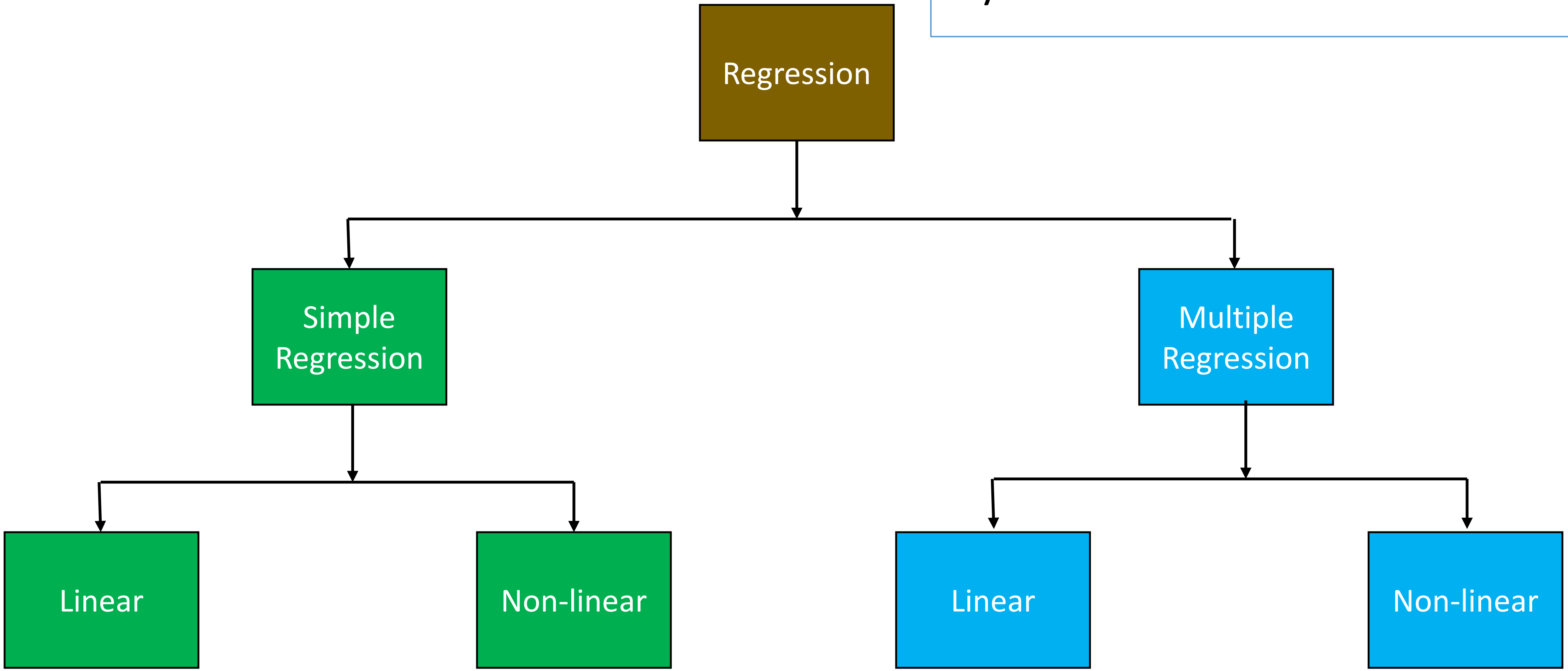
Linear Regression

Introduction

- Linear regression is a powerful and widely used method to estimate values, such as
  - the price of a house
  - the value of a certain stock
  - the life expectancy of an individual
  - the amount of time a user will watch a video or spend on a website

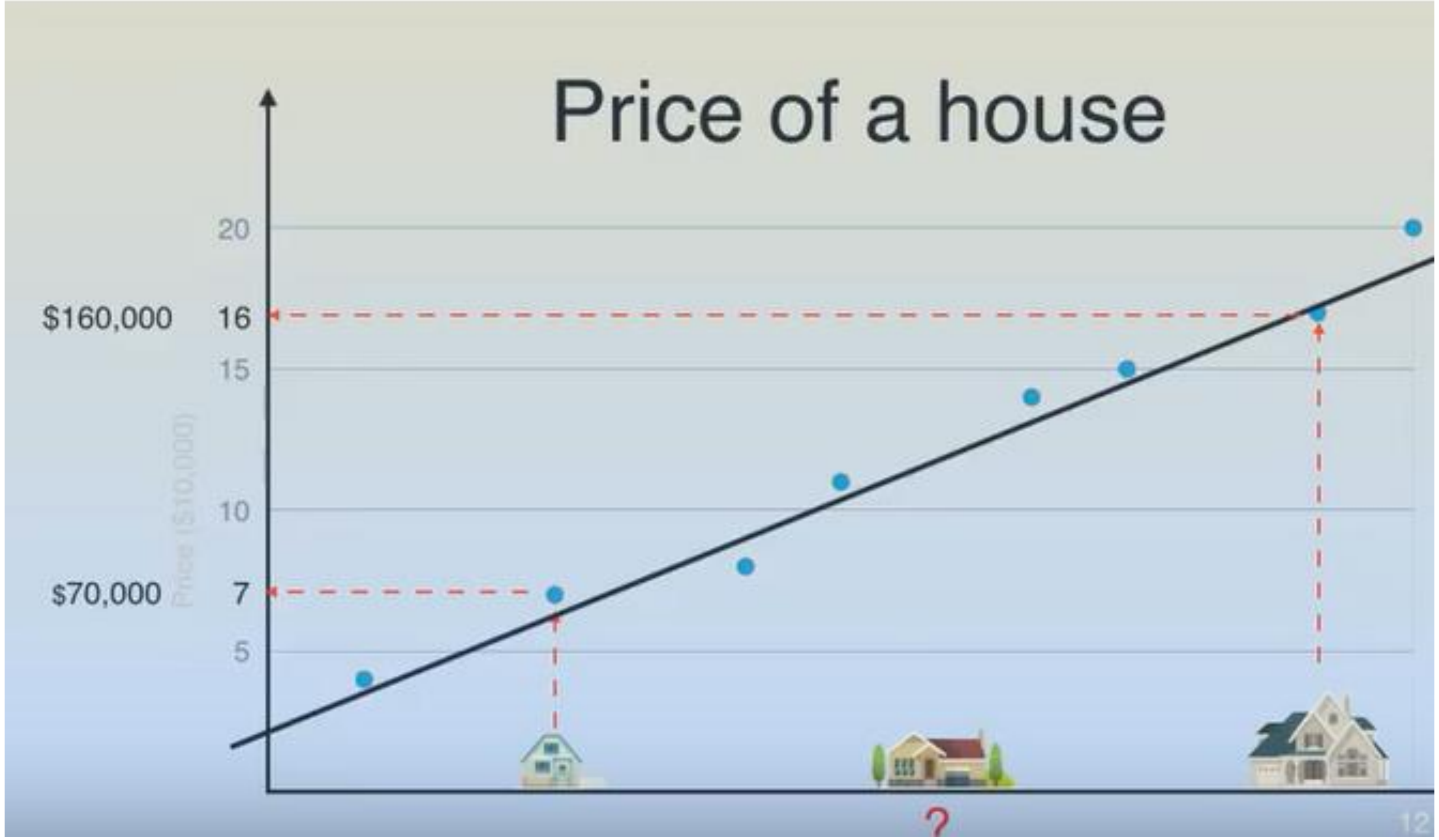
Regression

- It is a supervised learning approach
- If x is given, you can predict y = f(x)
- y is continuous value.



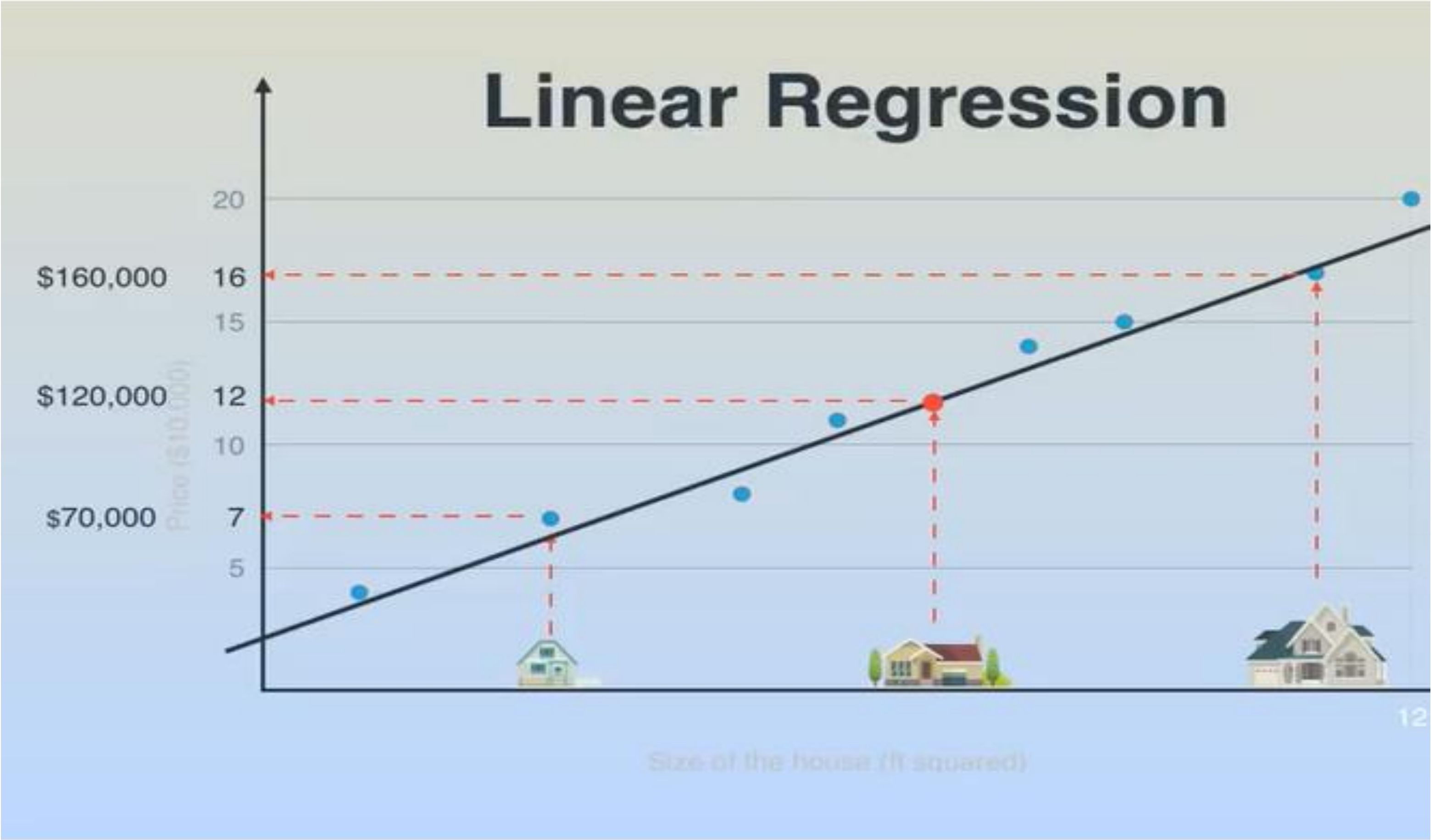
Linear Regression Intuition

House Price Estimation



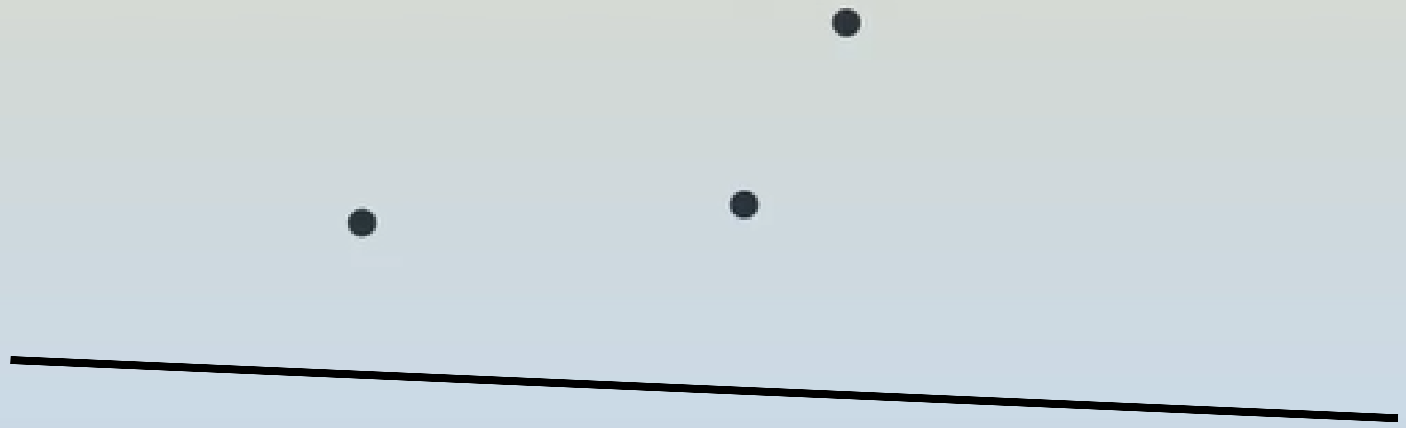
What is the best estimate for the price of the house?

What is the best estimate for the price of the house?

