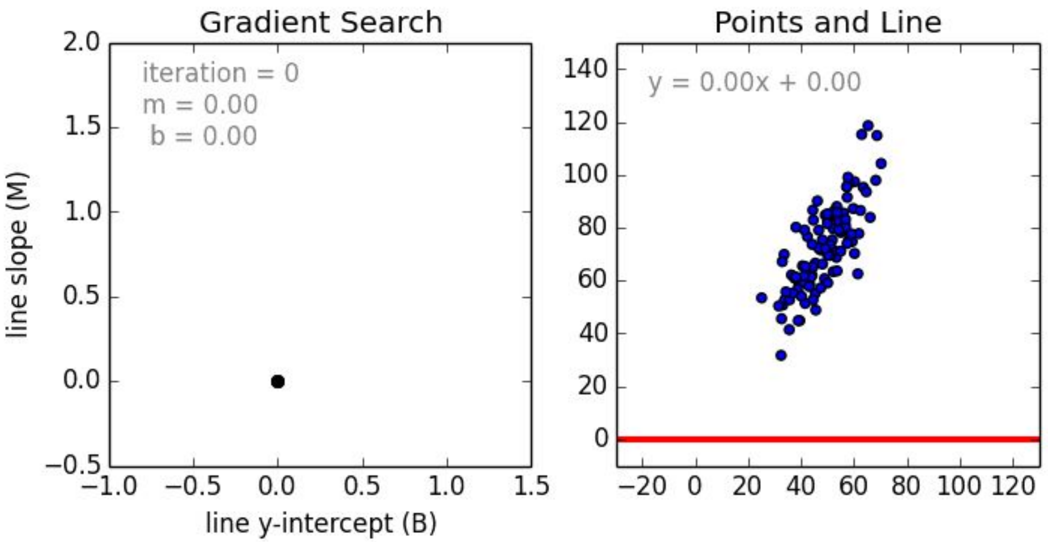
- What is Regression ?
- Regression Use – Case
- Types of Regression – Linear vs Logistic Regression
- What is Linear Regression ?
- Finding best regression line using Least Square Method
- Checking goodness of fit using R squared Method
- Implementation of Linear Regression using Python
    - Linear Regression Algorithm using Python from Scratch
    - Linear Regression Algorithm using Python (scikit lib)



What is Regression?

"Regression analysis. is a form of predictive modelling technique which investigates the relationship the relationship between a dependent and independent variable."

Uses of Regression

Three major uses of regression analysis are

- Determining the strength of predictors
- Forecasting an effect, and
- Trend Forecasting

Linear vs Logistic **Regression**

Basis	Linear Regression	Logistic Regression
Core Concept	The data is modelled using a straight line	The probability of some obtained event is represented as a linear function of a combination of predictor variables.
Used with	Continuous Variables	Categorical Variable
Output / prediction	Values of the Variable	Probability of occurrence of the event.
Accuracy and Goodness of fit	Measured by Loss, R squared, Adjusted R squared etc.	Accuracy, Precision, Recall, F1 Score , ROC curve, Confusion Matrix , etc.

Linear Regression

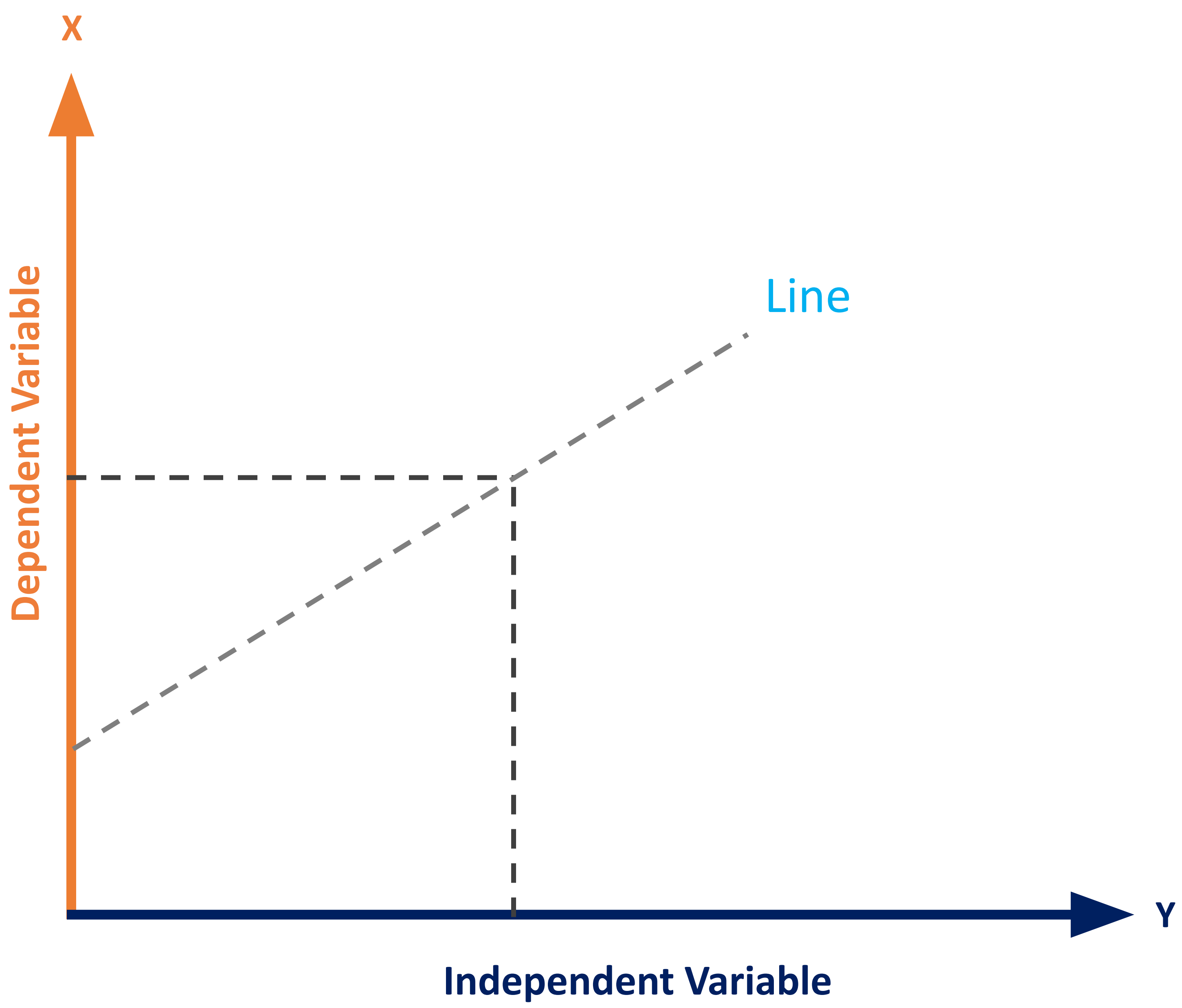
Selection Criteria

- Classification and Regression Capabilities
- Data Quality
- Computational Complexity
- Comprehensible and Transparent

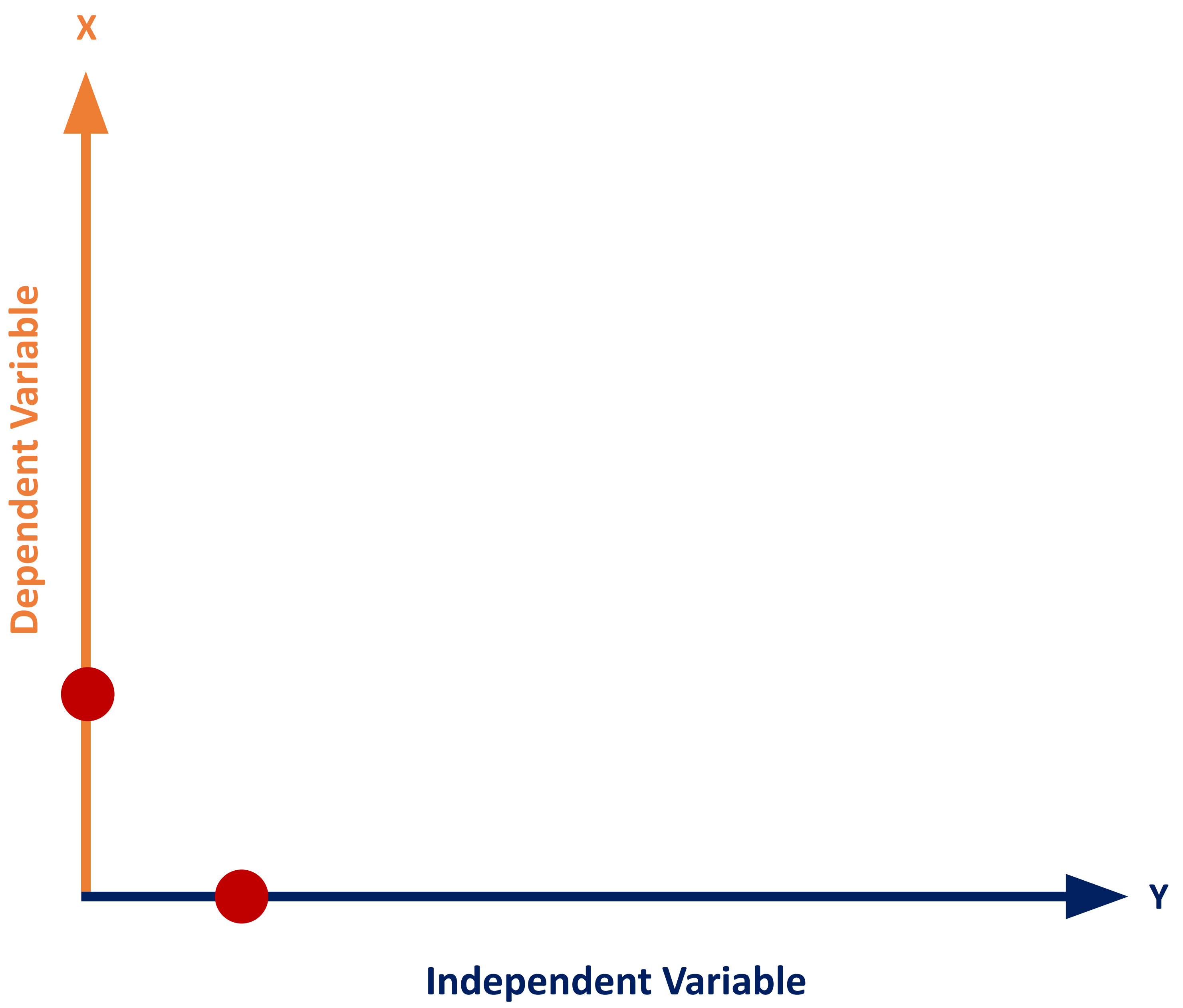
Where is Linear Regression Used ?

- Evaluating Trends and Sales estimates
- Analyzing the Impact of price Changes
- Assessment of risk in financial services and insurance domain

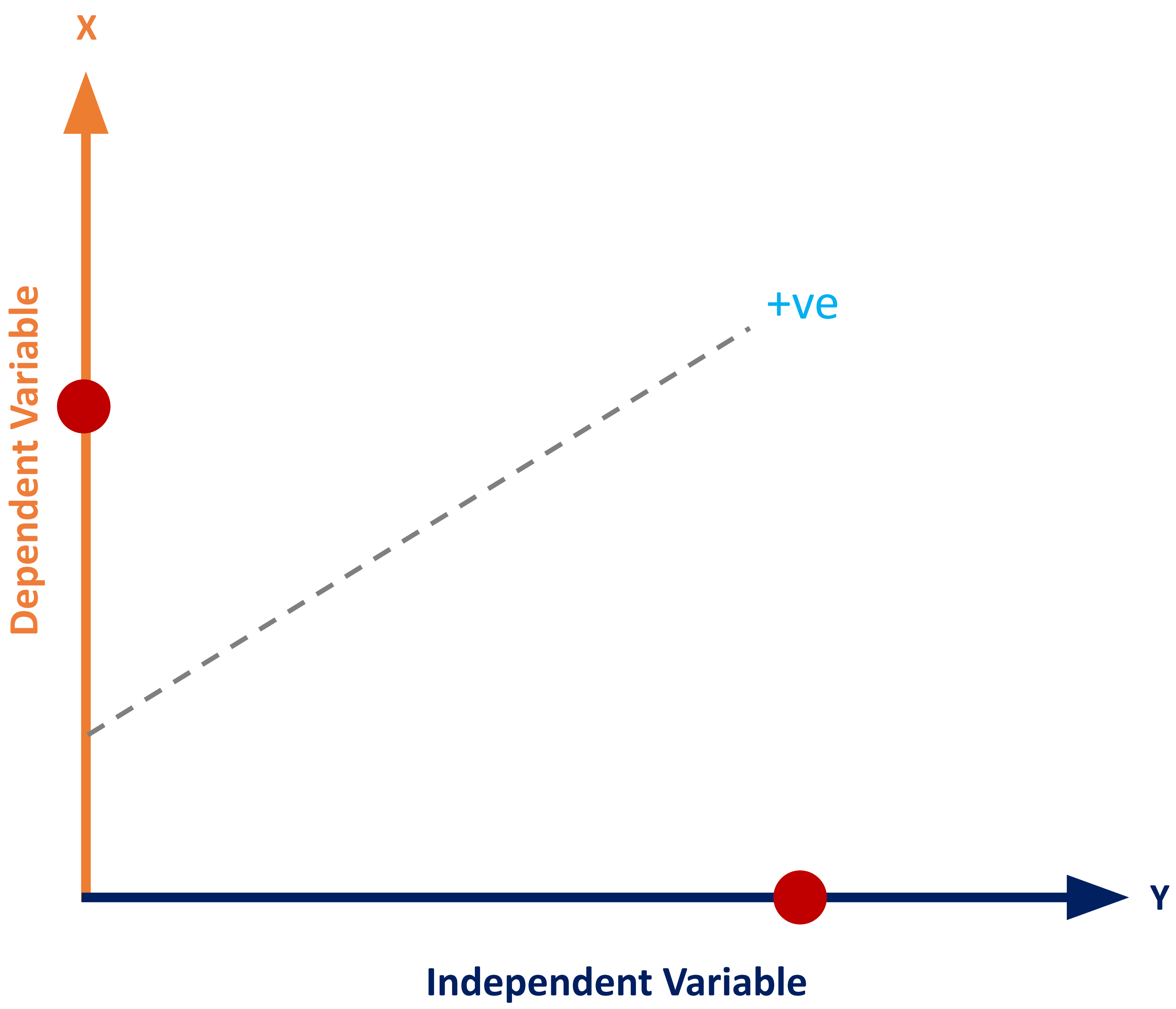
Understanding Linear Regression Algorithm



