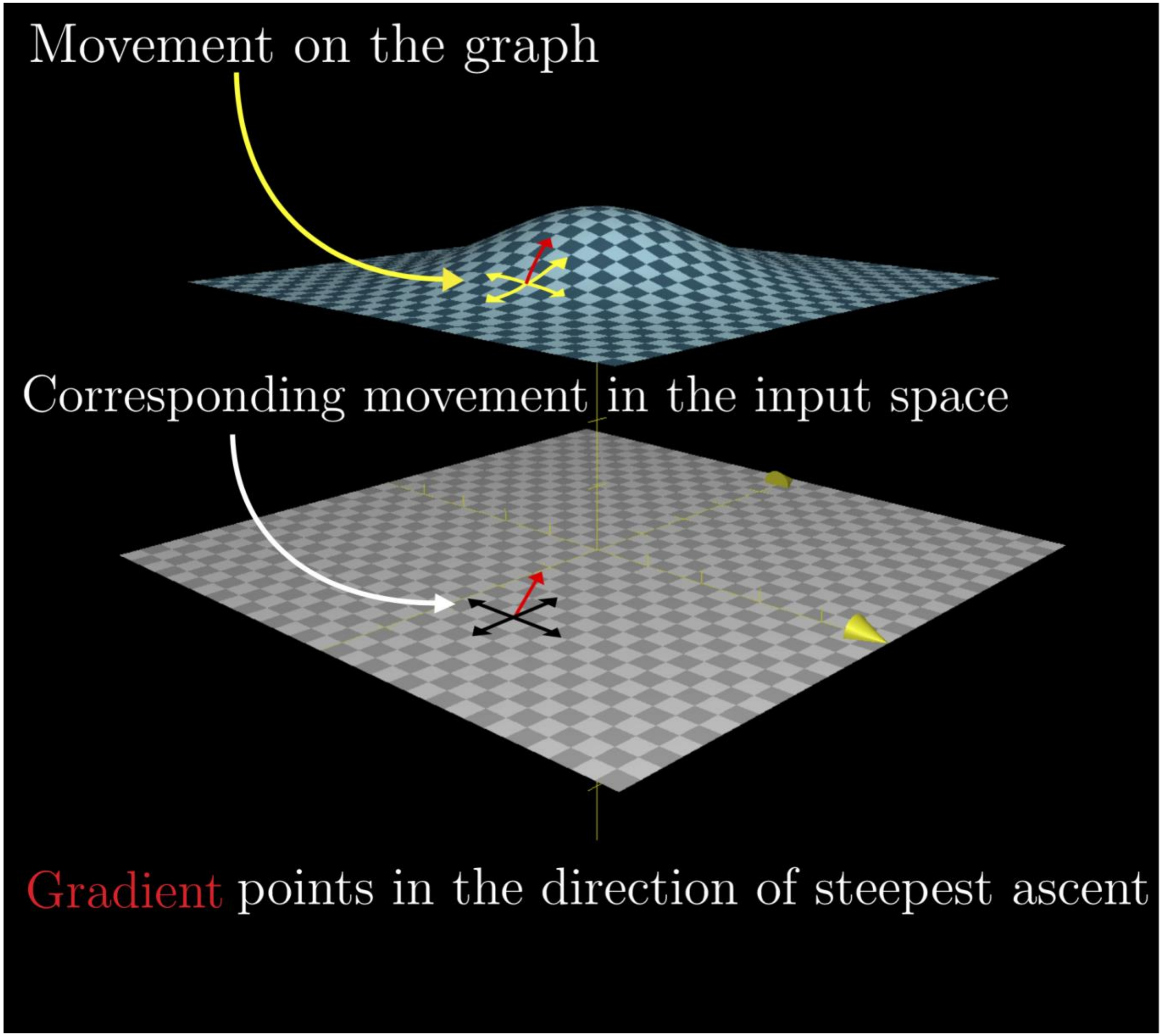
INTERPRETATION

- Think of the graph of f as a hilly terrain. If you are standing on the part of the graph directly above—or below—the point (x_0, y_0)—, the slope of the hill depends on which direction you walk.

INTERPRETATION

- For example, if you step straight in the positive x direction, the slope is $\frac{\partial f}{\partial x}$
- if you step straight in the positive y-direction, the slope is $\frac{\partial f}{\partial y}$
- But most directions are some combination of the two