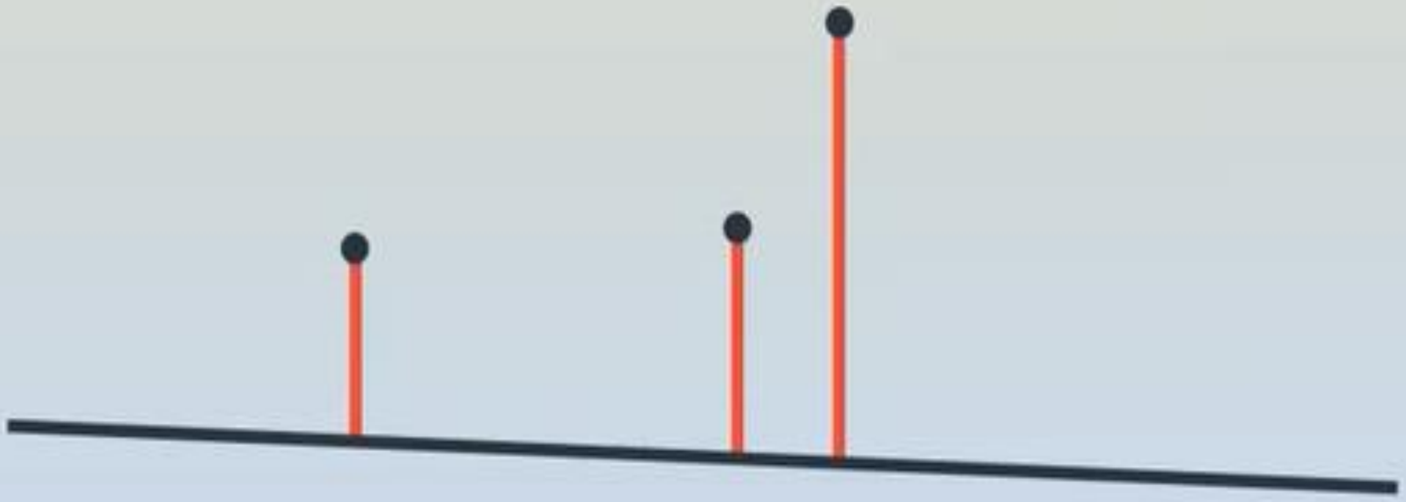


What is the best estimate for the price of the house?

Linear Regression



Linear Regression

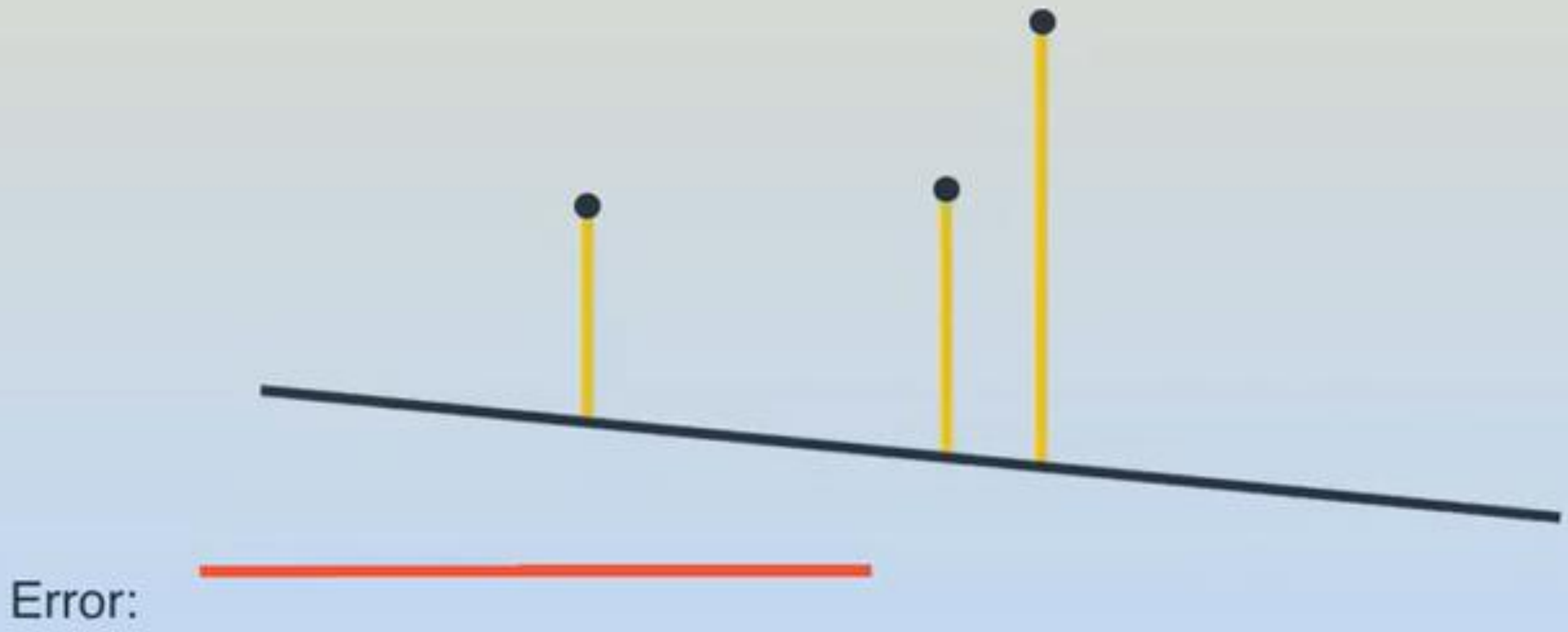


Error:

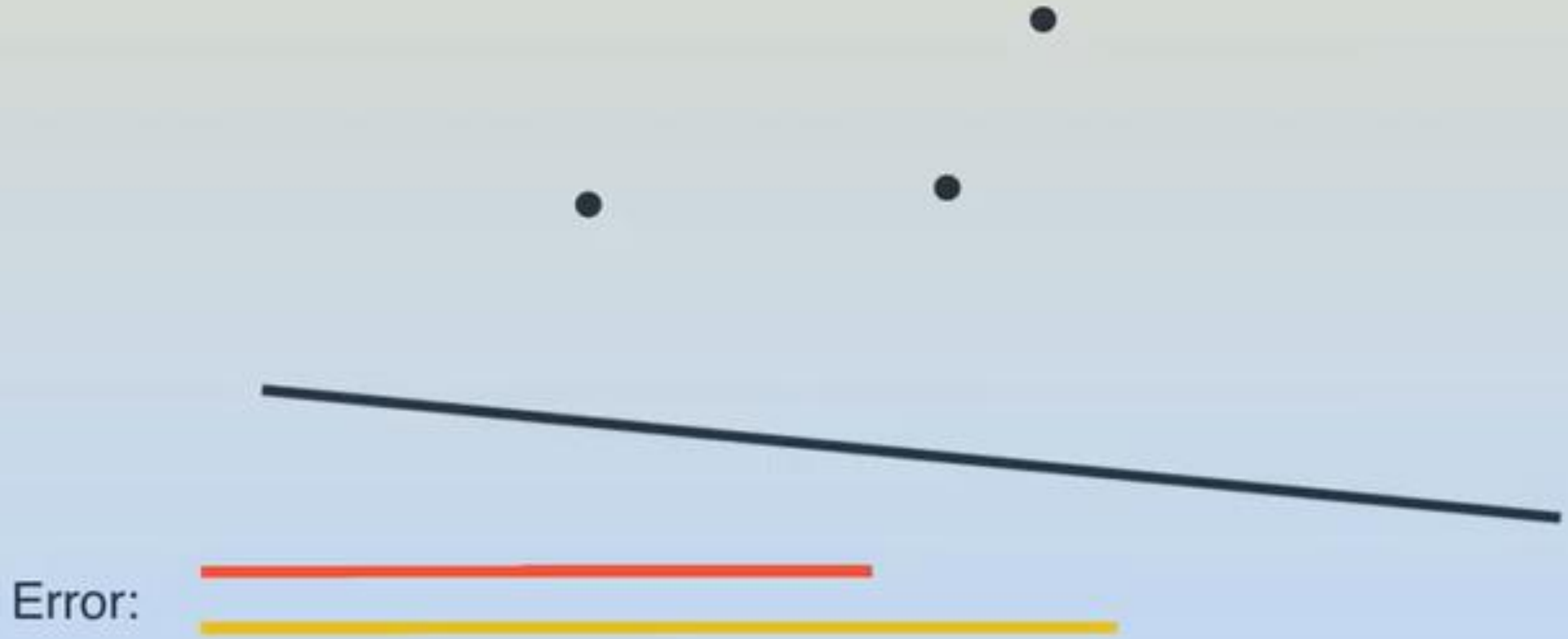
Linear Regression

Error:

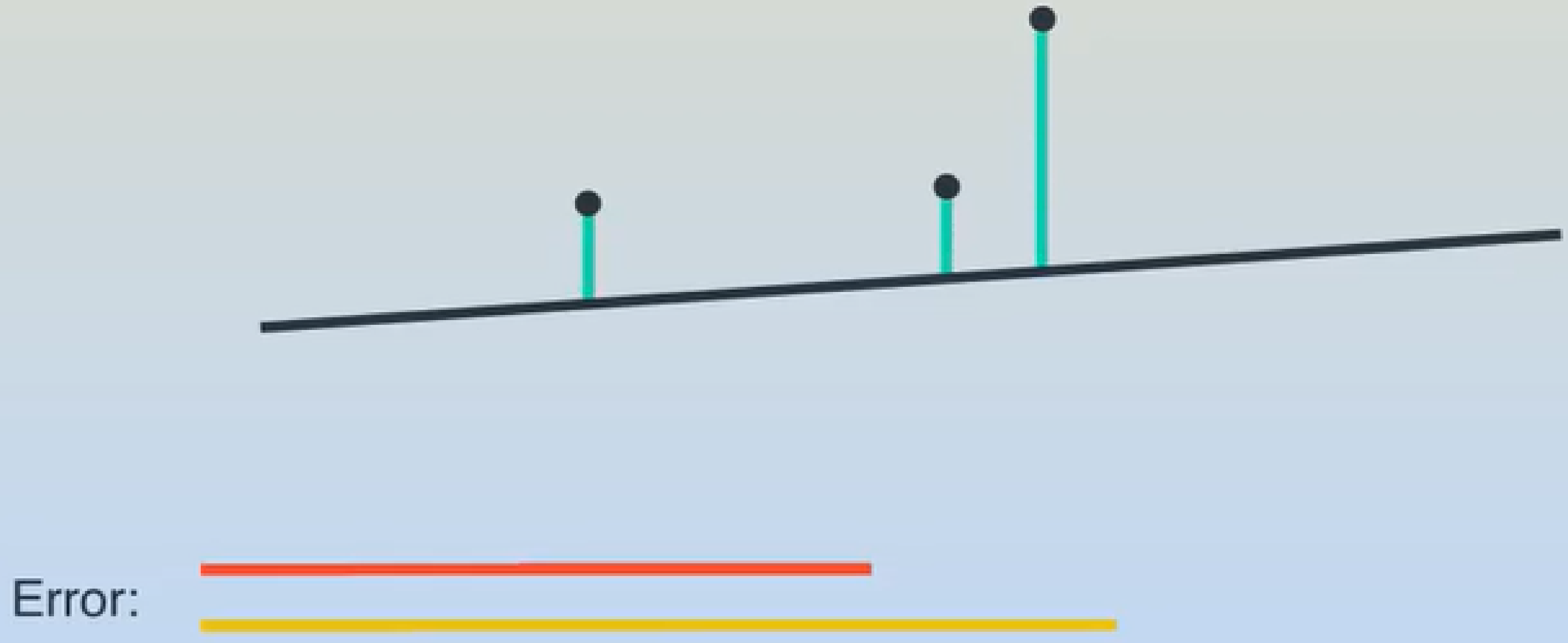
Linear Regression



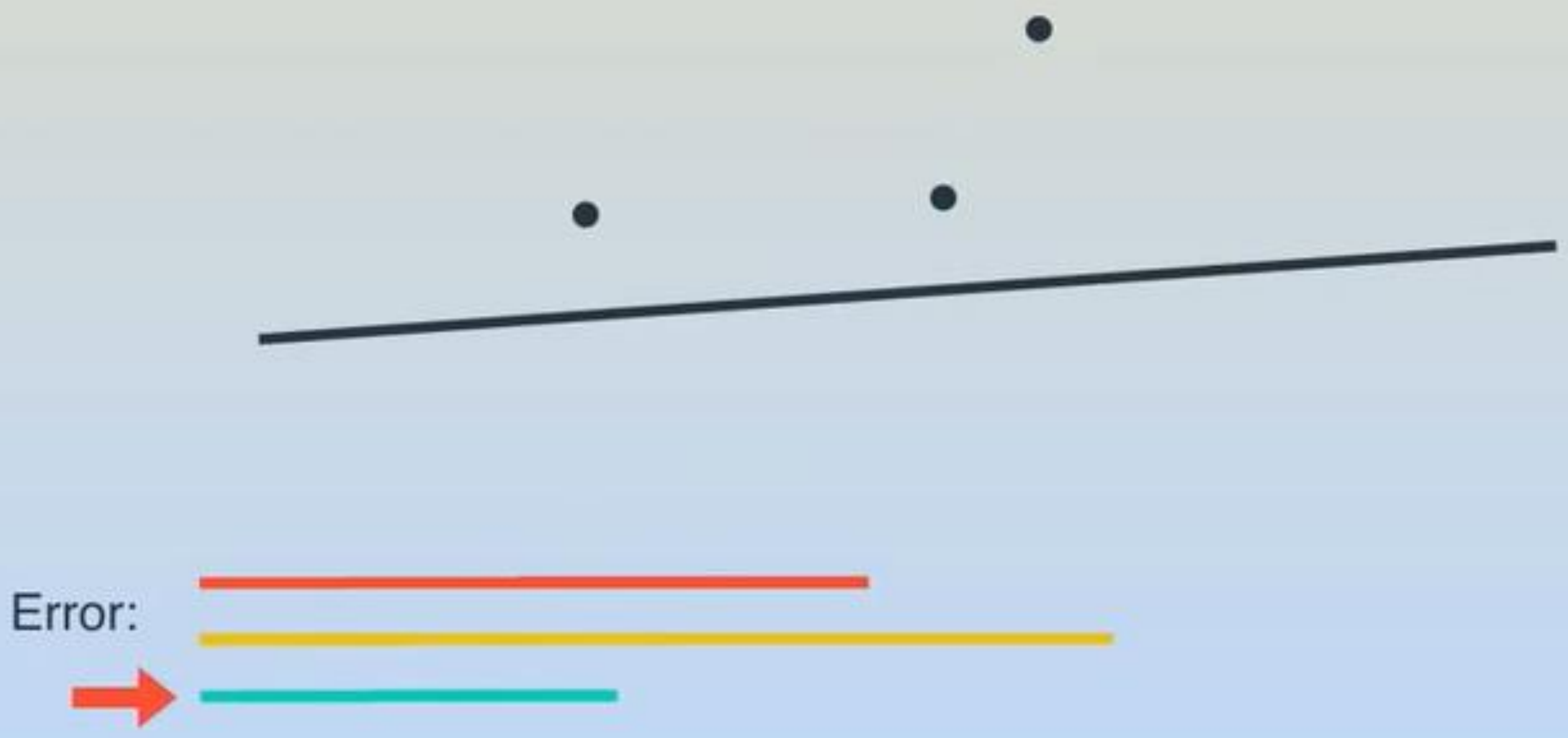
Linear Regression

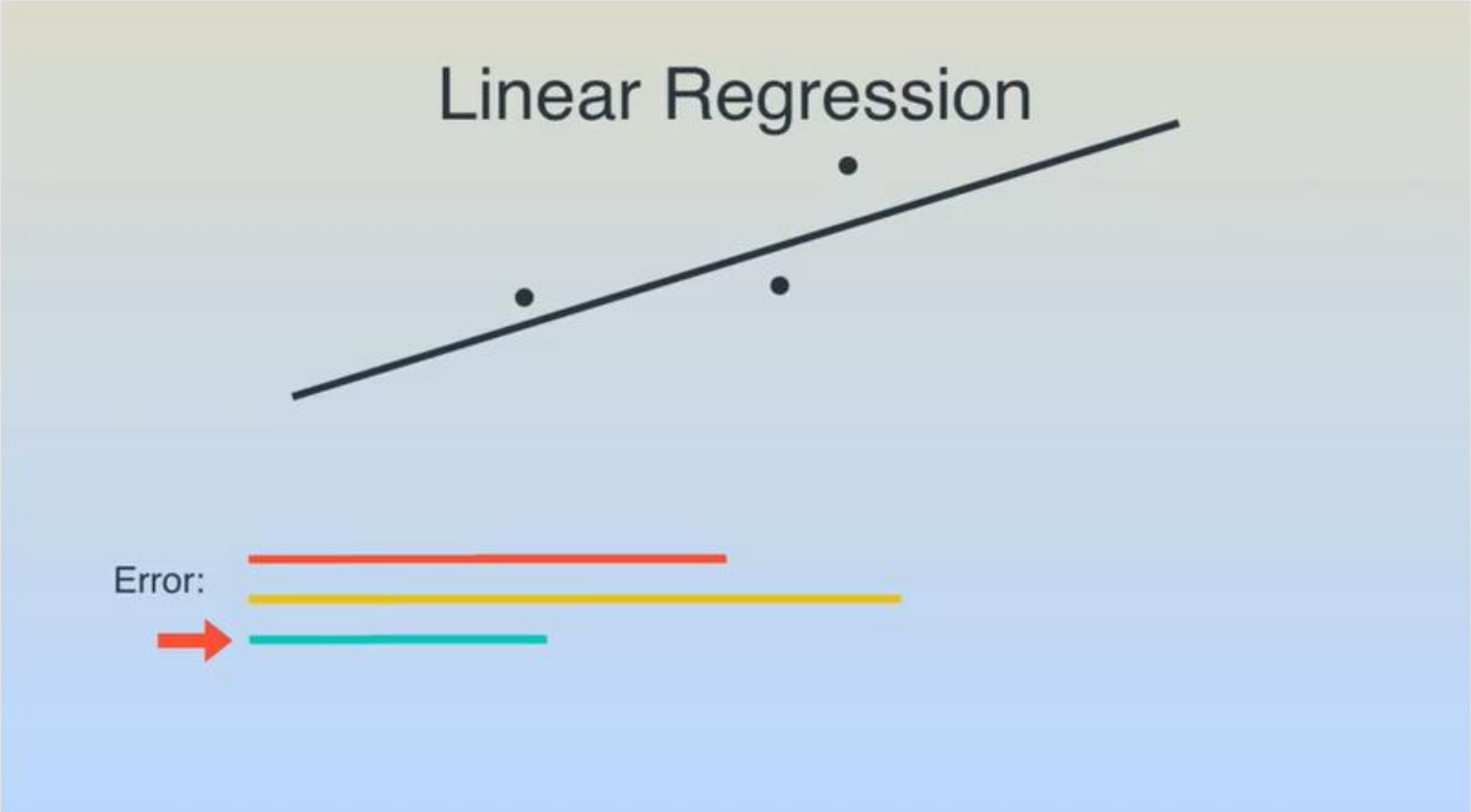


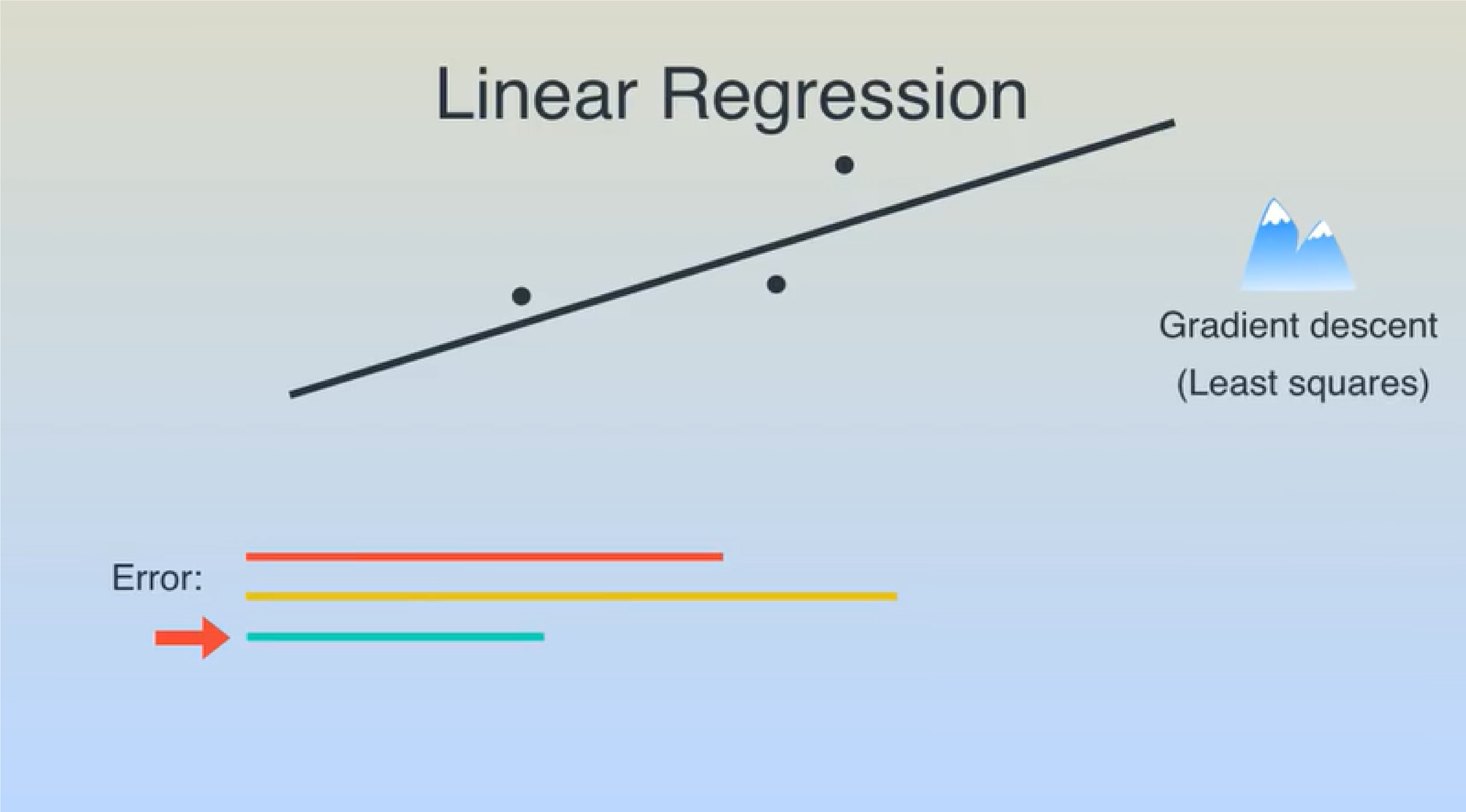
Linear Regression

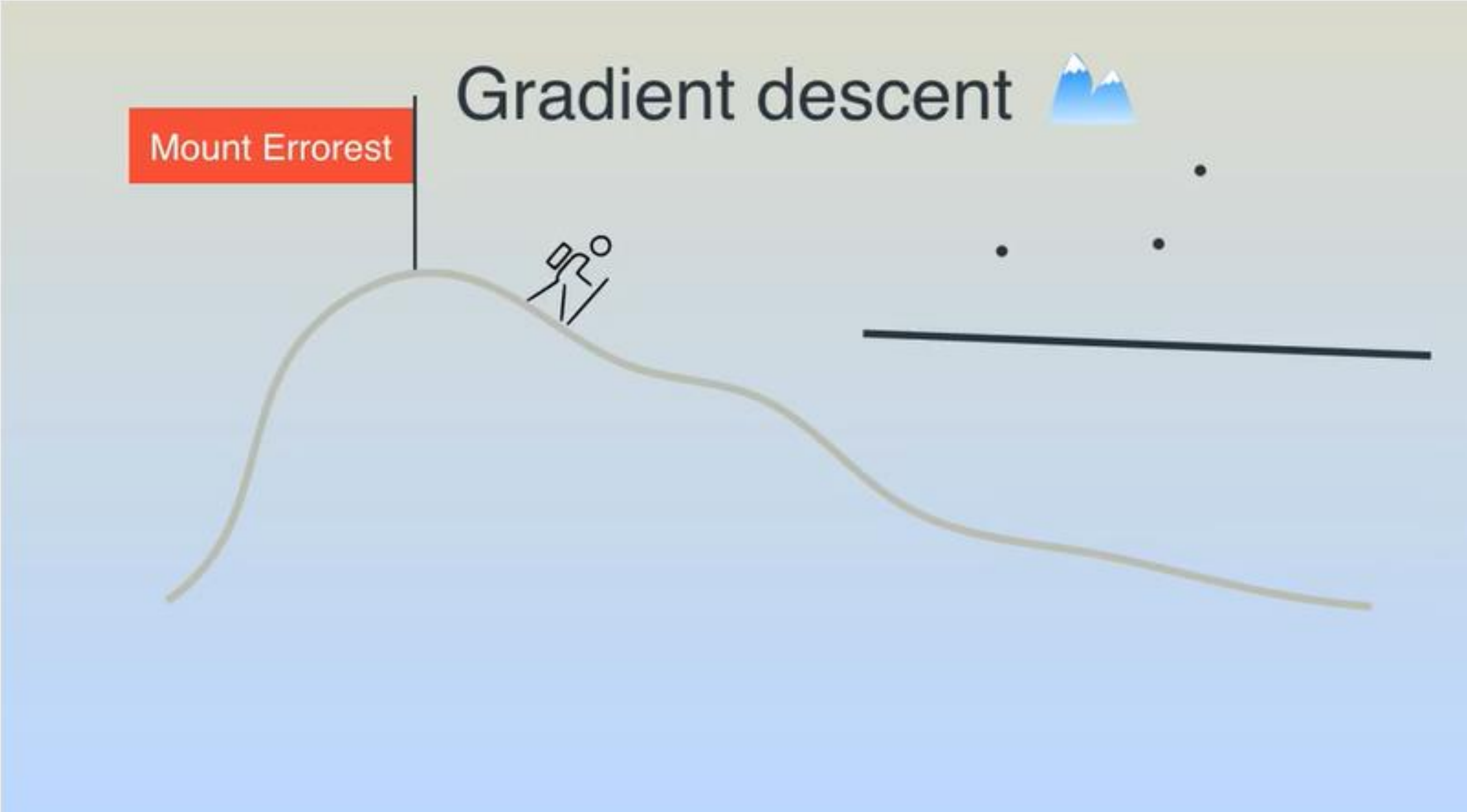


Linear Regression

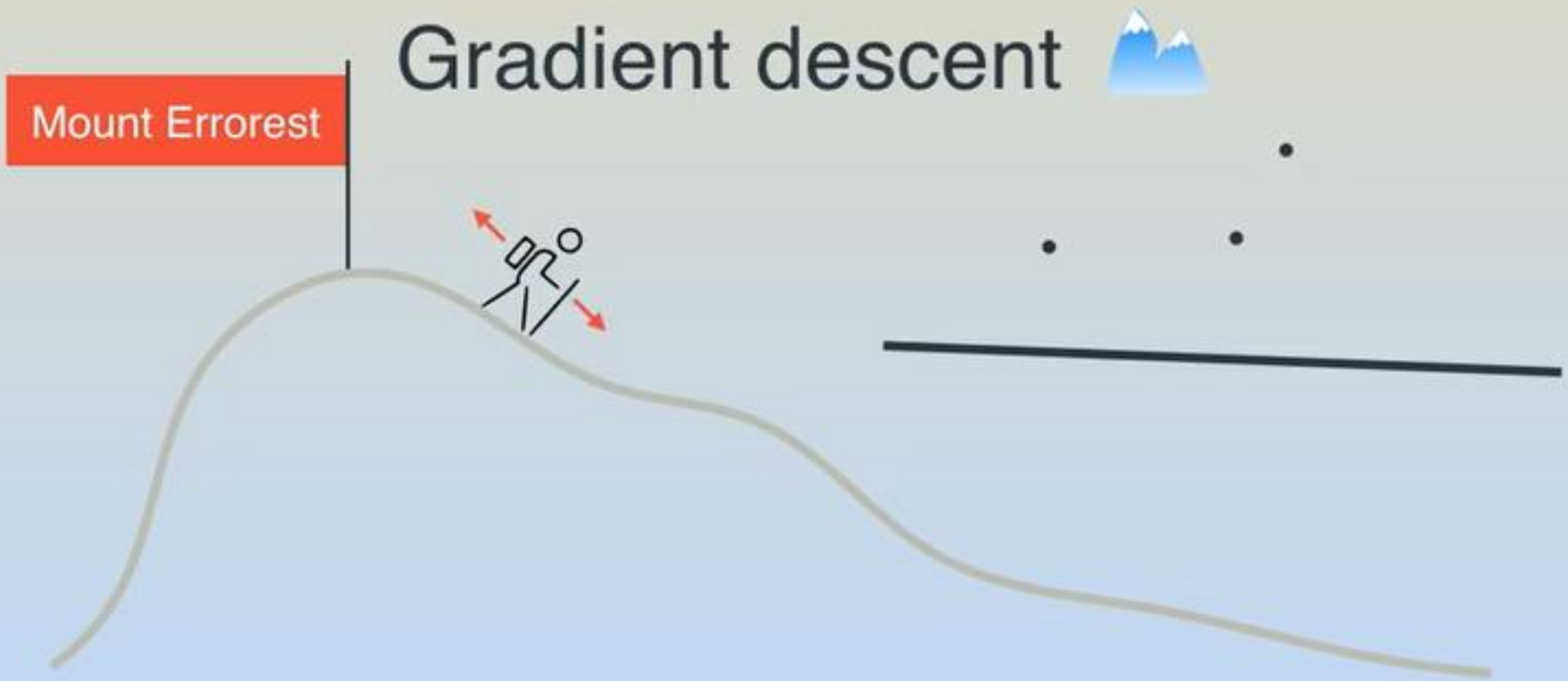



Linear Regression
Error:


Linear Regression
Gradient descent
(Least squares)
Error:


Gradient descent
Mount Errorest

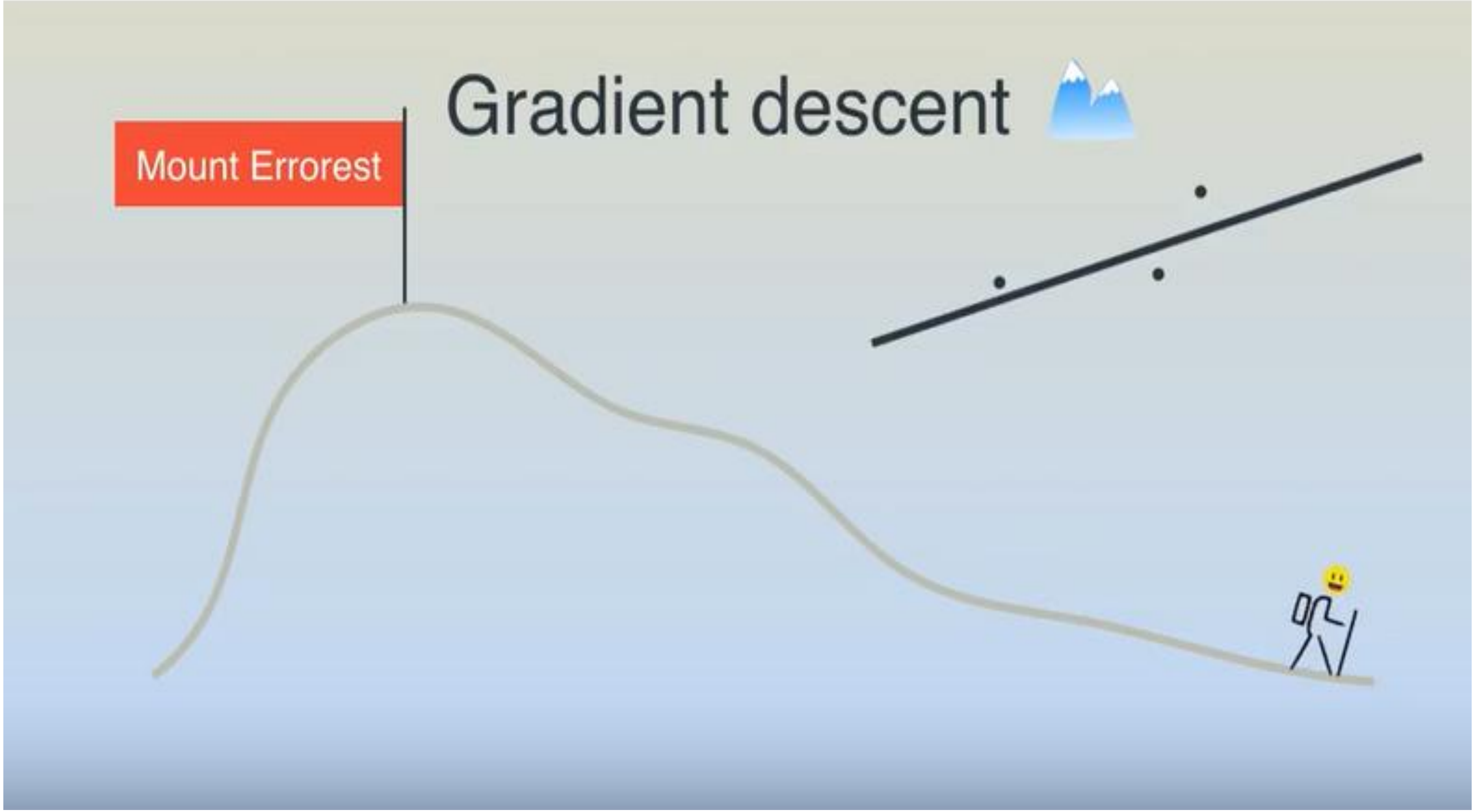
Gradient descent



Gradient descent 🏔️

Gradient descent 🏔

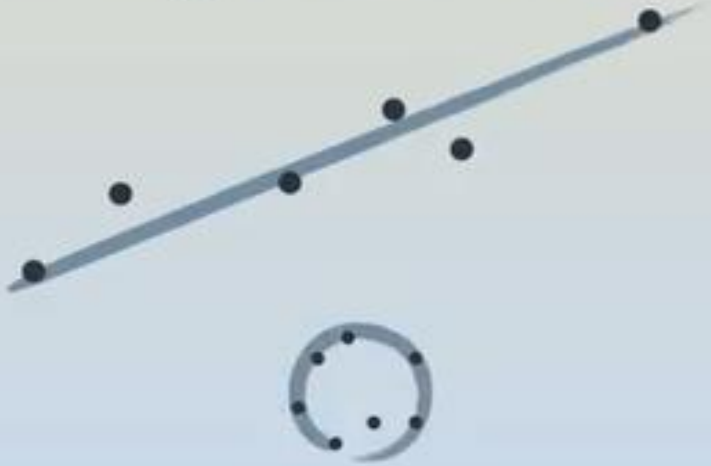
Gradient descent 🏔️


Gradient descent
Mount Errorest

Linear and Polynomial Regression

Linear and Polynomial Regression

Linear and Polynomial Regression



Linear and Polynomial Regression

Introduction

- **Regression models** describe the relationship between variables by fitting a line to the observed data.
- Linear regression models use a straight line, while logistic and nonlinear regression models use a curved line.
- Regression allows you to estimate how a dependent variable changes as the independent variable(s) change.

Simple Linear Regression

- **Simple linear regression** is used to estimate the relationship between **two quantitative variables**.

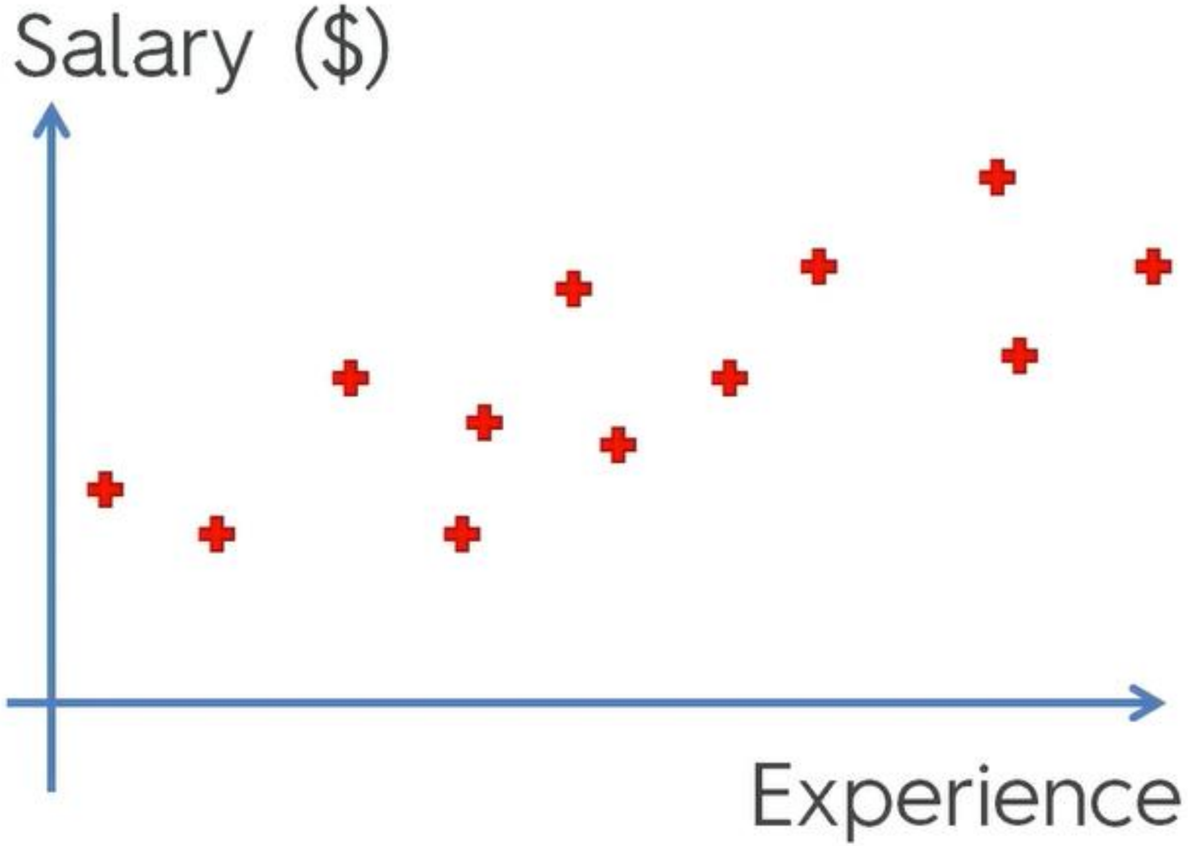
- You can use simple linear regression when you want to know:
  - **The relationship is between two variables**
  - e.g. the relationship between salary and experience.
  - The value of the dependent variable at a certain value of the independent variable
  - e.g. the salary at a certain level of experience.

Experience (years) (IDV)	Salary (DV)
1.1	39,343
1.3	46,205
1.5	37,731
2	43,525
2.2	39,891
2.9	56,642
3	60,150

Simple Linear Regression:

Experience (years) (IDV)	Salary (DV)
1.1	39,343
1.3	46,205
1.5	37,731
2	43,525
2.2	39,891
2.9	56,642
3	60,150
:	:
:	:

Simple Linear Regression:



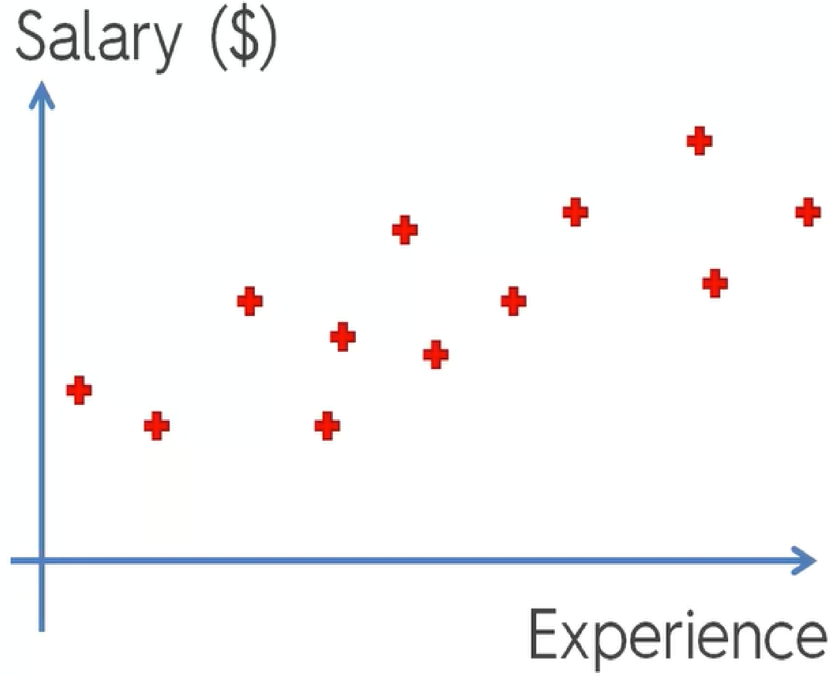
Experience (years) (IDV)	Salary (DV)
1.1	39,343
1.3	46,205
1.5	37,731
2	43,525
2.2	39,891
2.9	56,642
3	60,150
:	:
:	:

How to perform a simple linear regression?