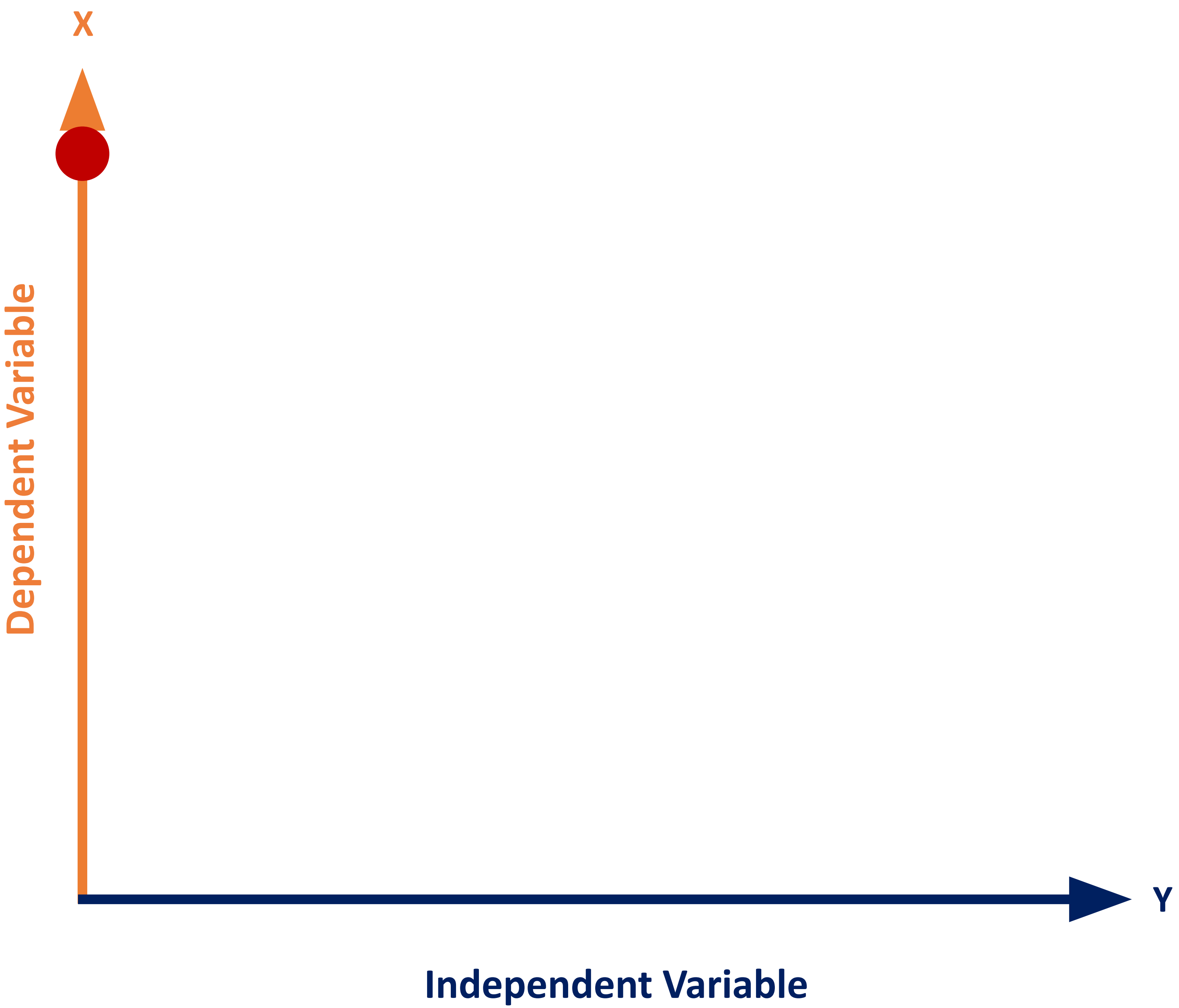
Understanding Linear Regression Algorithm



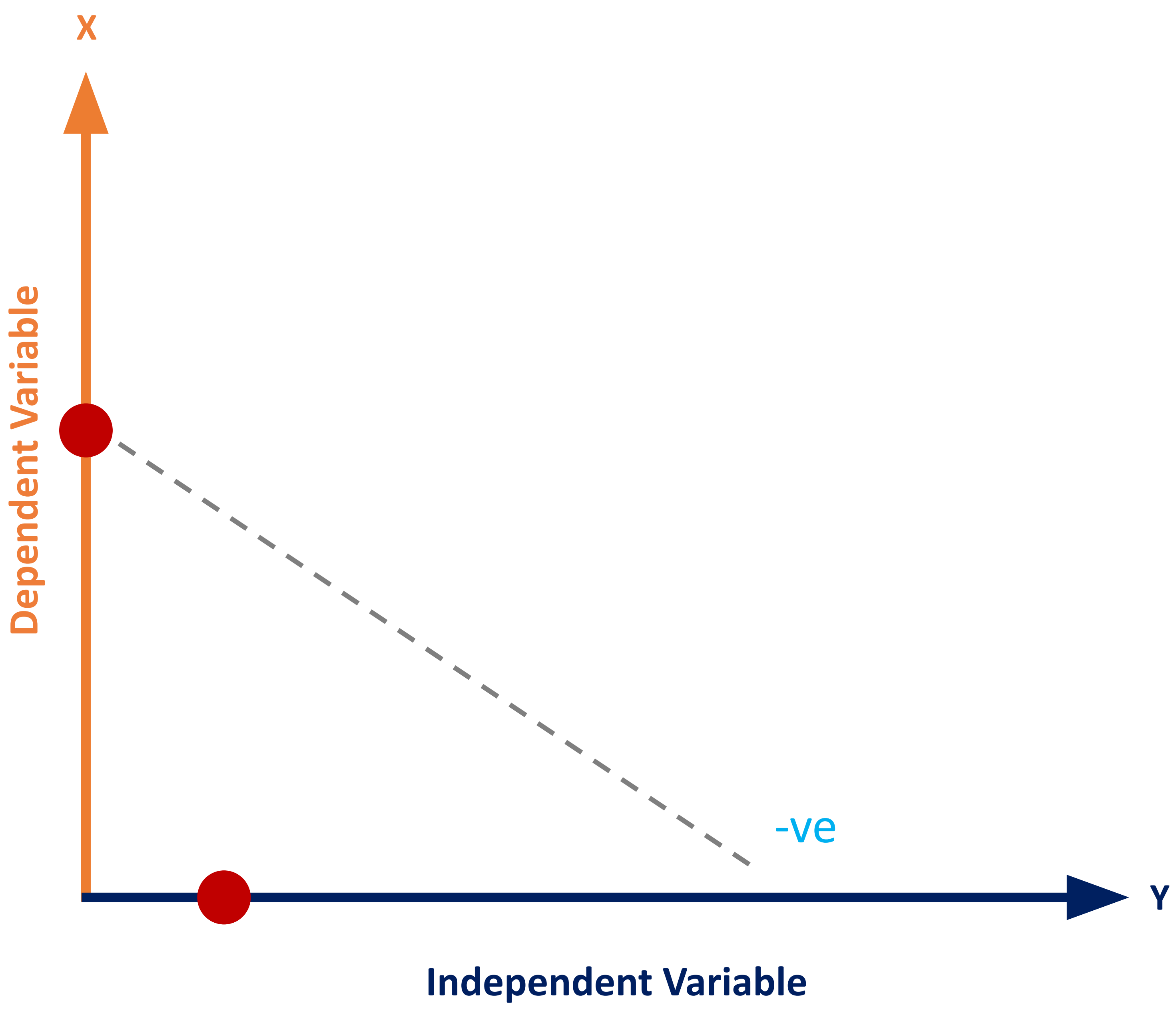
Understanding Linear Regression Algorithm



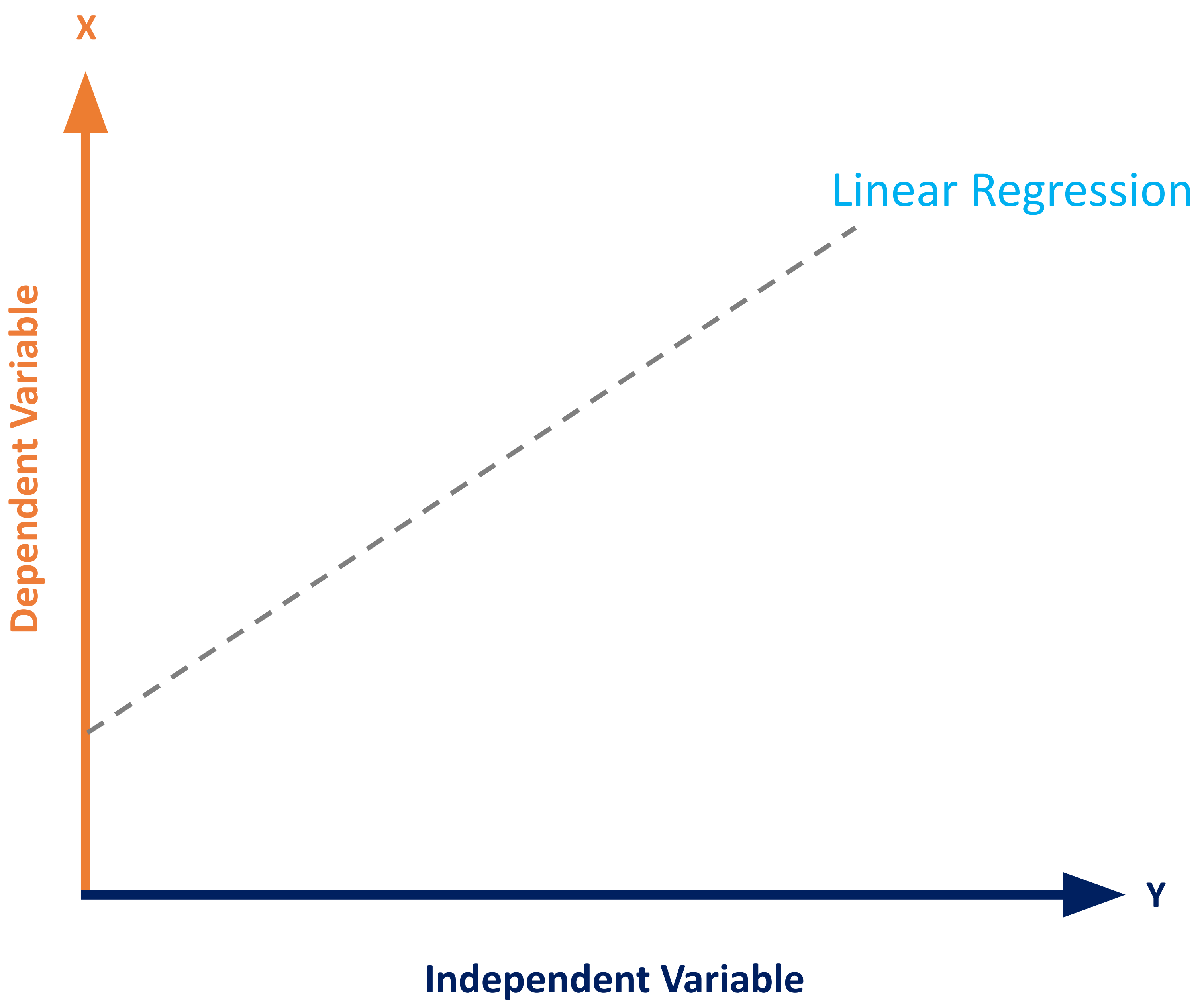
Understanding Linear Regression Algorithm



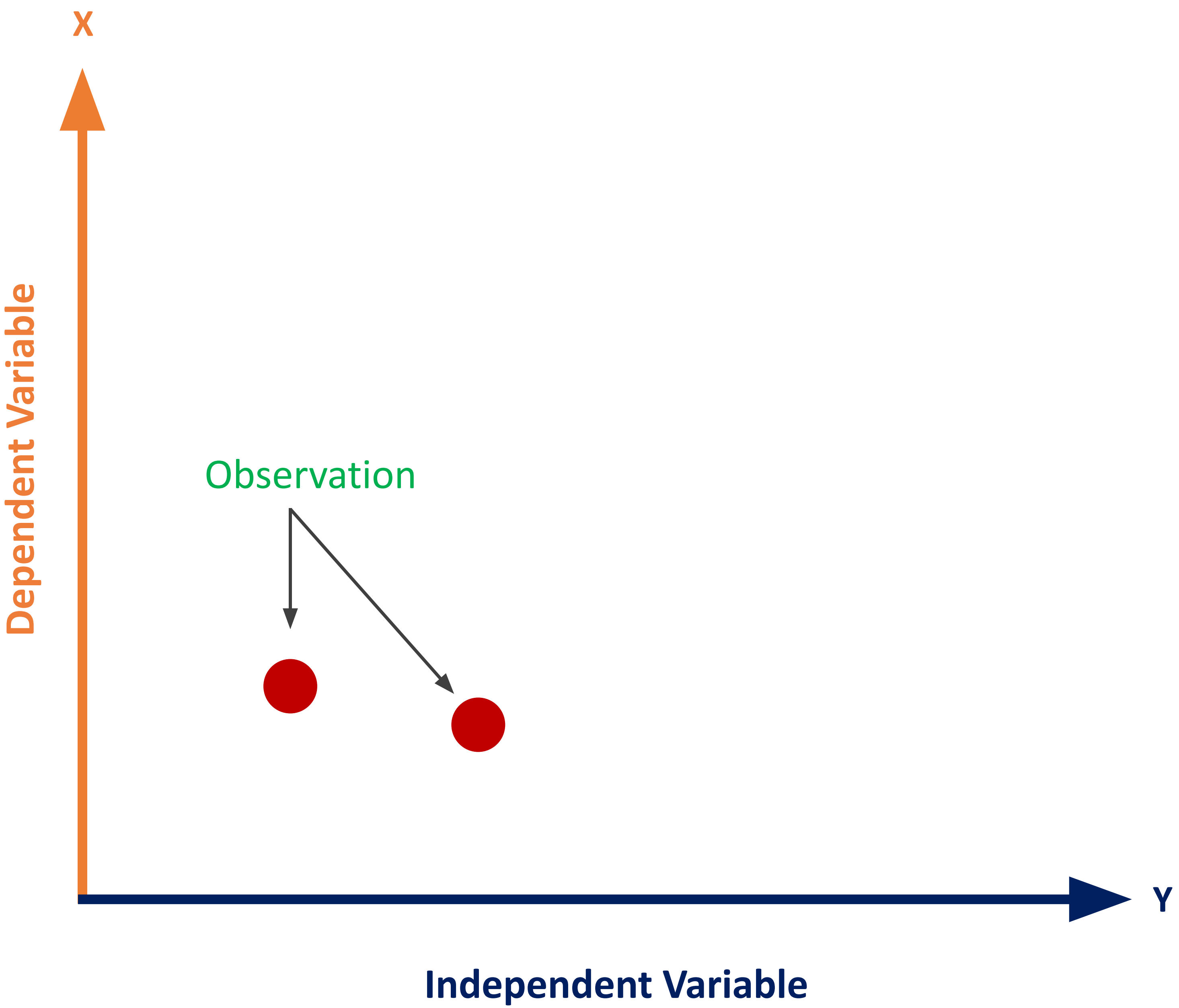
Understanding Linear Regression Algorithm