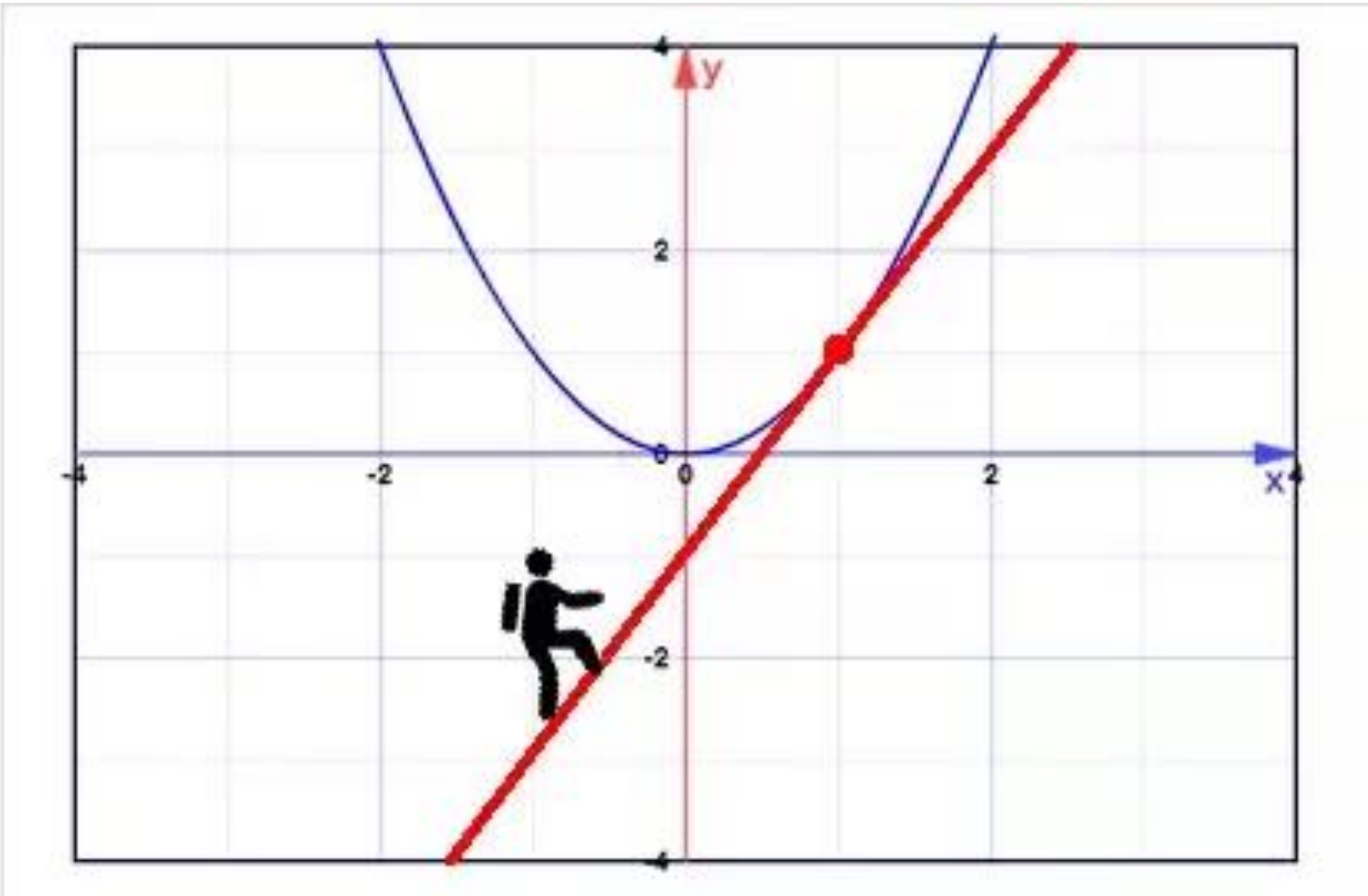
INTERPRETATION



GRADIENT IS STEEPEST ASCENT

- **The most important thing to remember about the gradient**: The gradient of f at a points (x_0, y_0) is in the direction of steepest ascent.

INTERPRETATION



PROGRAMMING ASSIGNMENT 01

- Write a python program that calculate the gradient vector of a function at any given point. Multiply it with 2, and calculate new point in direction of the gradient vector)
- Draw the Function on Graph and Highlight the starting point and new calculated point

HINT FOR ASSIGNMENT 01

Step 01: Define a function

- First you need to define the function with two argument for that you need to find the gradient vector.
- This function take two arguments and return the value of $f(B_0,B_1)$ according to the passed parameters.
- Let the function is
- $f(B_0,B_1) = x^2 - y^2$

HINT FOR ASSIGNMENT 01

Step 02: functions to calculate

- Define two separate functions that calculates the partial derivative of the given functions with respect to B_0 and B_1
- Each of these functions will take one argument and return the value of the partial derivative at given point.

HINT FOR ASSIGNMENT 01

Step 03: function to find the Gradient

- Write function to find the Gradient vector at any given point. Take two values as input and return two value in form of pair (gradient vector)

HINT FOR ASSIGNMENT 01

Step 04: Visualization

- Draw the graph of the function with some values of B0 and B1 let within the interval [-10,10]
- Draw the graph the point at which we want to find the gradient vector
- Draw the gradient vector point on the graph (you can multiply gradient vector with 2)
- (Optional: for visualization you can use http://www.math3d.org/)

PROGRAMMING ASSIGNMENT 02

- Extend the previous code that take starting point and move toward direction of the ascending slope and find the point until it start decreasing or until 20 iterations (whatever come first)
- Draw the Function on Graph and Highlight the starting point and final calculated point

PROGRAMMING ASSIGNMENT 02

- Extend the previous code that take starting point and move toward direction of the ascending slope and find the point until it start decreasing or until 20 iterations (whatever come first)
- Hint: From starting point find the gradient vector, calculate new point on graph using the gradient vector and keep iterating until either of two conditions meet.