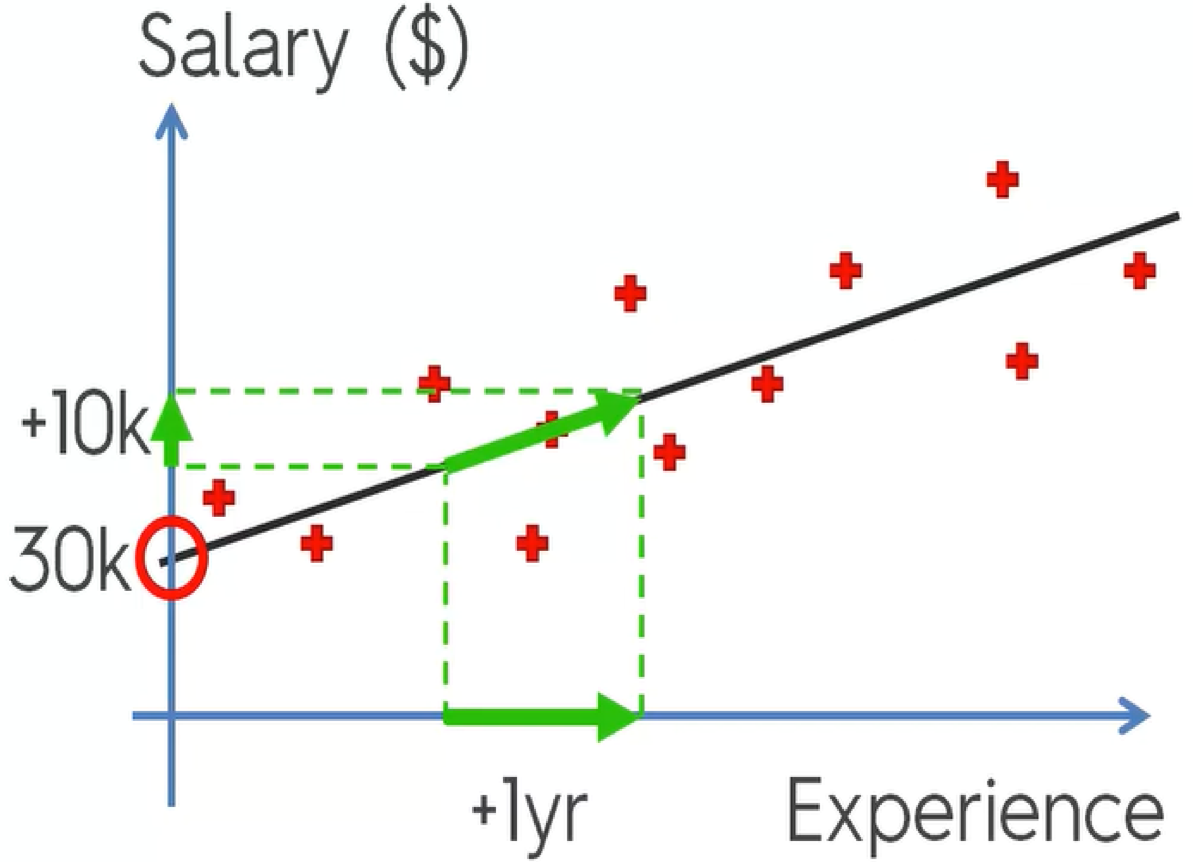
$$y = b_0 + b_1 * x_1$$

$$y = b_0 + b_1 * x_1$$

Dependent variable (DV) Independent variable (IV)



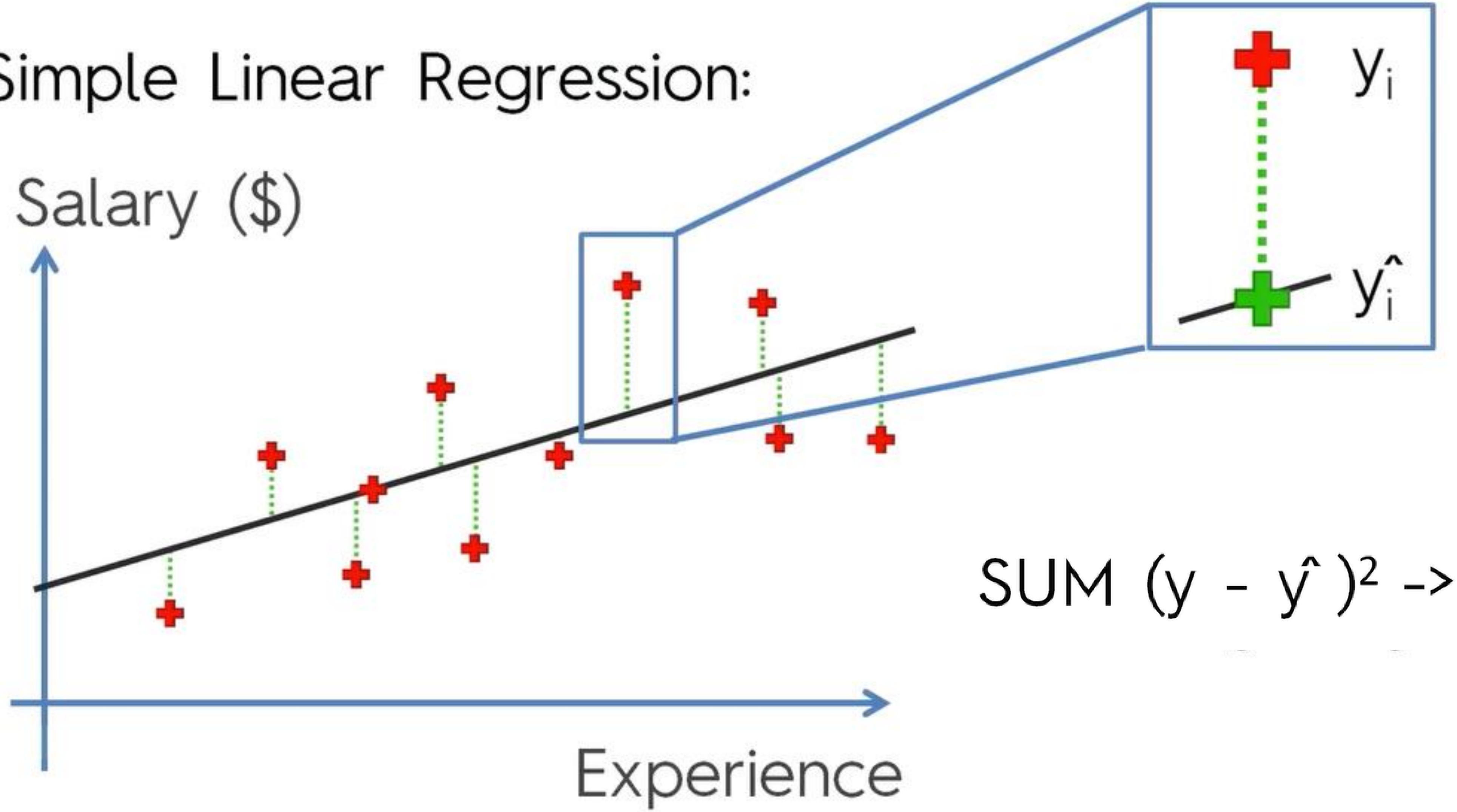
Simple Linear Regression:



$$y = b_0 + b_1{*}x$$

$$\text{Salary} = b_0 + b_1{*}\text{Experience}$$

Simple Linear Regression:



SUM $(y - \hat{y})^2$ -> min

Population versus Sample

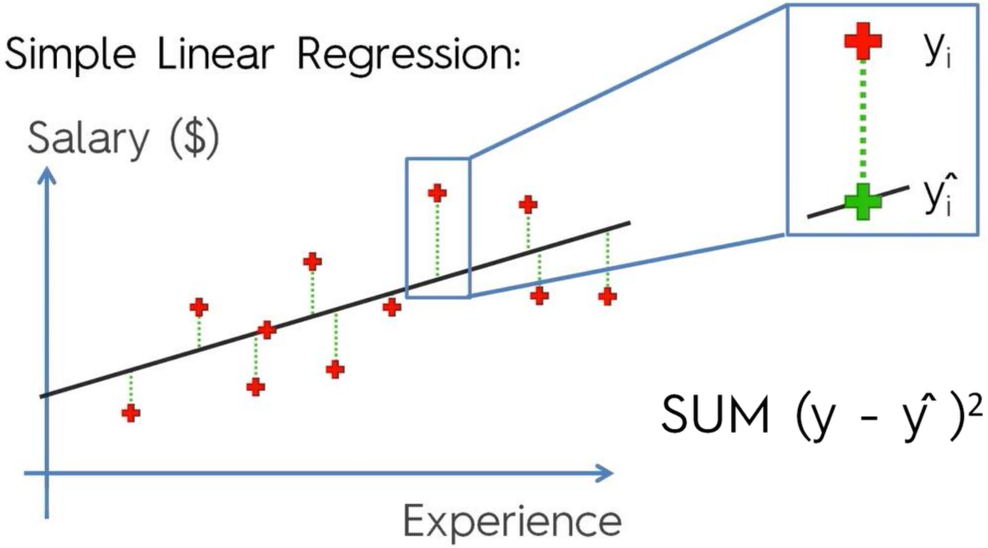
- Population Regression Function

  - Deterministic Component

$$y = f(x)$$

  - Stochastic Component

$$y = f(x) + \epsilon$$



Population versus Sample

- Normally distributed error component
- Univariate function $y = wx + b + \epsilon$
- Multivariate function

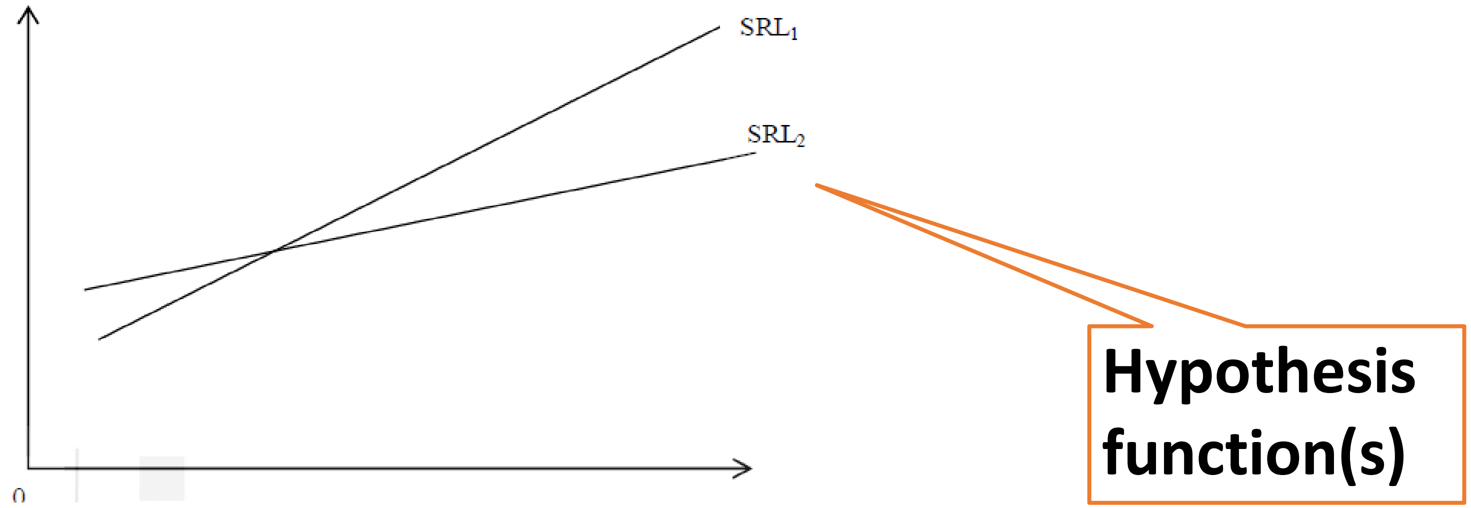
$$y = w_n x_n + ...w_1 x_1 + w_0 + \epsilon$$
$$= \mathbf{w}^T \mathbf{x} + \epsilon$$

$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ .. \\ w_n \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ .. \\ x_n \end{bmatrix}$$

Population versus Sample

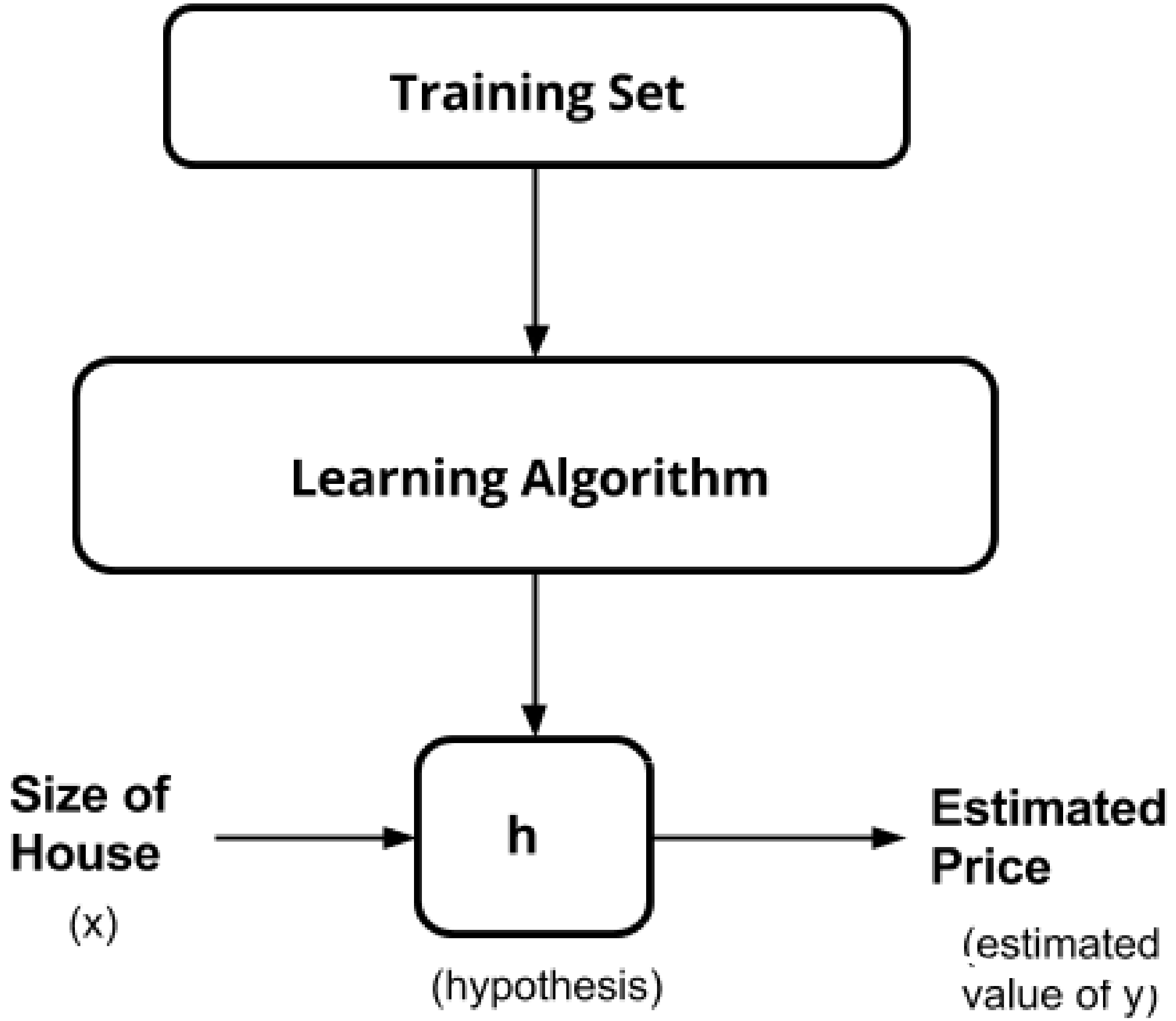
- Sample Regression Functions
  - Different Regression Line/Plane/Hyperplane