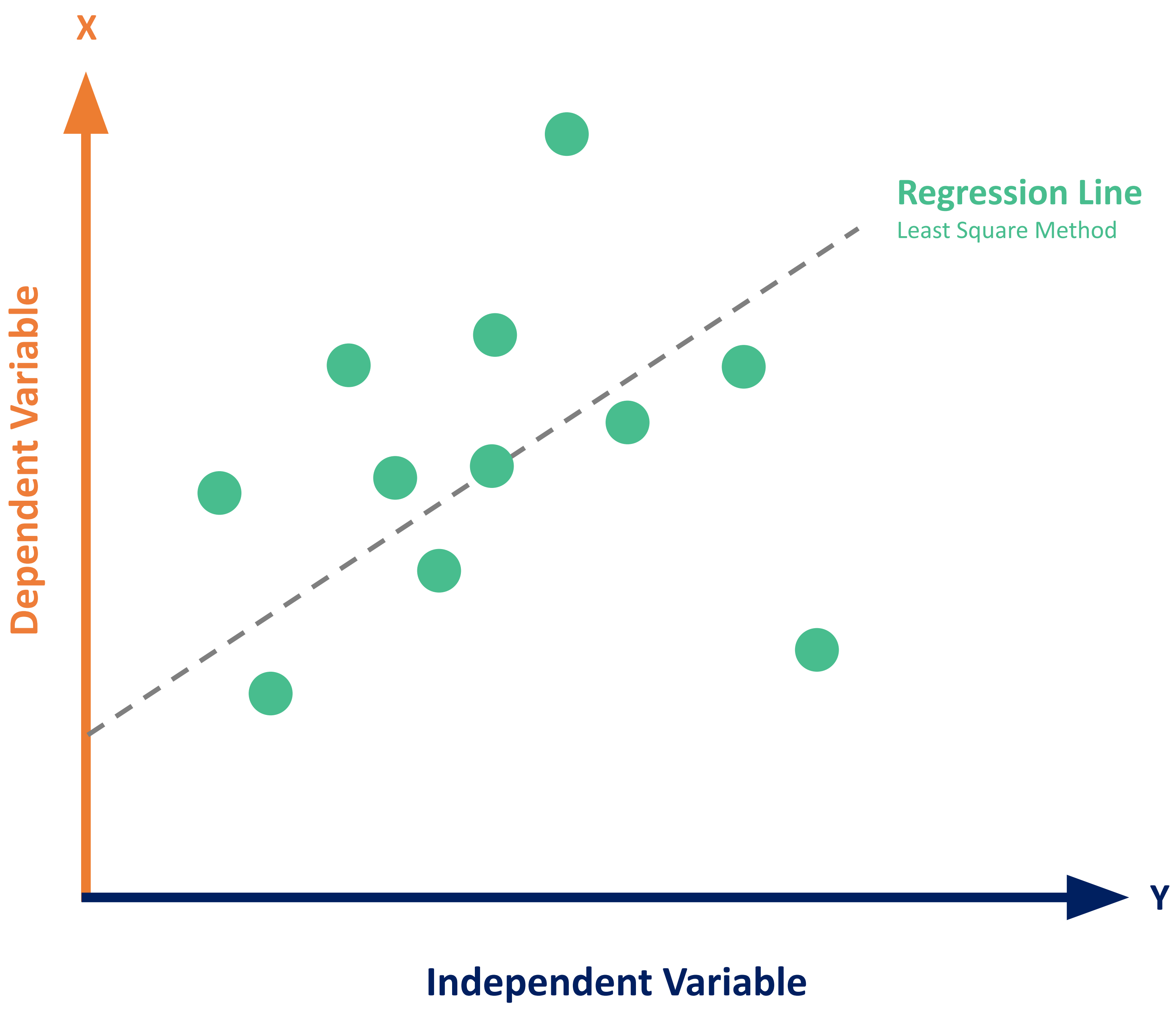
Understanding Linear Regression Algorithm



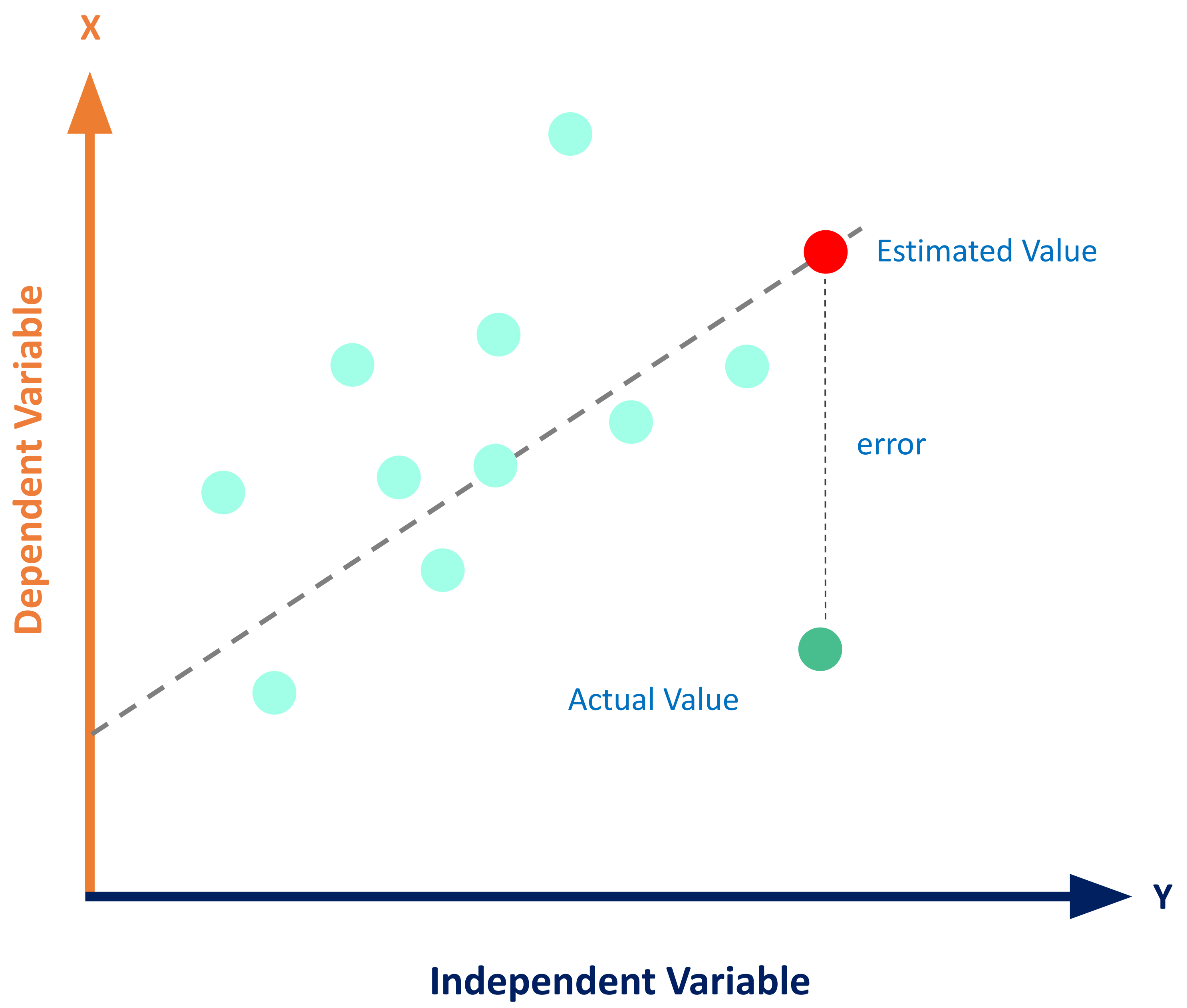
Understanding Linear Regression Algorithm



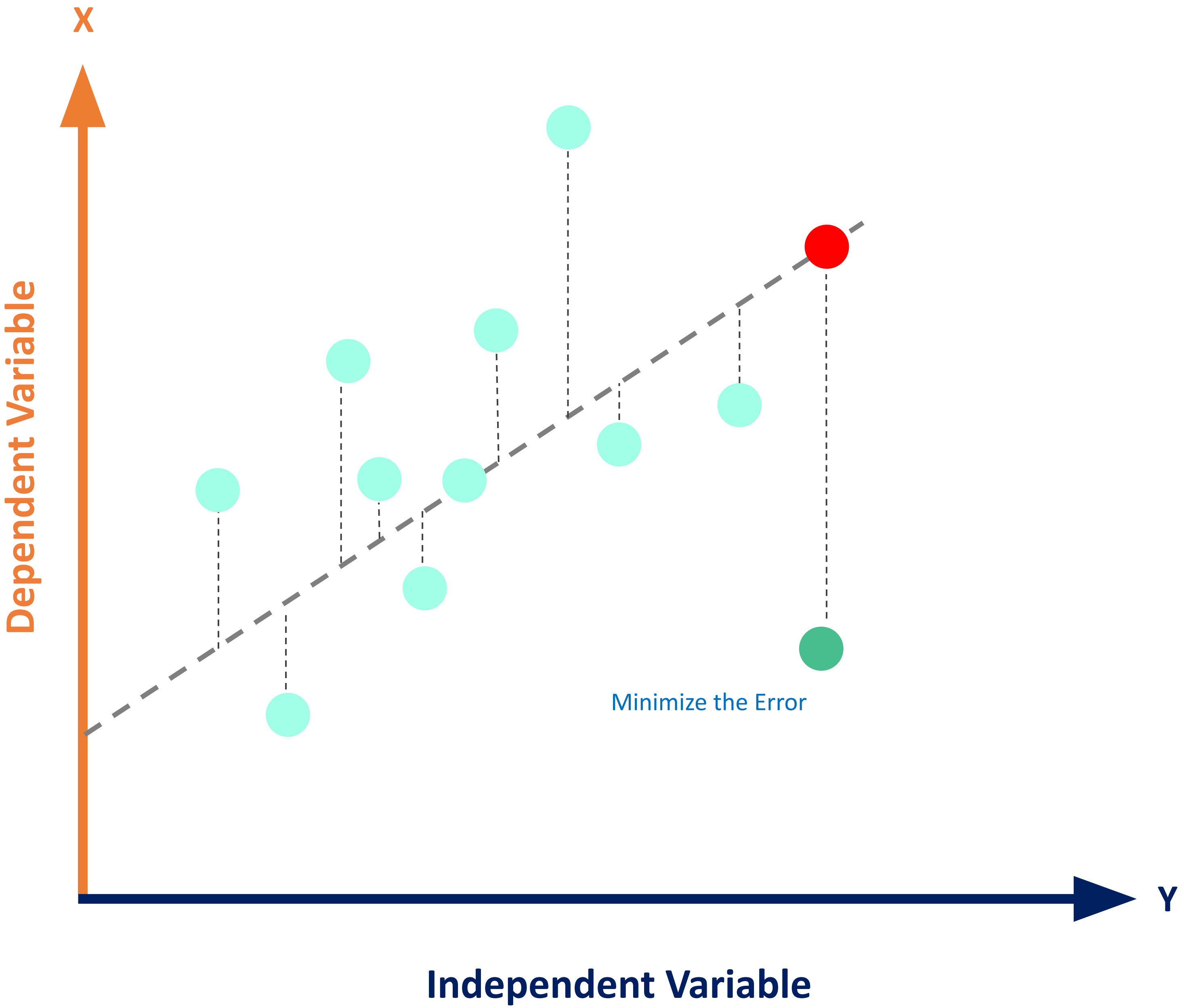
Understanding Linear Regression Algorithm



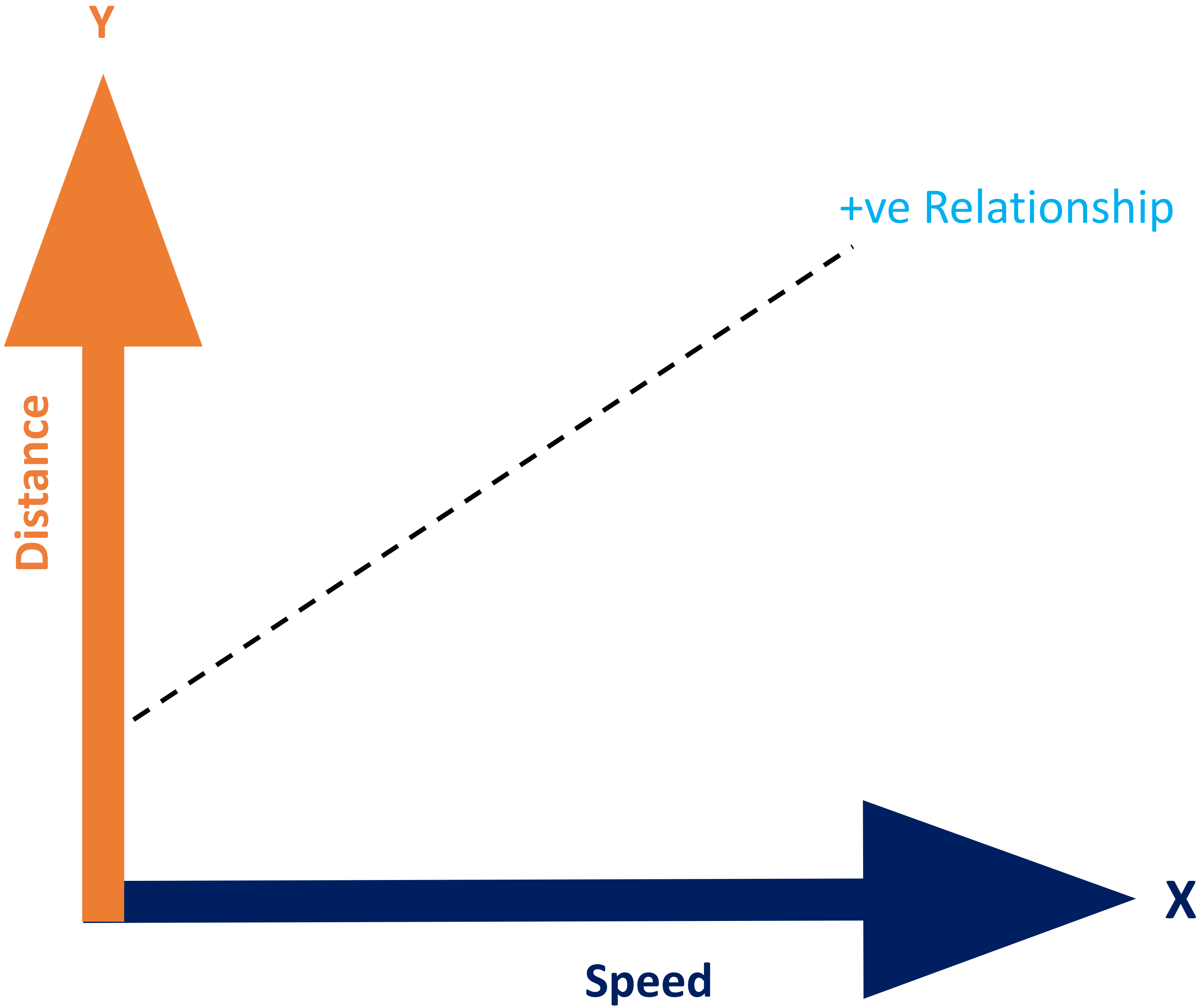
Understanding Linear Regression Algorithm