- Univariate function $\hat{y} = h(x) = wx + b$

- Multivariate function $\hat{y} = h(x) = \mathbf{w}^T\mathbf{x}$

Hypothesis function(s)



Linear Regression algorithm

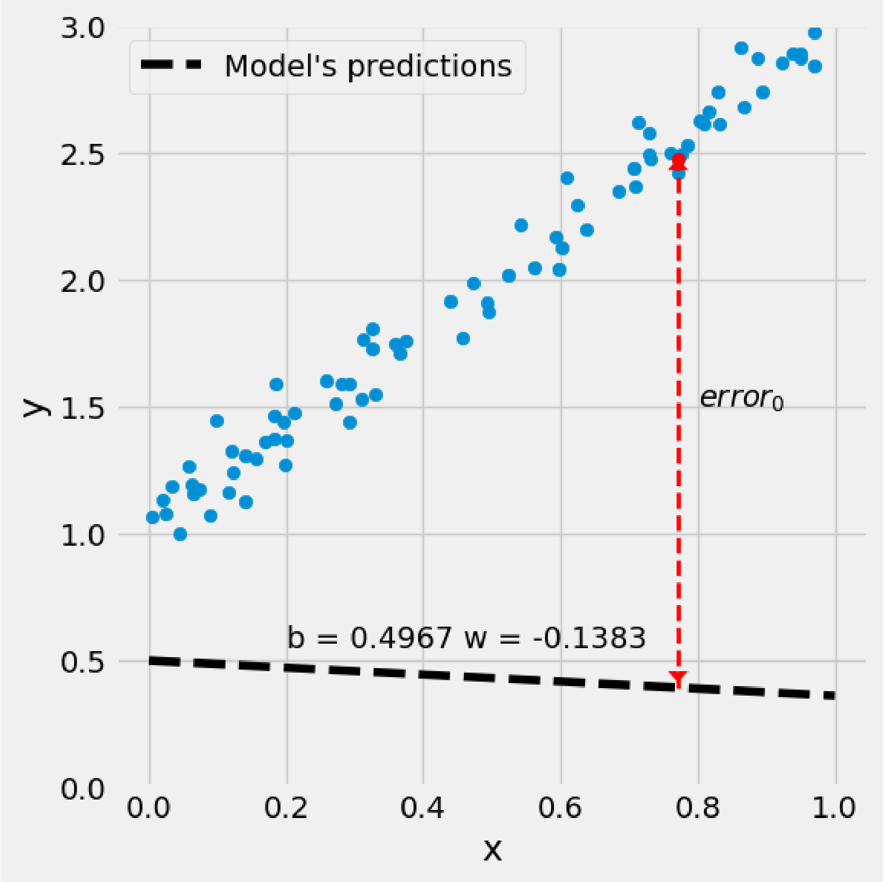
Step 1. Initialization

- Assume parametric form $\hat{y} = wx + b$

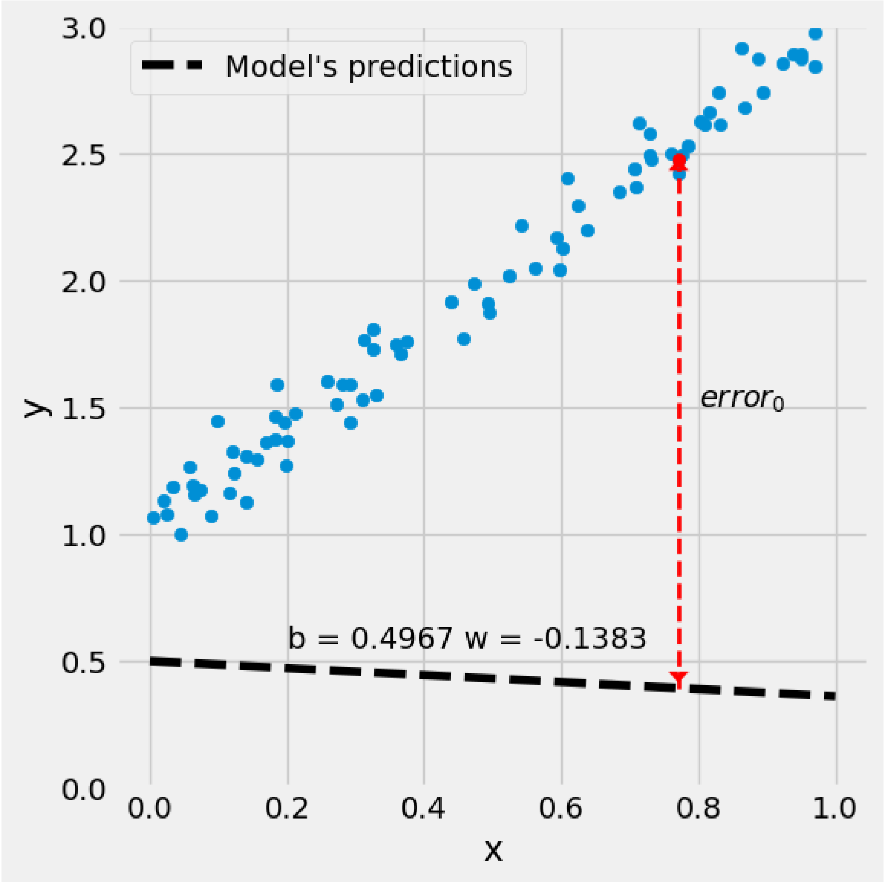
Parametric form of the hypothesis function

- Calculate error using the initial w & b

$$error^{(i)} = \hat{y}^{(i)} - y^{(i)}$$



Formulate Objective function



$$error^{(i)} = \hat{y}^{(i)} - y^{(i)}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{error}^{(i)^2}$$

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

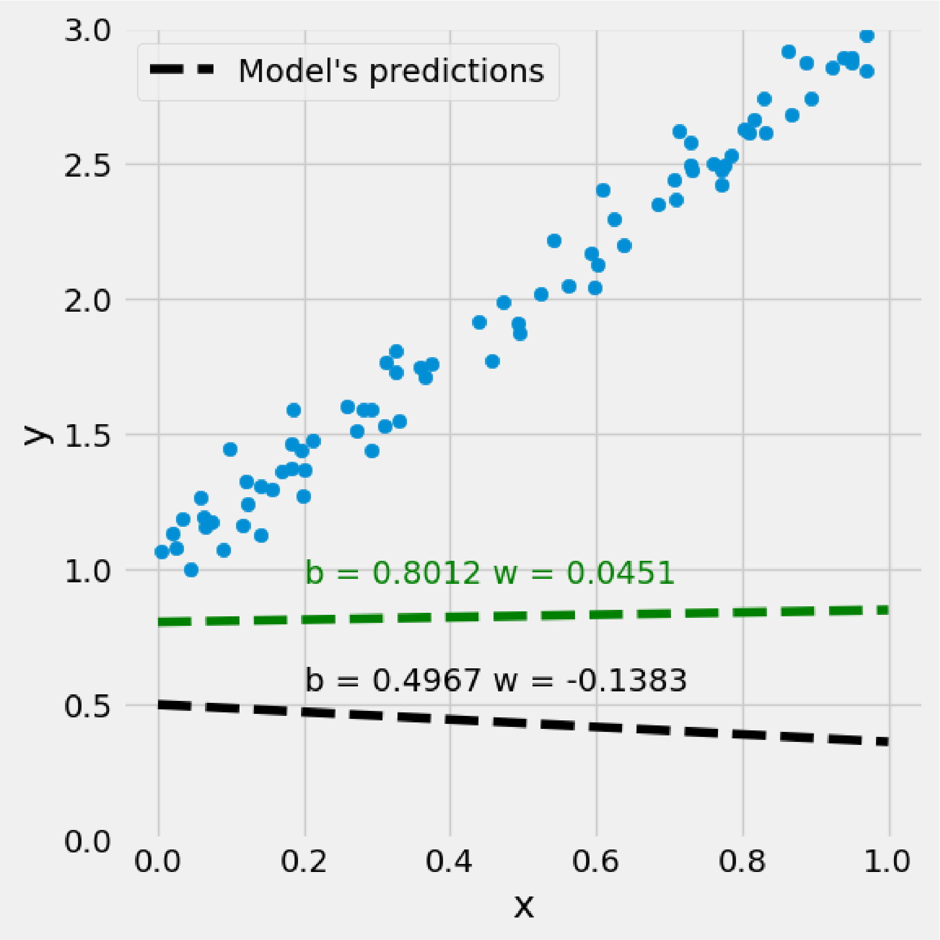
$$\mathcal{J}(w, b) = \frac{1}{n}\sum_{i=1}^{n} \left(b + wx^{(i)} - y^{(i)}\right)^2$$

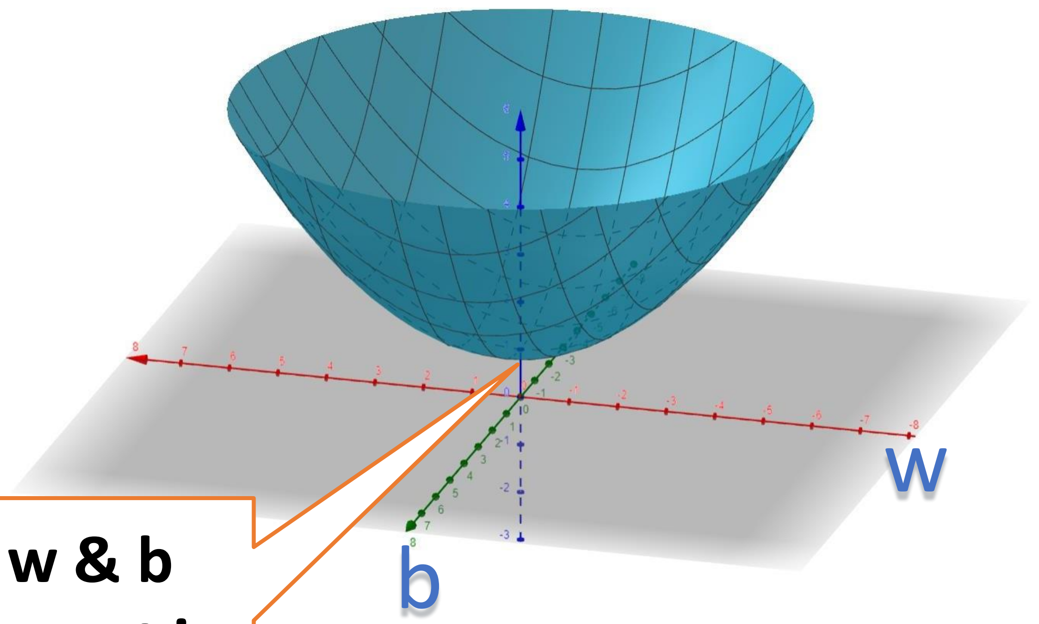
Also called cost / loss function

Feature space versus parameter space

$$\hat{y} = wx + b$$

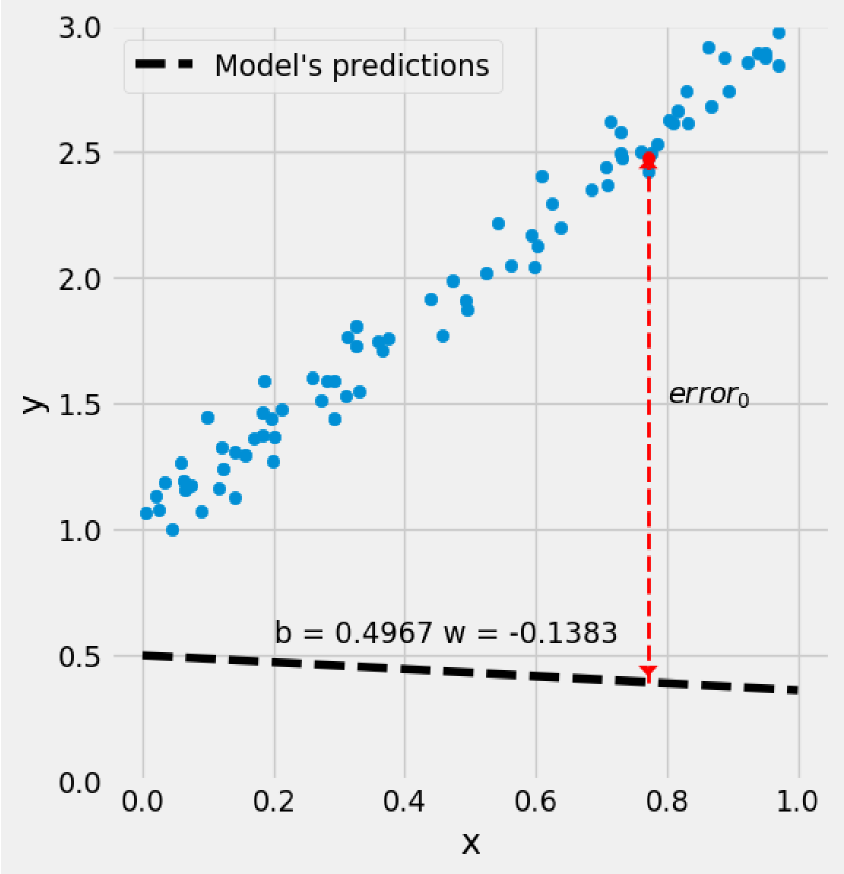
$$\mathcal{J}(w,b) = \frac{1}{n}\sum_{i=1}^{n}\left(b + wx^{(i)} - y^{(i)}\right)^2$$





Find the w & b when the cost is minimum

Step 2. Evaluate y-hat & evaluate objective function

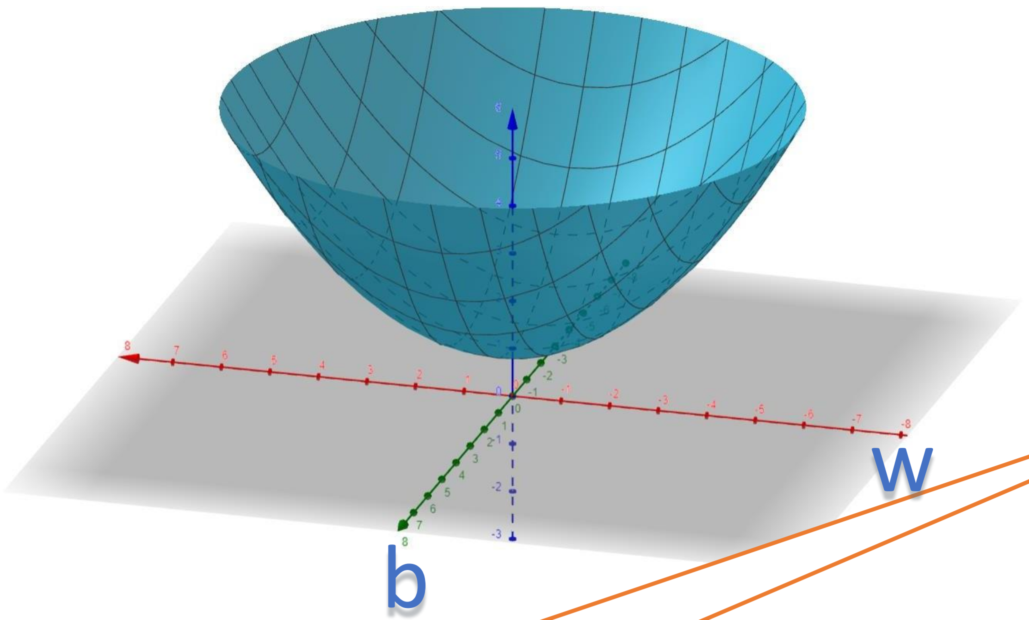


- Evaluate y-hat $\hat{y} = wx + b$
- Evaluate objective function (also known as forward pass)

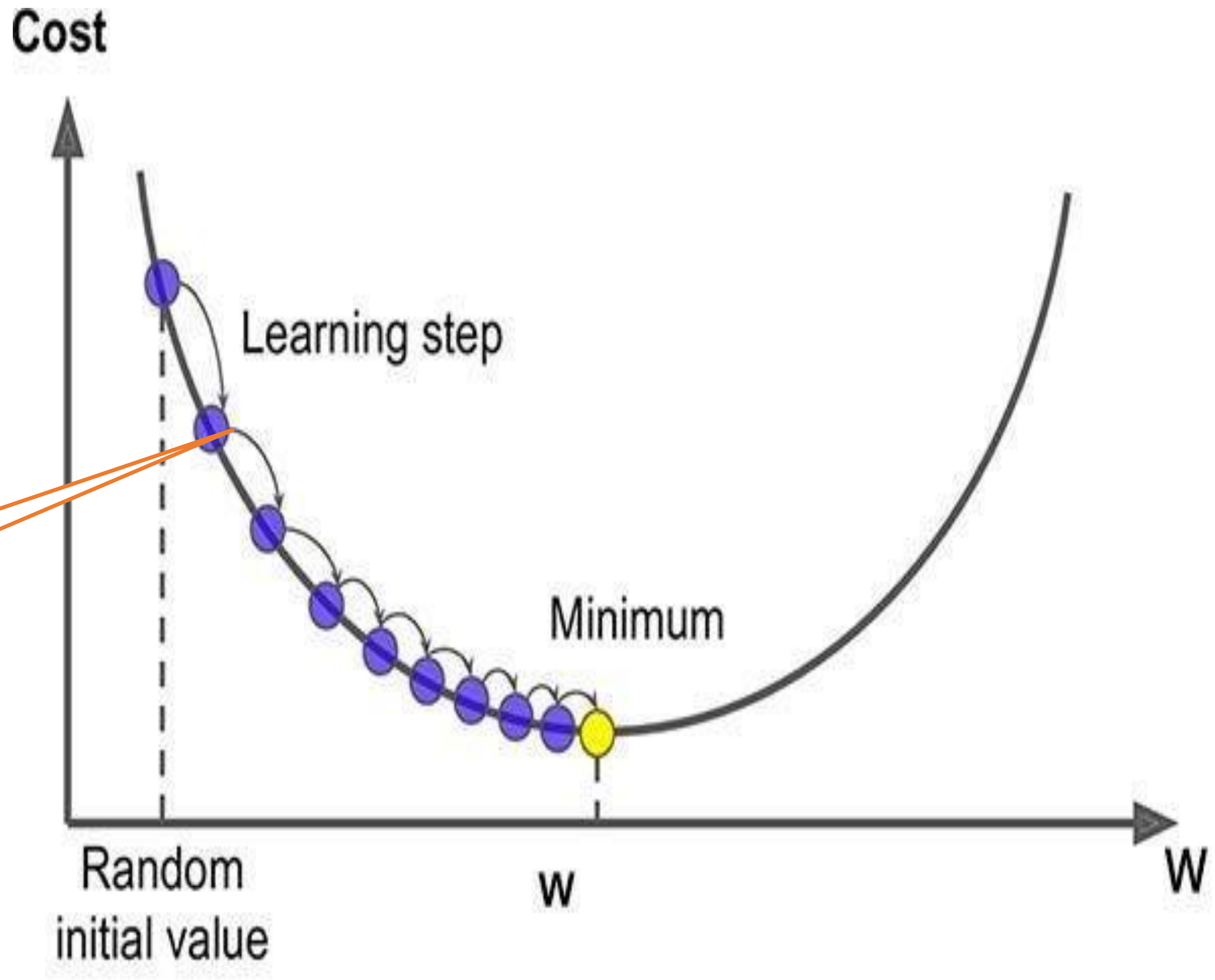
$$\mathcal{J}(w,b) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$$\mathcal{J}(w,b) = \frac{1}{n}\sum_{i=1}^{n}\left(b + wx^{(i)} - y^{(i)}\right)^2$$

Objective function plot



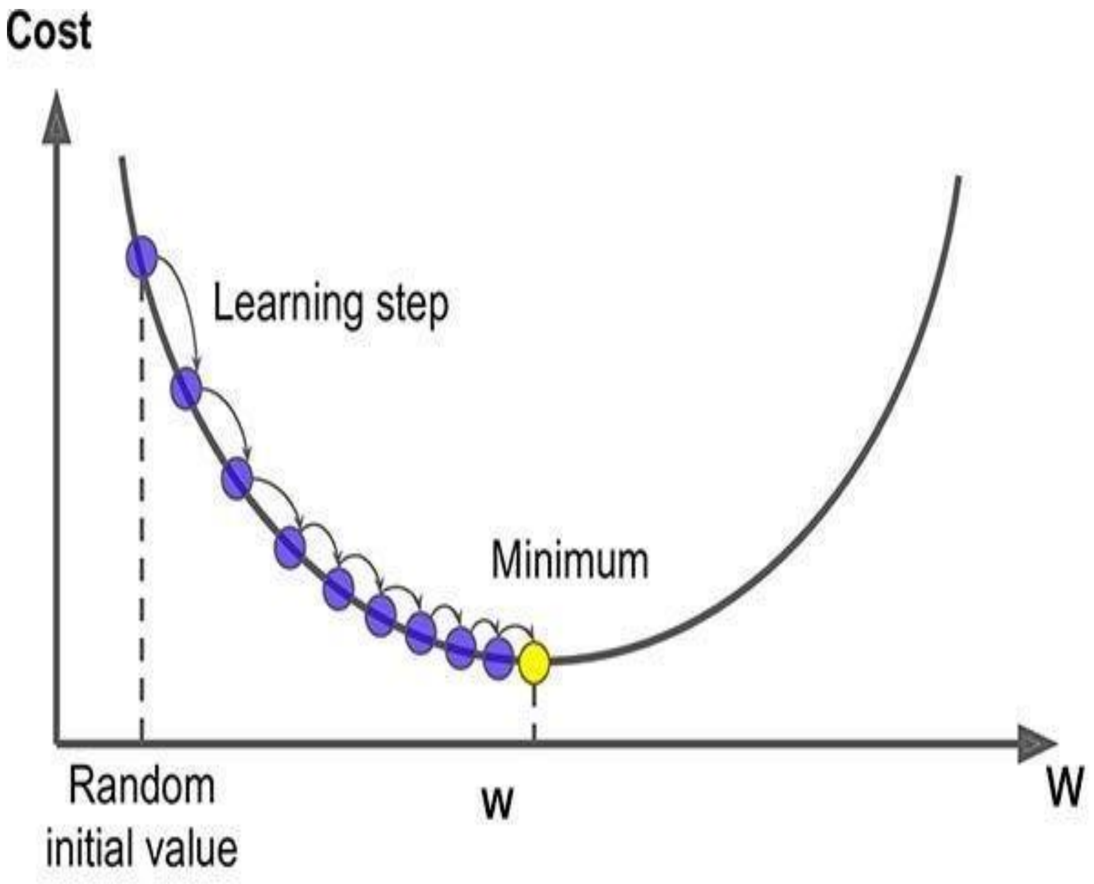
Take a little step towards lowest cost.
1. How little?
2. Which direction?



Step 3. Calculate analytical gradients

$$\mathcal{J}(w,b) = \frac{1}{n}\sum_{i=1}^{n}\left(b + wx^{(i)} - y^{(i)}\right)^2$$

$$\mathcal{J}(w,b) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$



Calculate gradient for m sample steps

$$\frac{\partial \mathcal{J}}{\partial b} = \frac{1}{m}\sum_{i=1}^{m} 2(b + wx^{(i)} - y^{(i)})$$

$$\frac{\partial \mathcal{J}}{\partial w} = \frac{1}{m}\sum_{i=1}^{m} 2x^{(i)}(b + wx^{(i)} - y^{(i)})$$

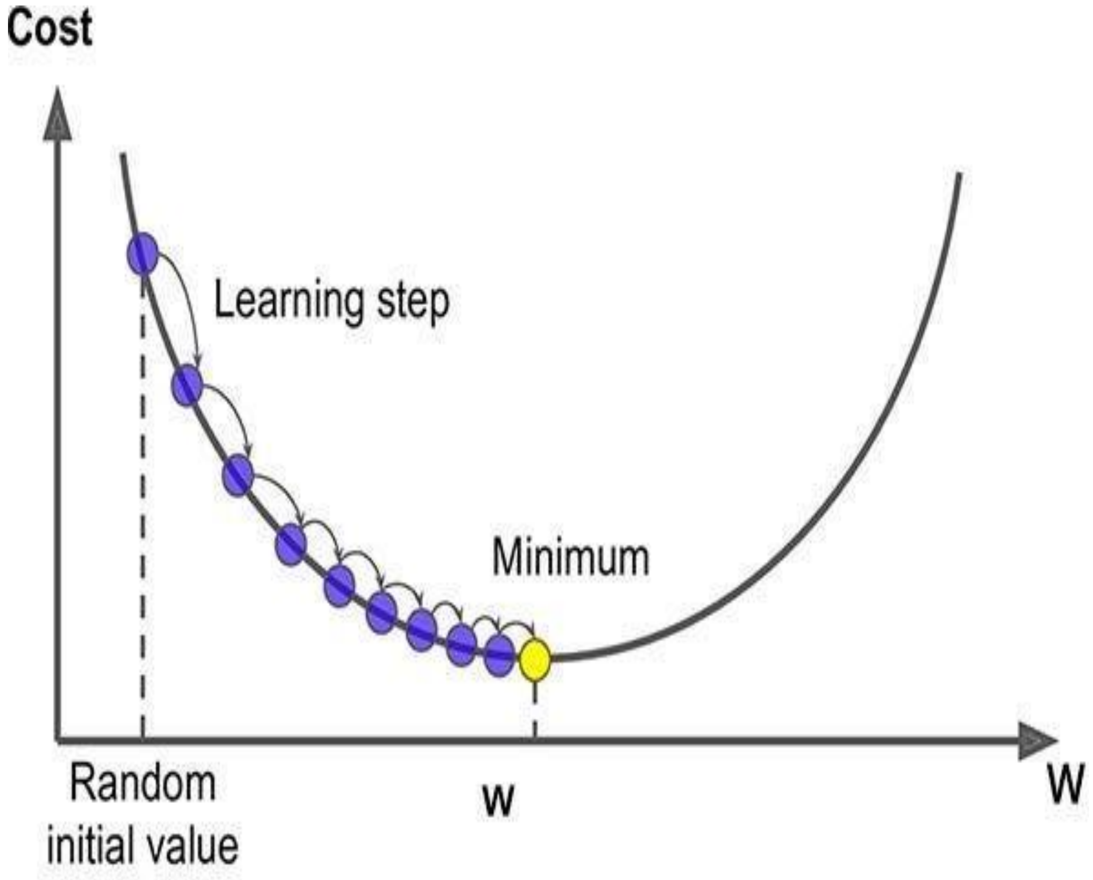
Step 4. Perform numerical gradient descent