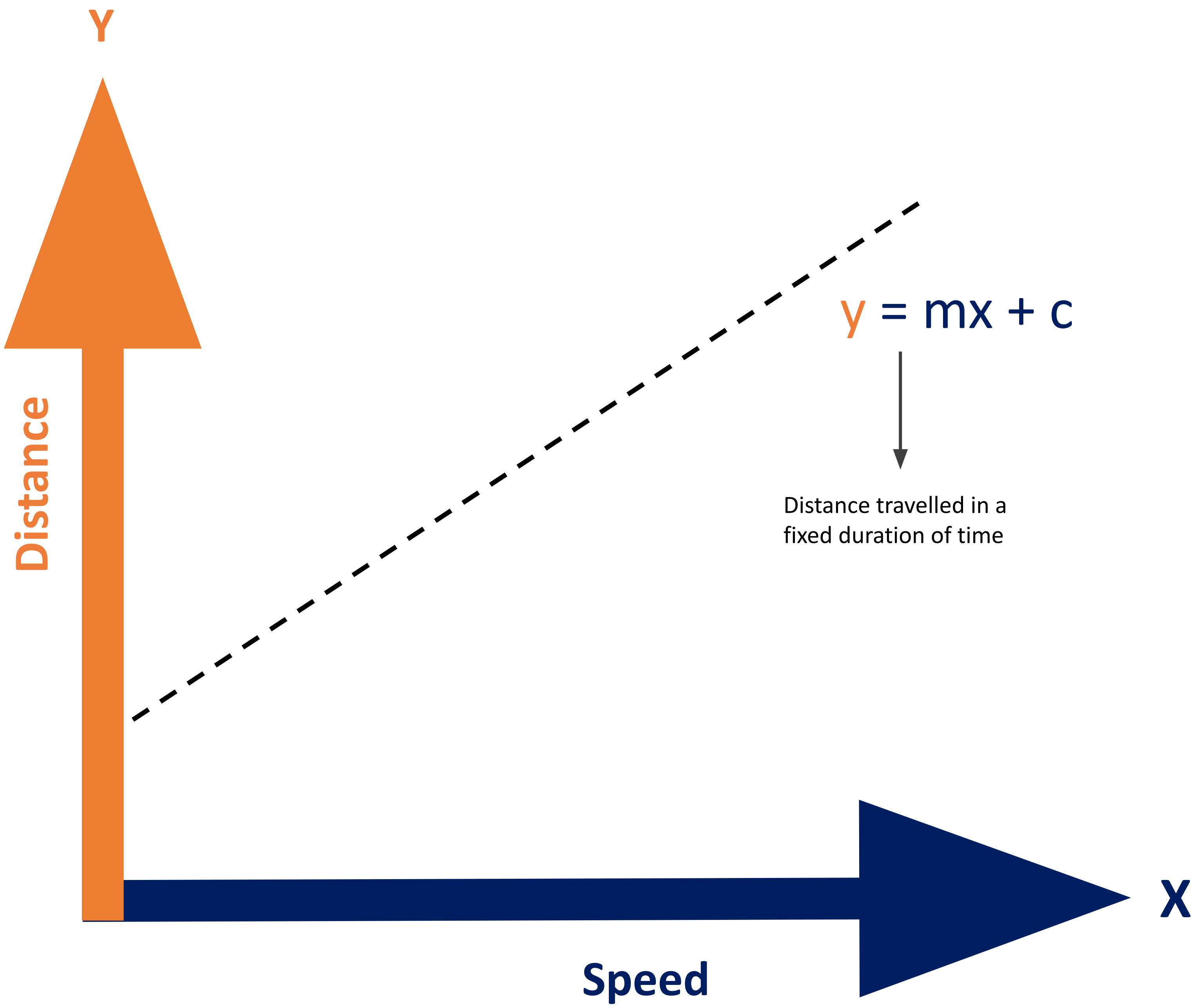
Understanding Linear Regression Algorithm



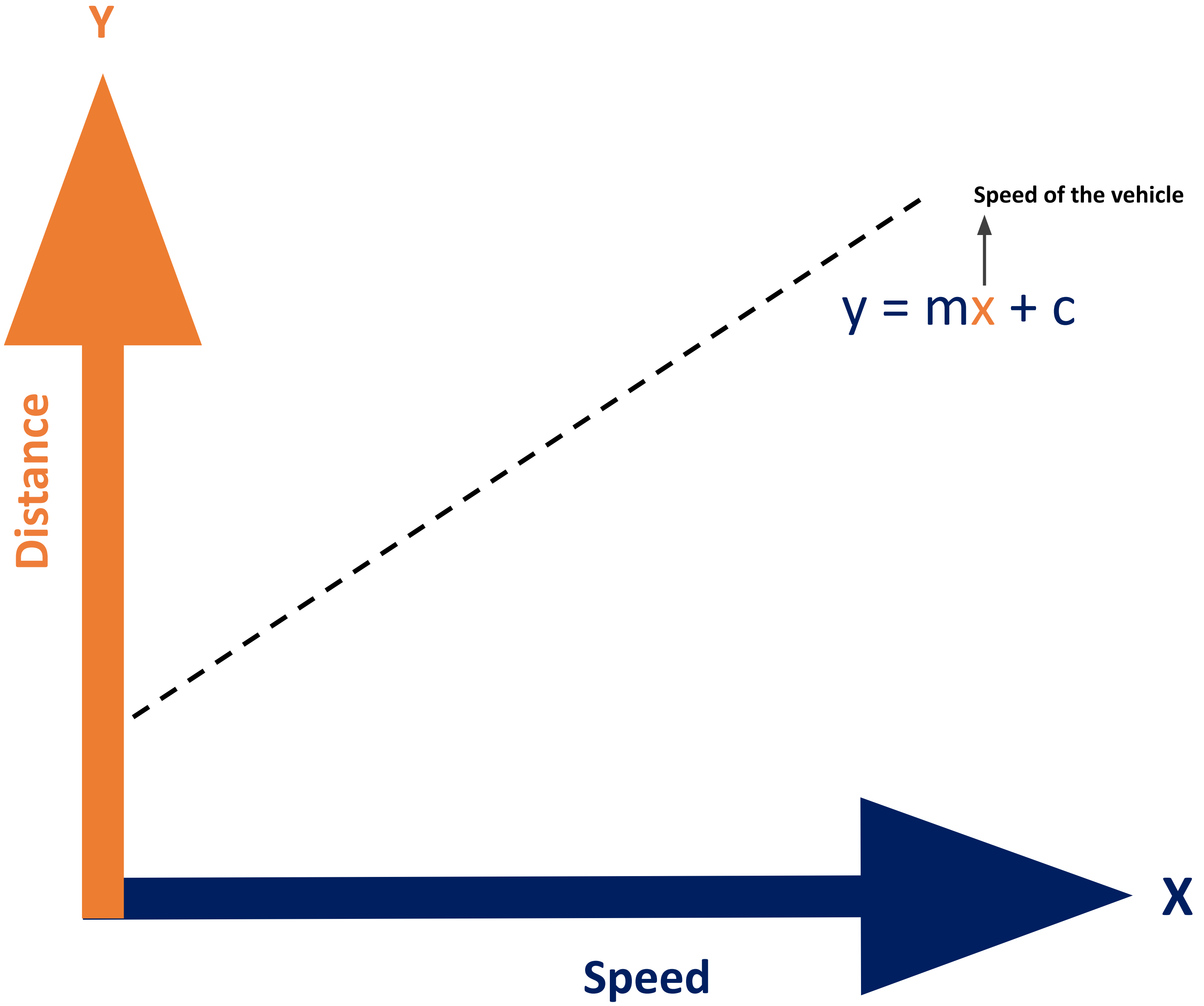
Understanding Linear Regression Algorithm



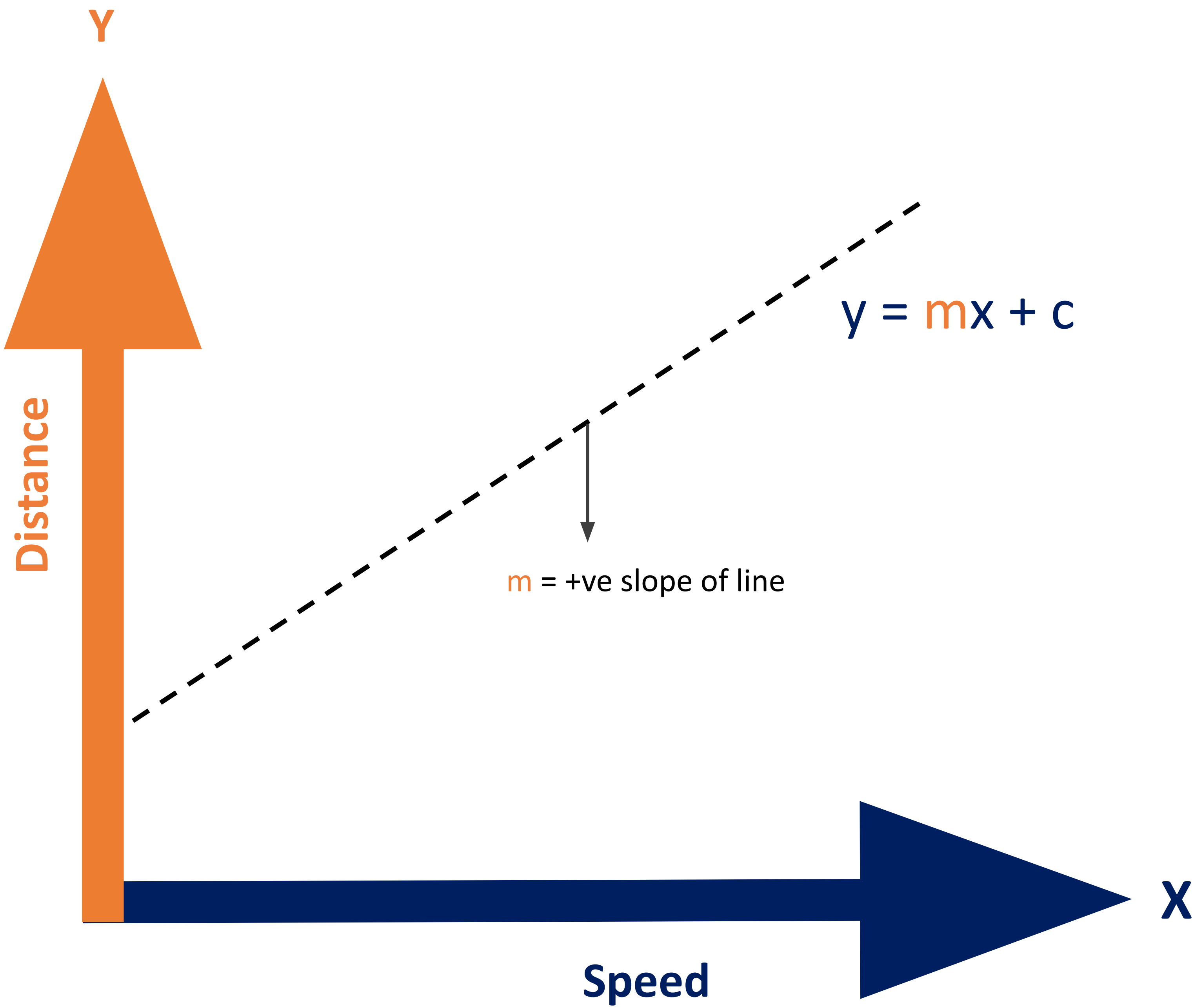
Understanding Linear Regression Algorithm



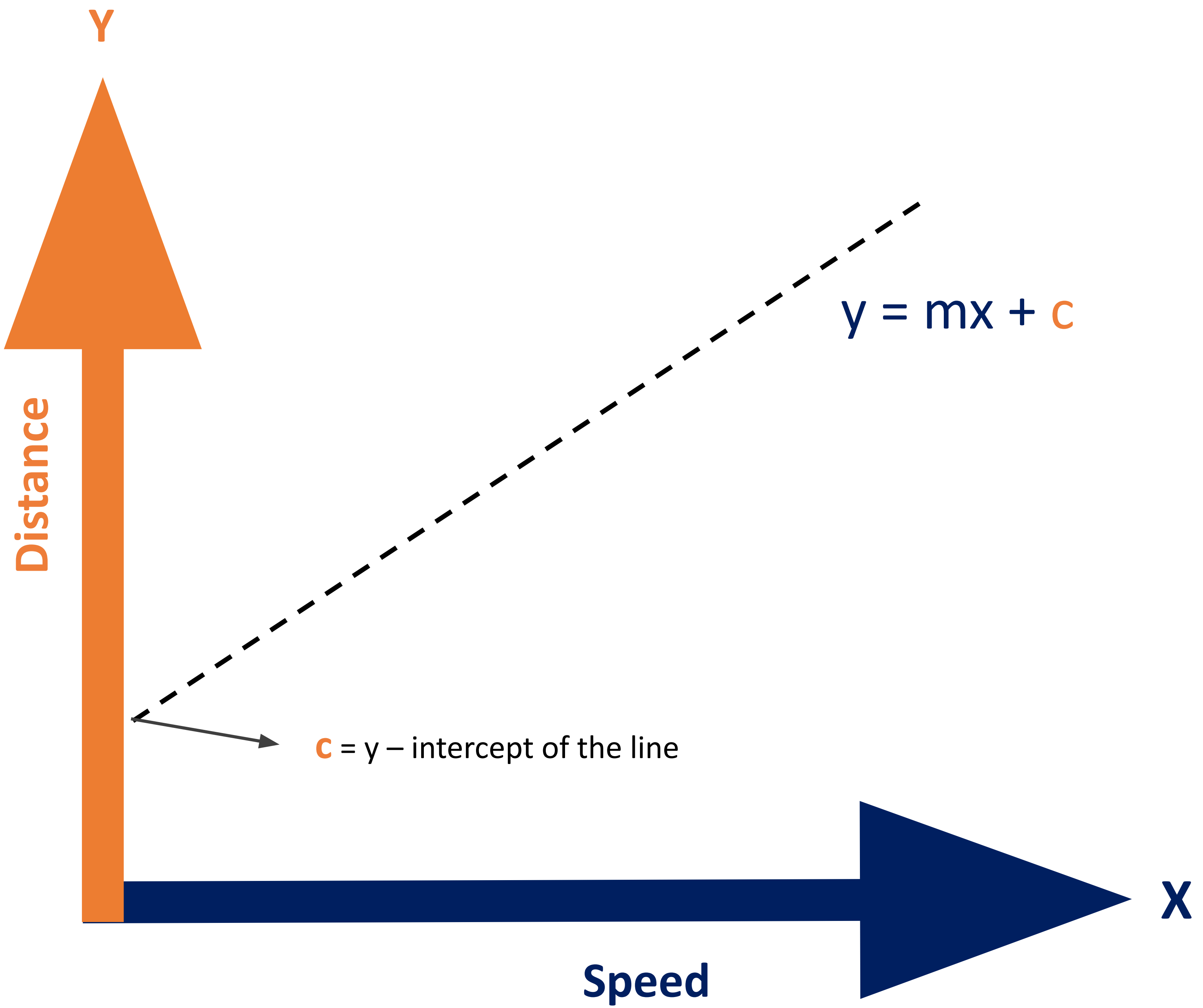
Understanding Linear Regression Algorithm



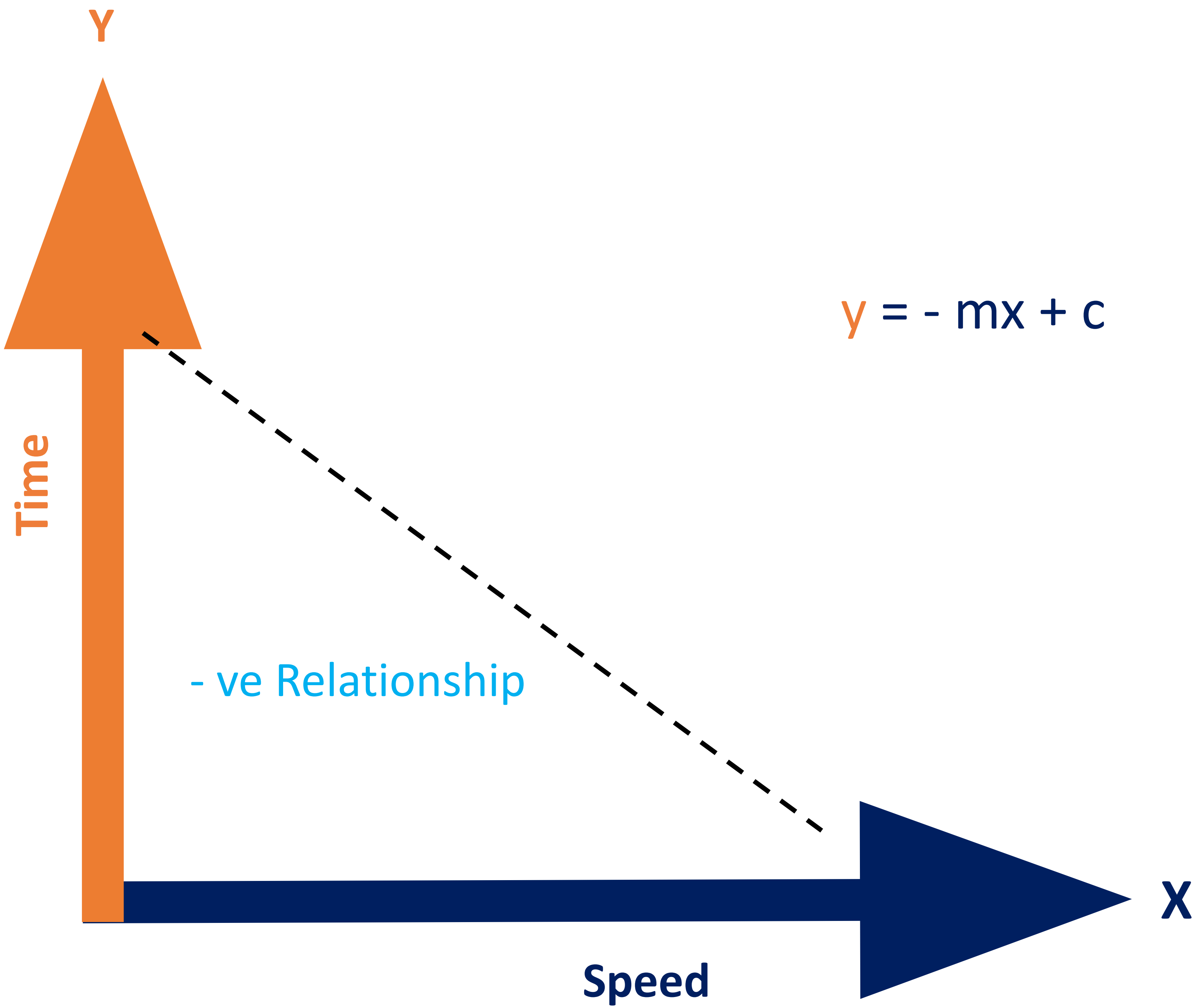
Understanding Linear Regression Algorithm



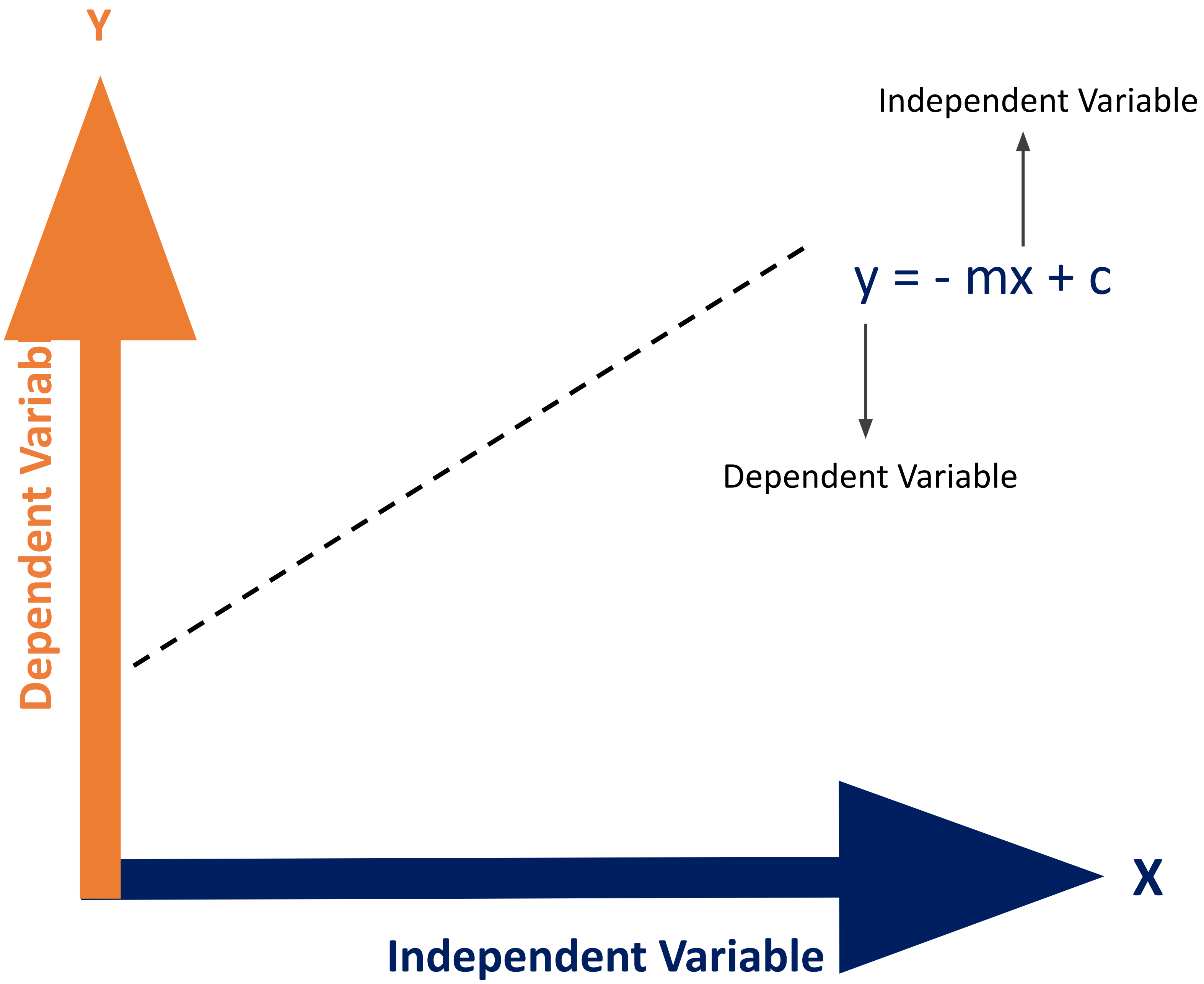
Understanding Linear Regression Algorithm



Understanding Linear Regression Algorithm



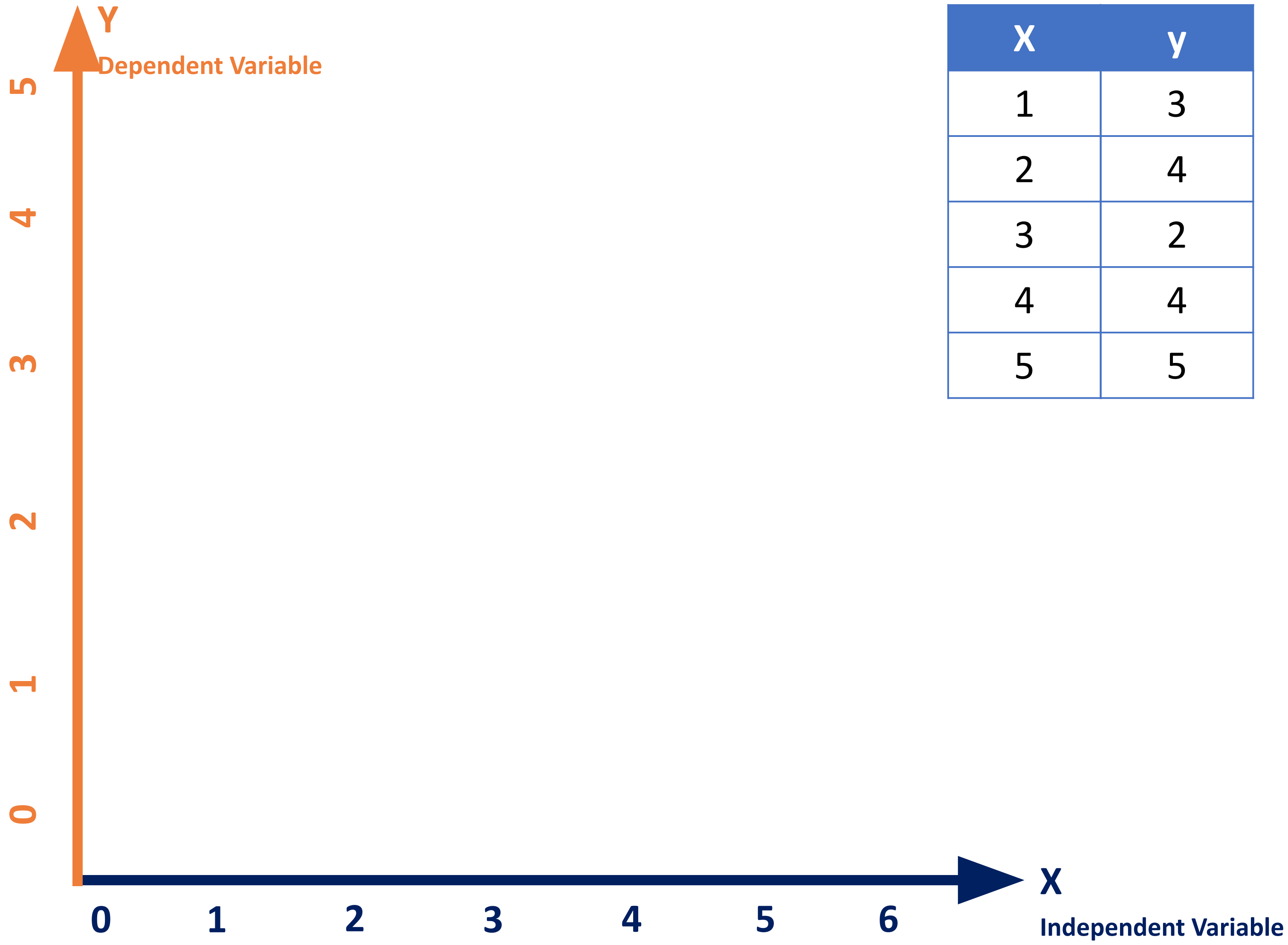
Understanding Linear Regression Algorithm