- Backward pass – update b and w

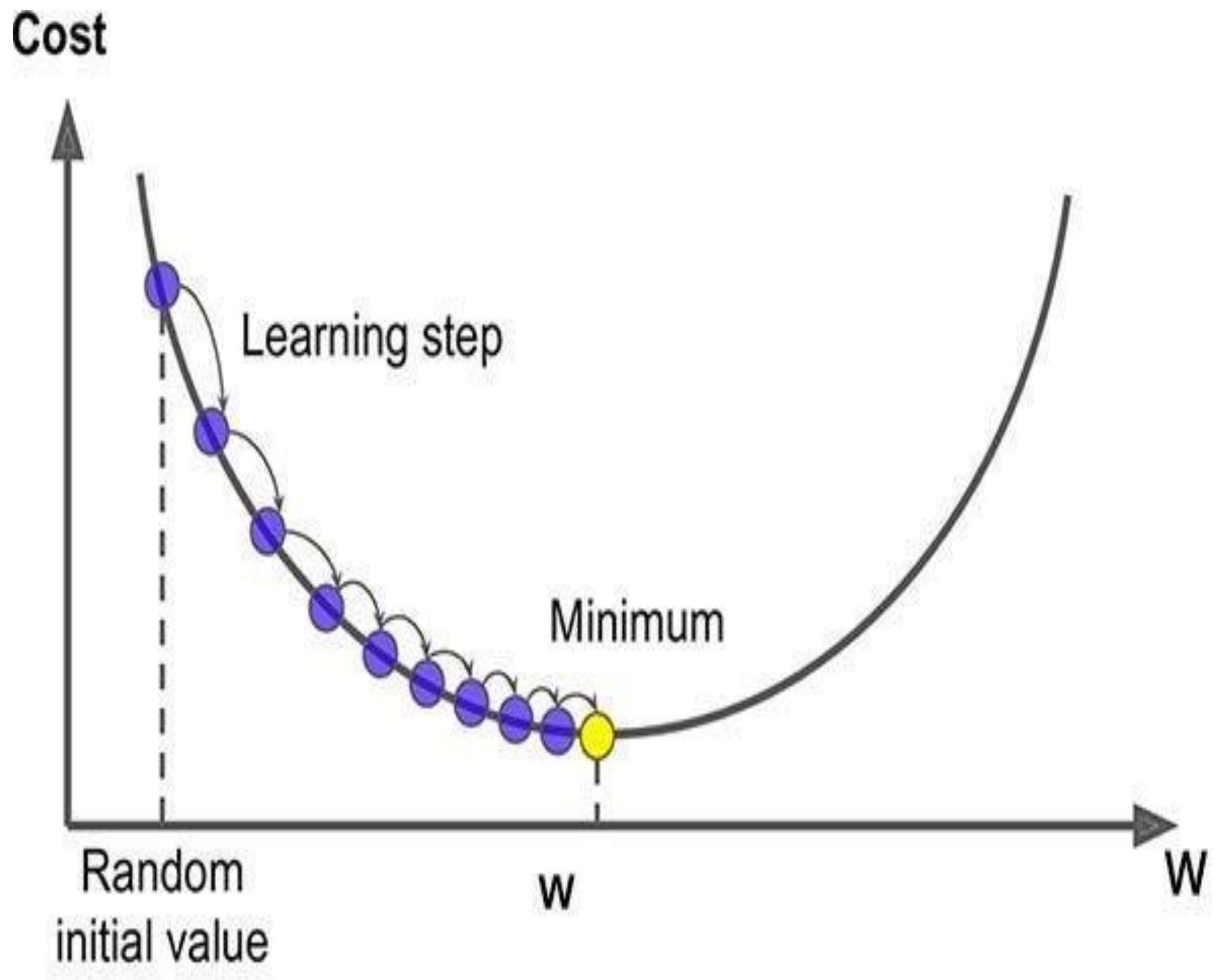
$$b = b - \eta \frac{\partial \mathcal{J}}{\partial b}$$

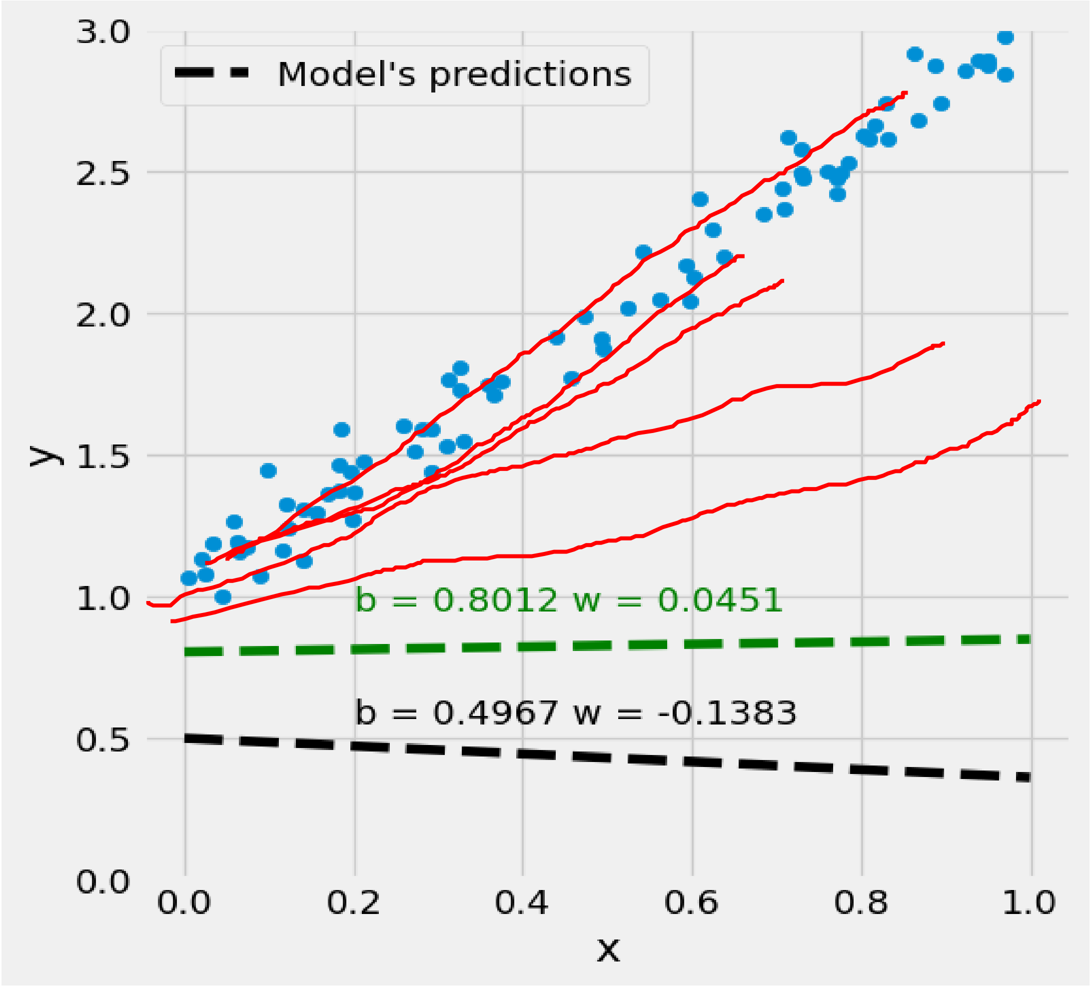
$$w = w - \eta \frac{\partial \mathcal{J}}{\partial w}$$

- Repeat Step 2, 3, 4

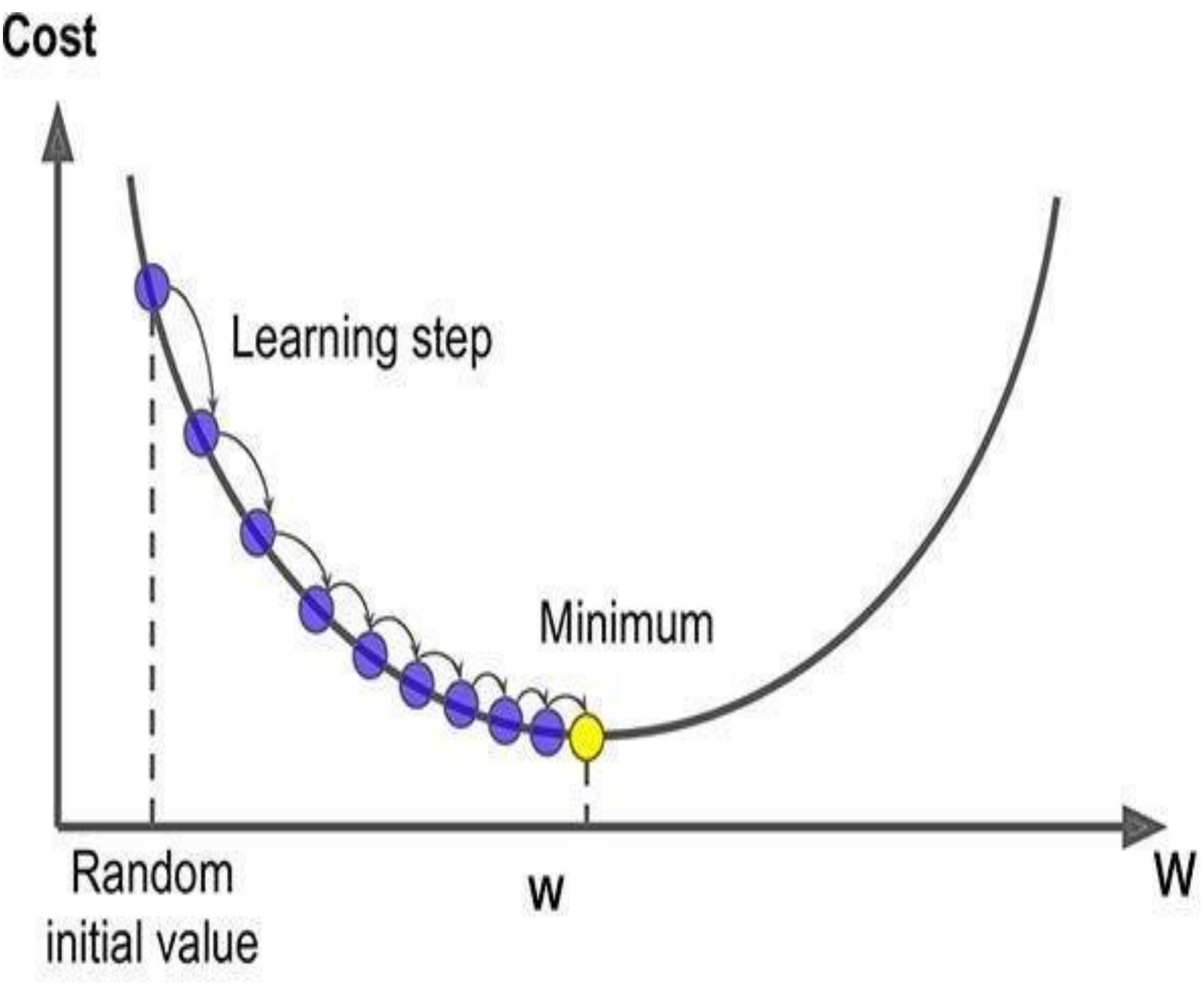


Change in w and b with gradient descent





Gradient descent summary



Iteration	w	b	Cost	dJ/dw	dJ/db
0	0.0110	0.0195	2.0443	-1.1077	-1.9538
1000	1.4309	1.2985	0.0178	-0.0407	0.02078
2000	1.7162	1.1527	0.0071	-0.0191	0.00976
3000	1.8502	1.0842	0.0047	-0.0090	0.00459
4000	1.9132	1.0520	0.0042	-0.0042	0.00215
5000	1.9430	1.0369	0.0041	-0.0020	0.00101
6000	1.9567	1.0298	0.0040	-0.0009	0.00047
7000	1.9632	1.0265	0.0040	-0.0004	0.00022
8000	1.9663	1.0249	0.0040	-0.0002	0.00010
9000	1.9677	1.0242	0.0040	-9.637e-05	4.925e-05

Example-1

Number of hours studied (x) IDV	**Marks scored (y) DV**	$\boldsymbol{X^2}$	$\boldsymbol{x.y}$	$(\hat{y} - y)^2$
2	**75**			
3	**82**			
4	**93**			
5	**89**			
6	**98**			

Example-1

$$\hat{y} = b_0 + b_1 * x$$

$$b_1 = \frac{\overline{xy} - \overline{x} * \overline{y}}{\overline{x^2} - (\overline{x})^2}$$

$$b_0 = \overline{y} - b_1 * \overline{x}$$

$$standard\ error = \sqrt{\frac{\sum(\hat{y} - y)^2}{n - 2}}$$

Assignment: Obtain a line that best fit the sample data given in the table. Evaluate the model by finding the standard error.

(x) IDV	Marks scored (y) DV		
1	2		
2	4		
3	5		
4	4		
5	5		

$$standard\ error = \sqrt{\frac{\sum(\hat{y} - y)^2}{n - 2}}$$

Thank you