Understanding Linear Regression Algorithm

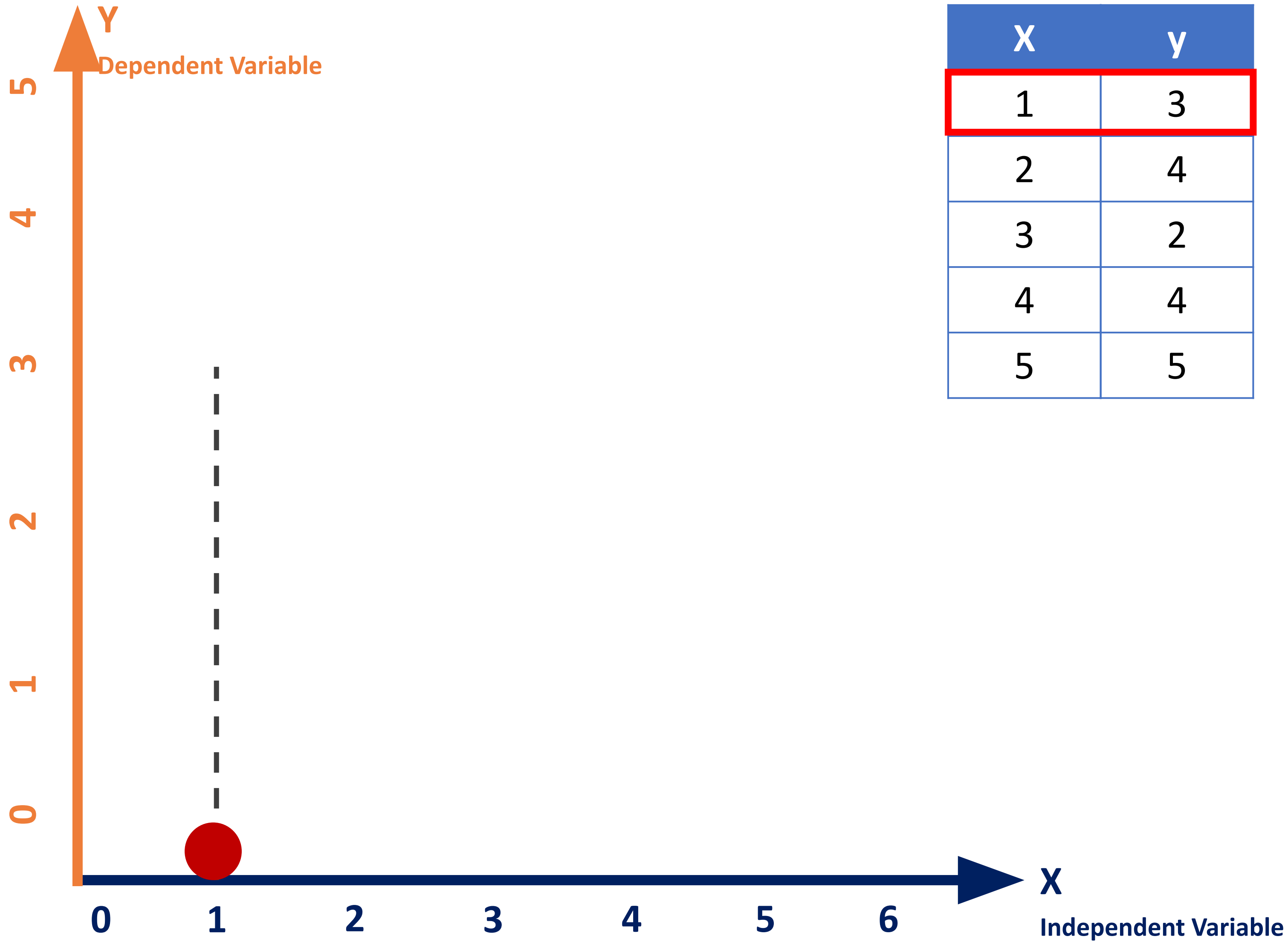
X
1
2
3
4
5

Understanding Linear Regression Algorithm



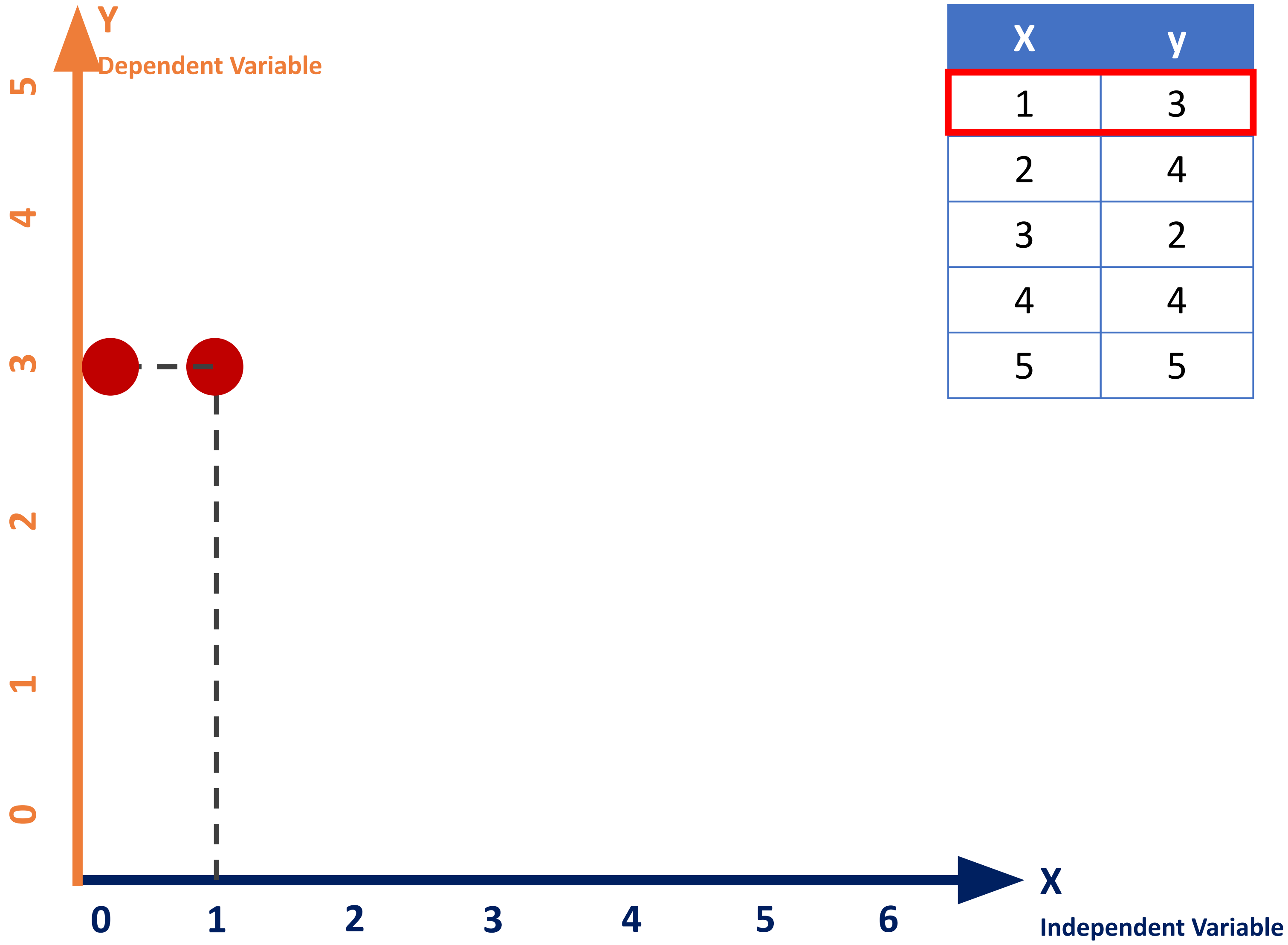
X	y
1	3
2	4
3	2
4	4
5	5

Understanding Linear Regression Algorithm



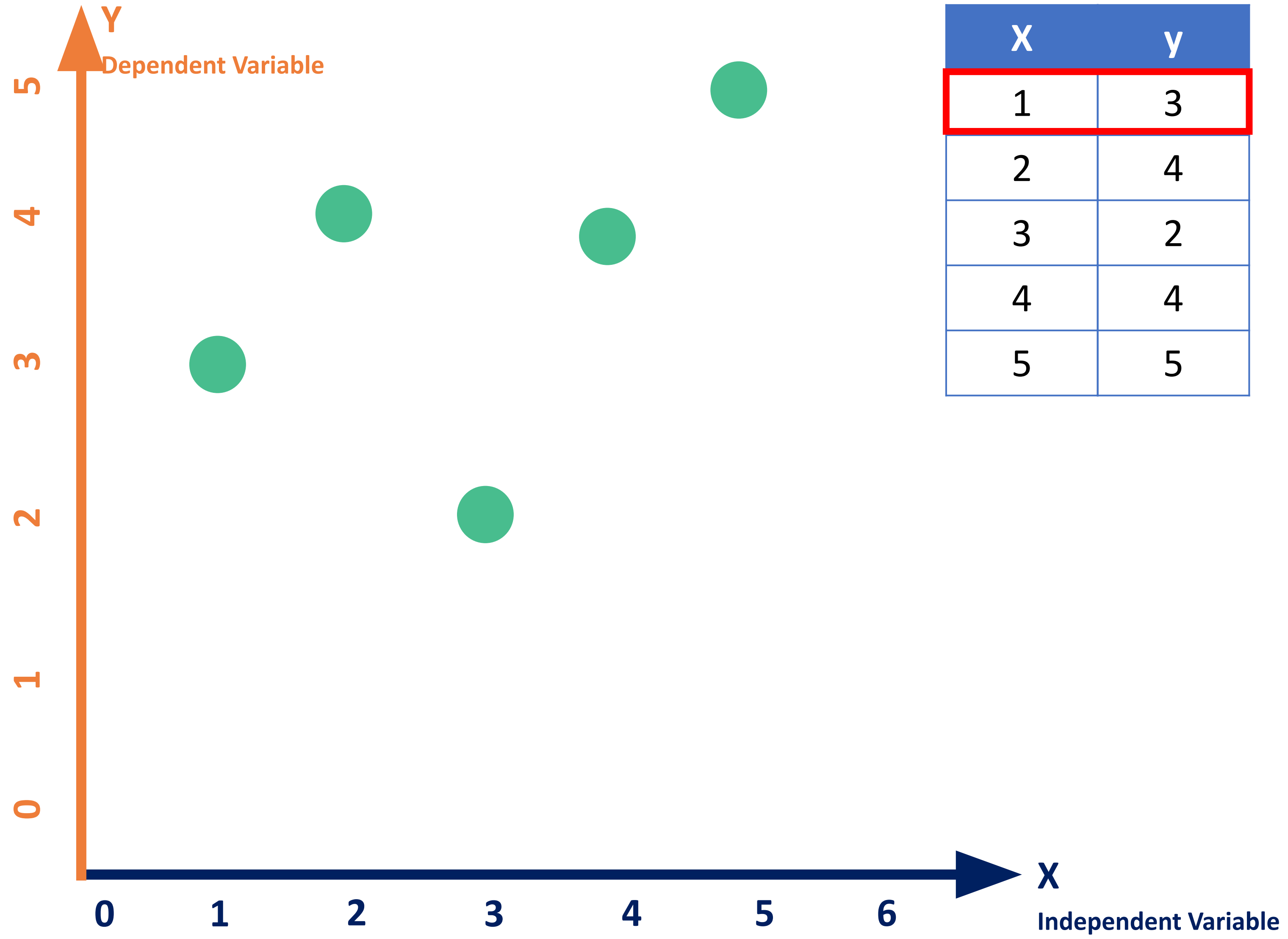
X	y
1	3
2	4
3	2
4	4
5	5

Understanding Linear Regression Algorithm



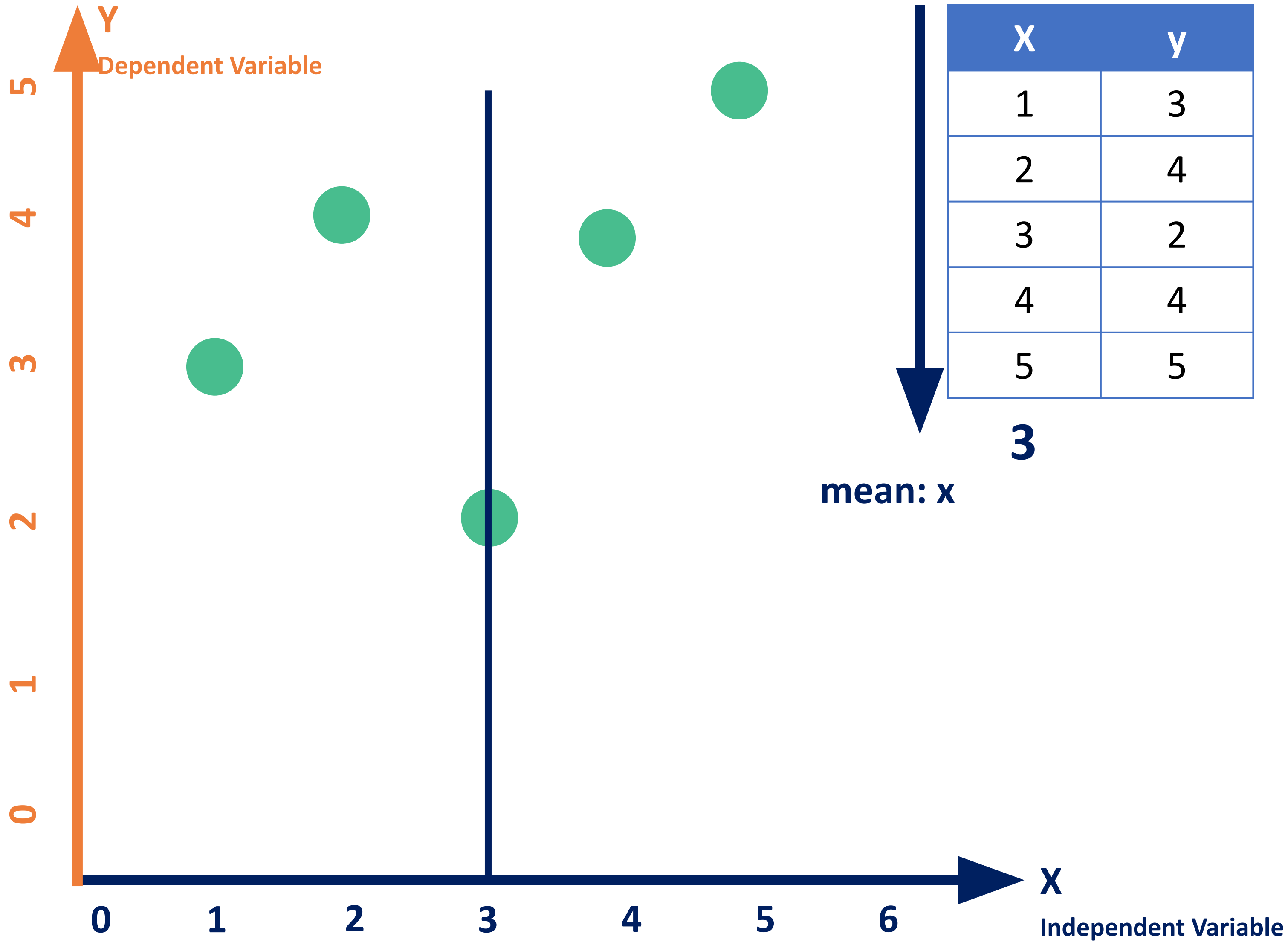
X	y
1	3
2	4
3	2
4	4
5	5

Understanding Linear Regression Algorithm



X	y
1	3
2	4
3	2
4	4
5	5

Understanding Linear Regression Algorithm

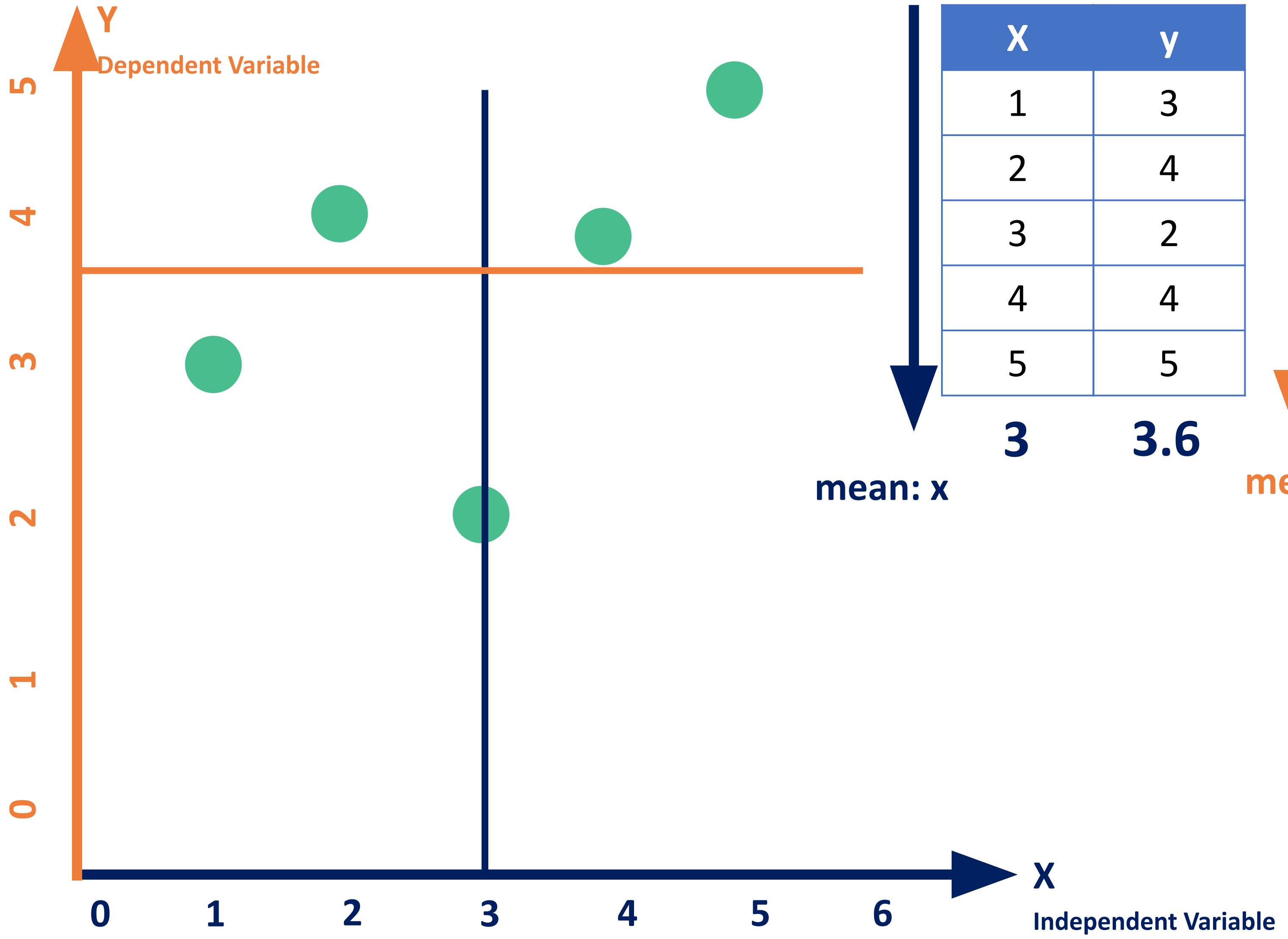


X	y
1	3
2	4
3	2
4	4
5	5

3

mean: x

Understanding Linear Regression Algorithm

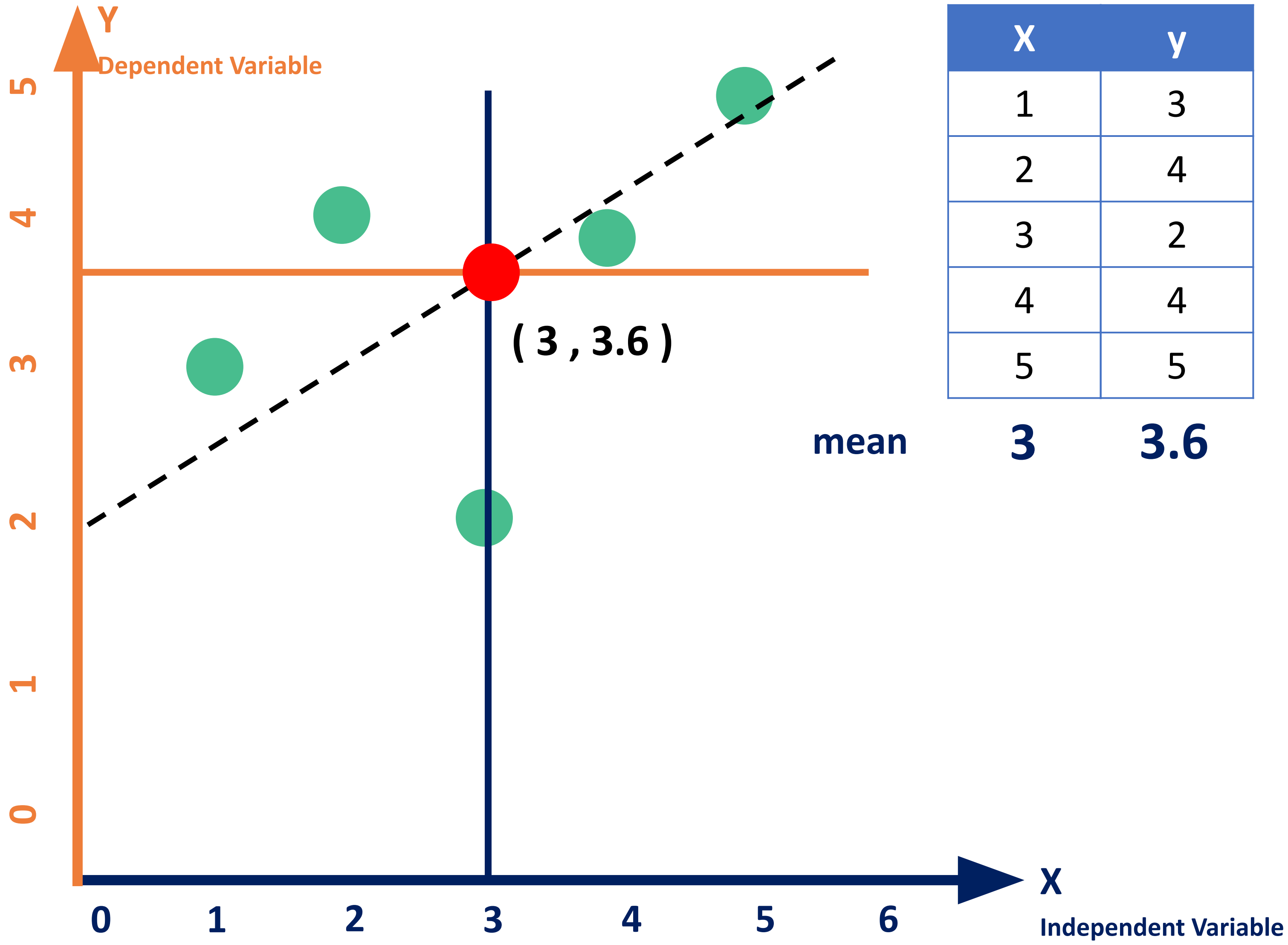


x	y
1	3
2	4
3	2
4	4
5	5
3	**3.6**

mean: x

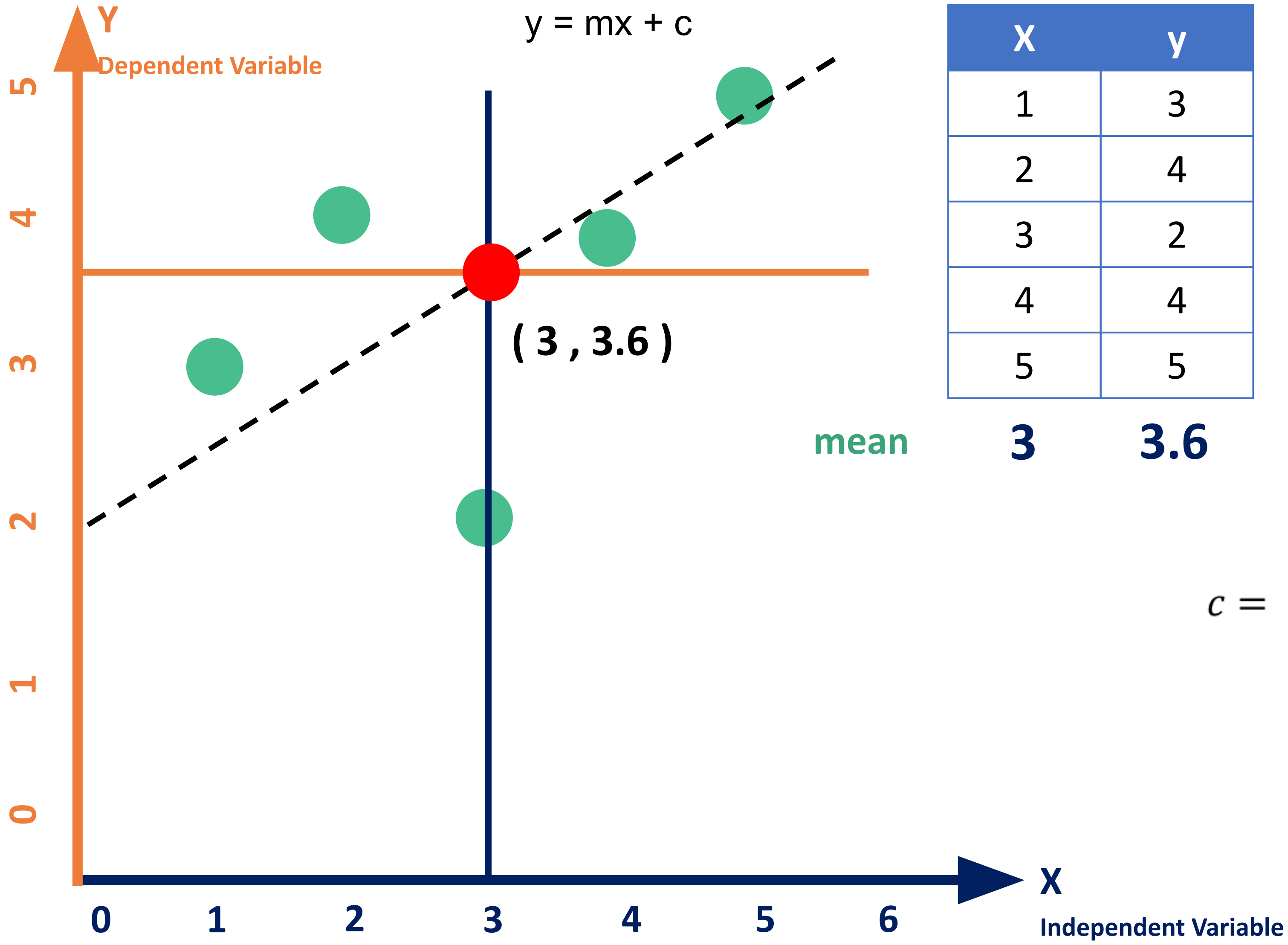
mean: y

Understanding Linear Regression Algorithm



	X	y
	1	3
	2	4
	3	2
	4	4
	5	5
mean	**3**	**3.6**

Understanding Linear Regression Algorithm

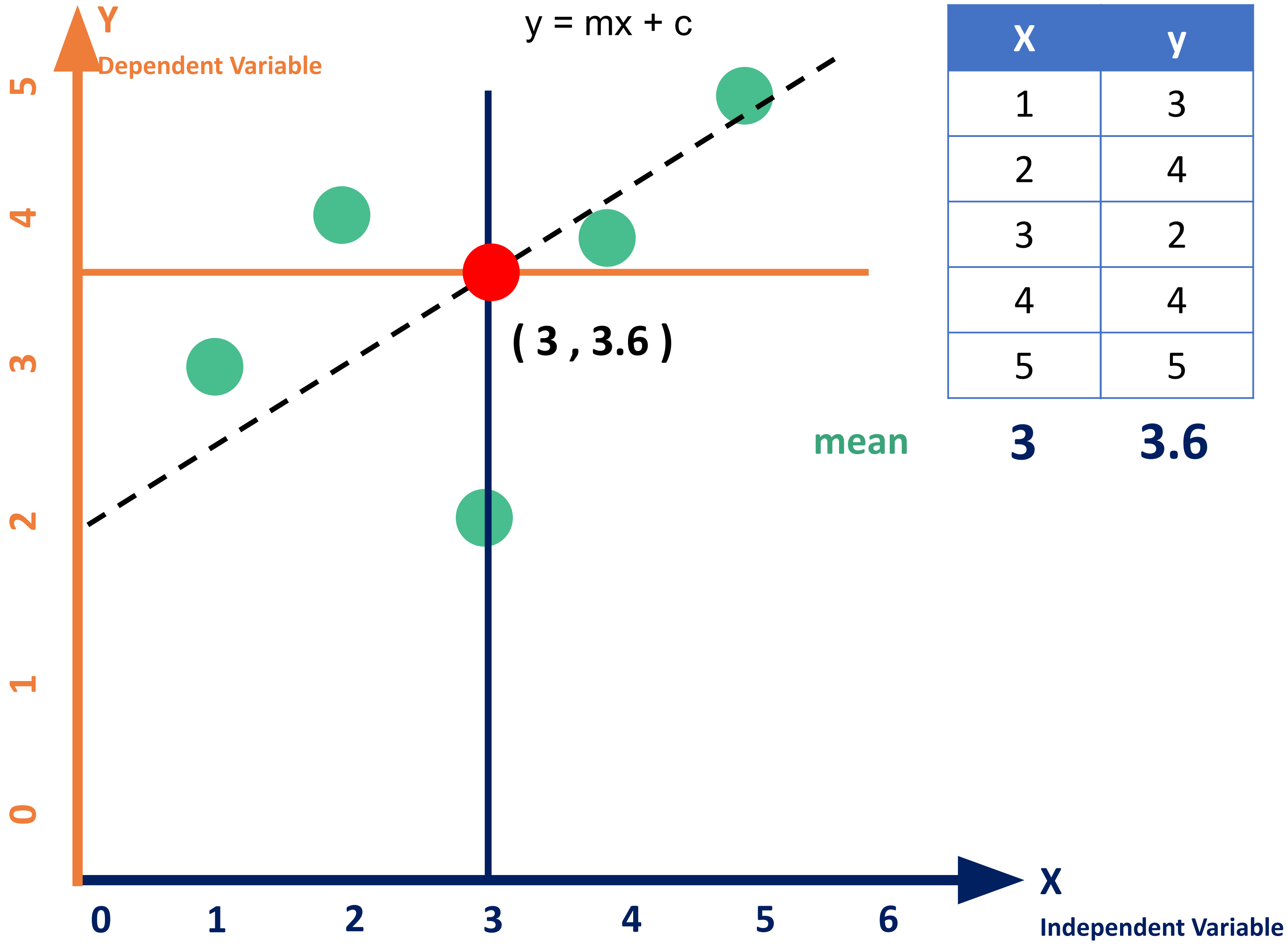


X	y
1	3
2	4
3	2
4	4
5	5
3	**3.6**

mean **3** **3.6**

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Understanding Linear Regression Algorithm



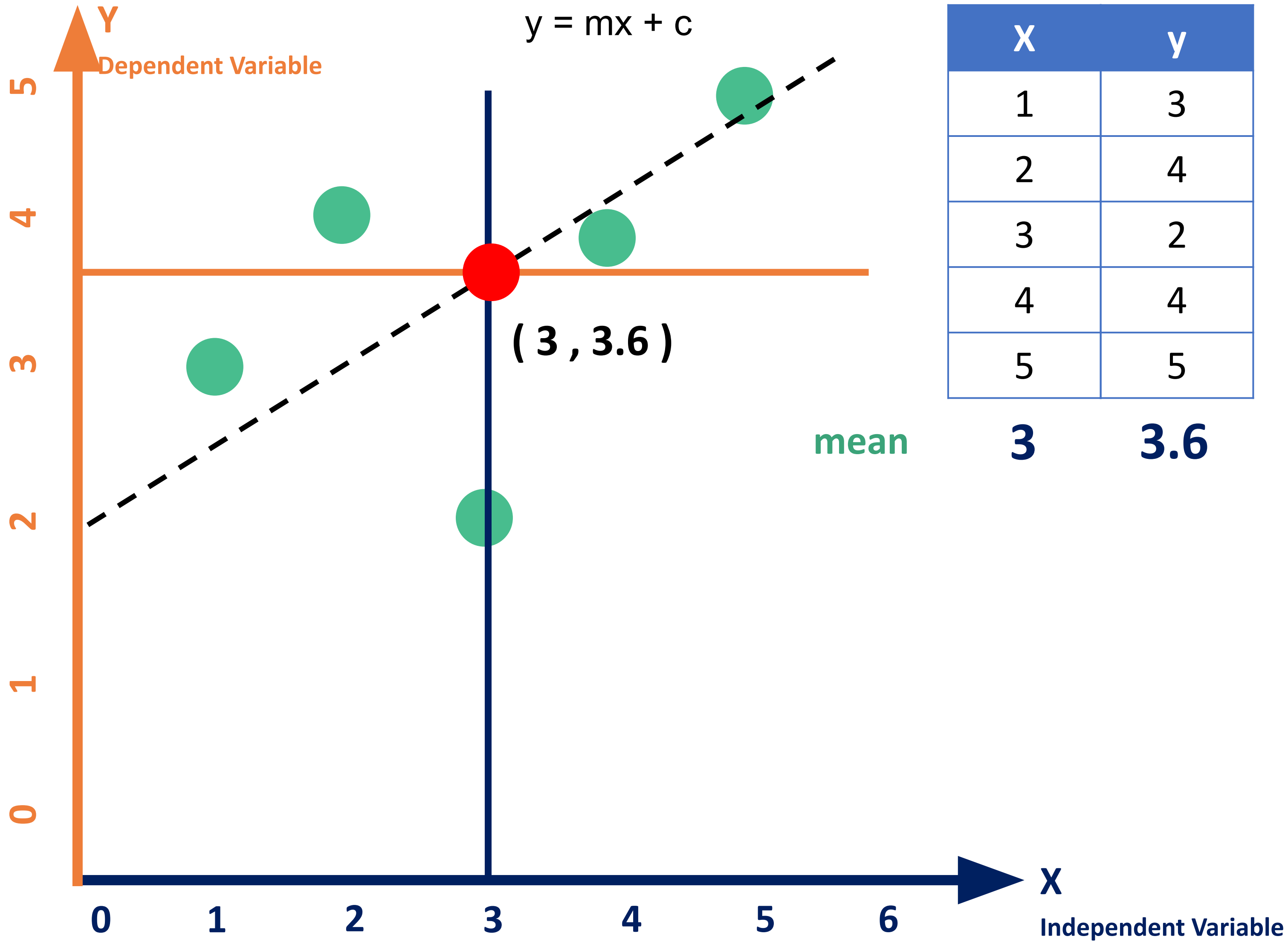
X	y
1	3
2	4
3	2
4	4
5	5
3	**3.6**

mean: **3**, **3.6**

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

(Σy) = 3 + 4 + 2 + 4 + 5 = **18**

Understanding Linear Regression Algorithm



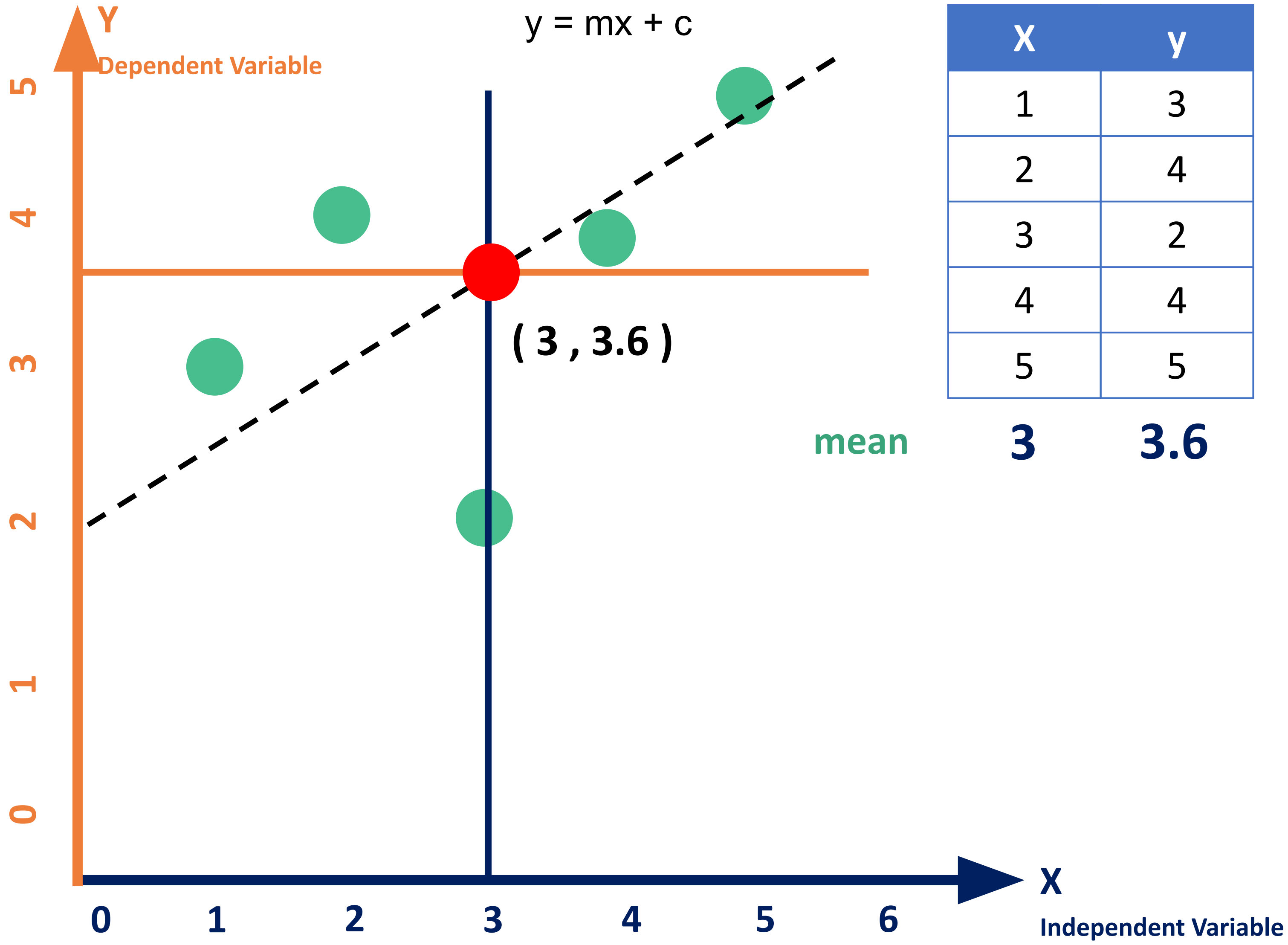
	X	y
	1	3
	2	4
	3	2
	4	4
	5	5
mean	**3**	**3.6**

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

(Σy) = **18**

(Σx) = 1 + 2 + 3 + 4 + 5 = **15**

Understanding Linear Regression Algorithm



X	y
1	3
2	4
3	2
4	4
5	5
mean **3**	**3.6**

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

(Σx) = **15**

(Σy) = **18**

Understanding Linear Regression Algorithm

X	y		
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25

y = mx + c

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

(Σx) = **15**

(Σy) = **18**

$(\boldsymbol{\Sigma xy})$ = 3 + 8 + 6 + 16 + 25 = **58**

Understanding Linear Regression Algorithm

X	y		
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

(Σx) = **15**

(Σy) = **18**

(Σxy) = **58**

(Σx^2) = 1 + 4 + 9 + 16 + 25 = **55**

Understanding Linear Regression Algorithm

X	y		
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

(Σx) = **15**

(Σy) = **18**

(Σxy) = **58**

(Σx^2) = **55**

Understanding Linear Regression Algorithm

X	y		
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$c = \frac{(18)(55) - (15)(58)}{5(55) - (225)}$$

Understanding Linear Regression Algorithm

X	y		
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$c = \frac{(18)(55) - (15)(58)}{5(55) - (225)} = \frac{990 - 870}{275 - (225)}$$

$$c = \frac{120}{50} = 12 / 5 = \mathbf{2.4}$$

y = mx + c

$c = 2.4$

Understanding Linear Regression Algorithm

X	y		
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25

y = mx + c

$$m = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$m = \frac{5(58) - (15)(18)}{5(55) - (15)^2}$$

$$m = \frac{290 - 270}{275 - 225}$$

$$m = \frac{20}{70} = \frac{2}{7} = 0.28$$

Understanding Linear Regression Algorithm

X	y		
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25
6	4	24	36
7	12	84	49
8	15	120	64
9	18	162	81
10	20	200	100

$$y = mx + c \,;$$

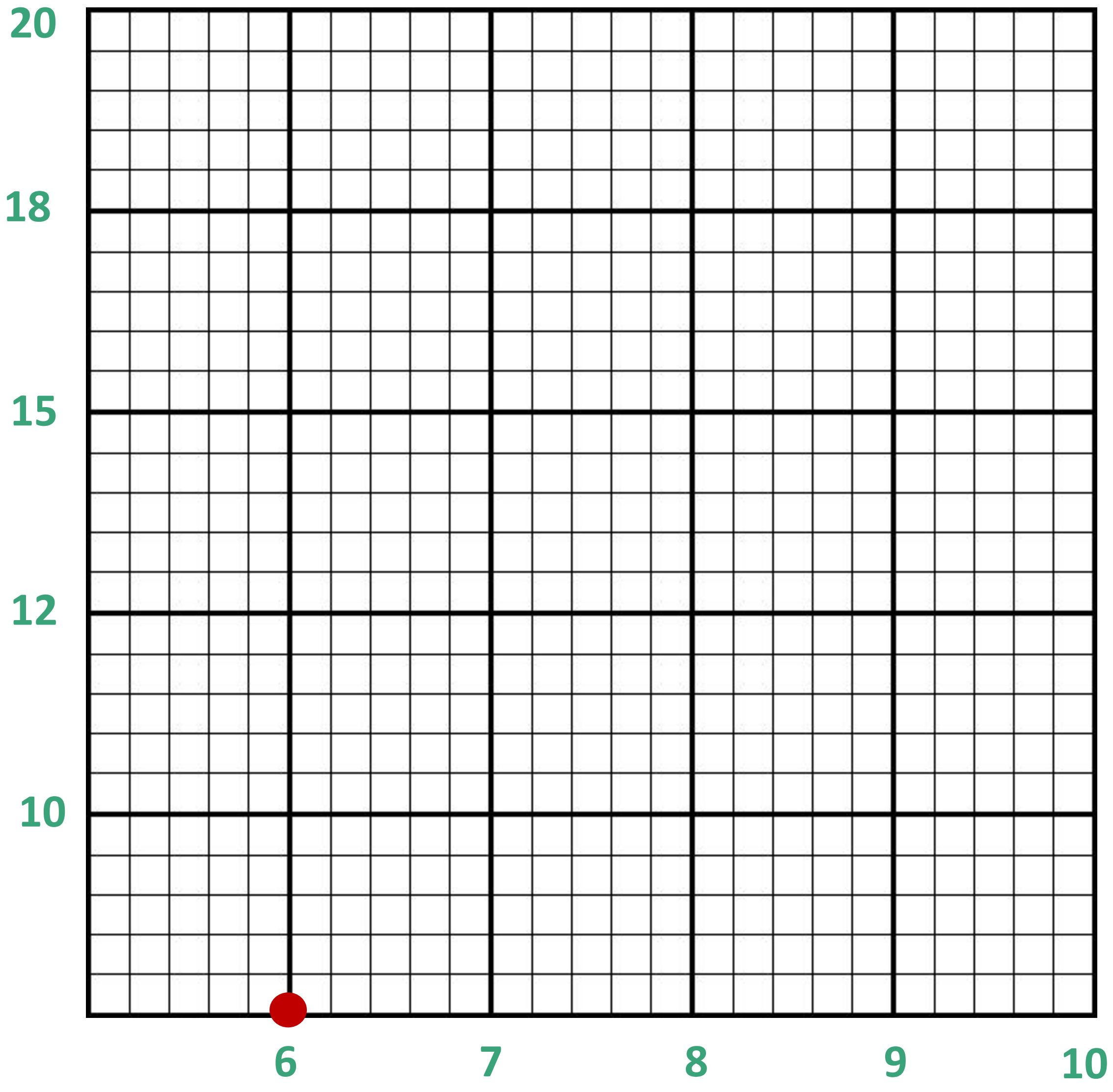
$$m = 0.28 \,;\; c = 2.4$$

$$y = 0.28x + 2.4$$

Let's calculate y for

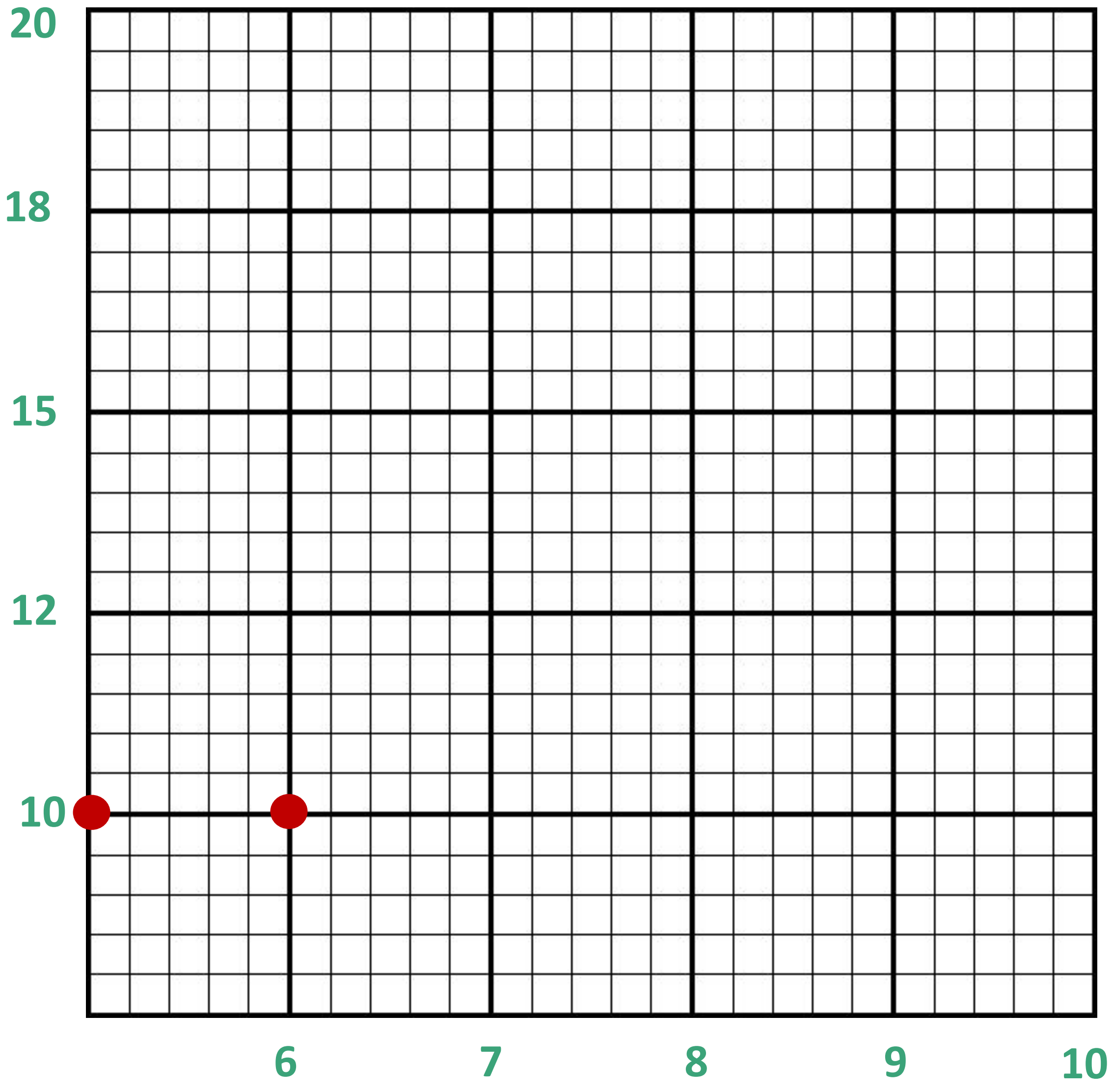
X = {6,7,8,9,10}

Understanding Linear Regression Algorithm



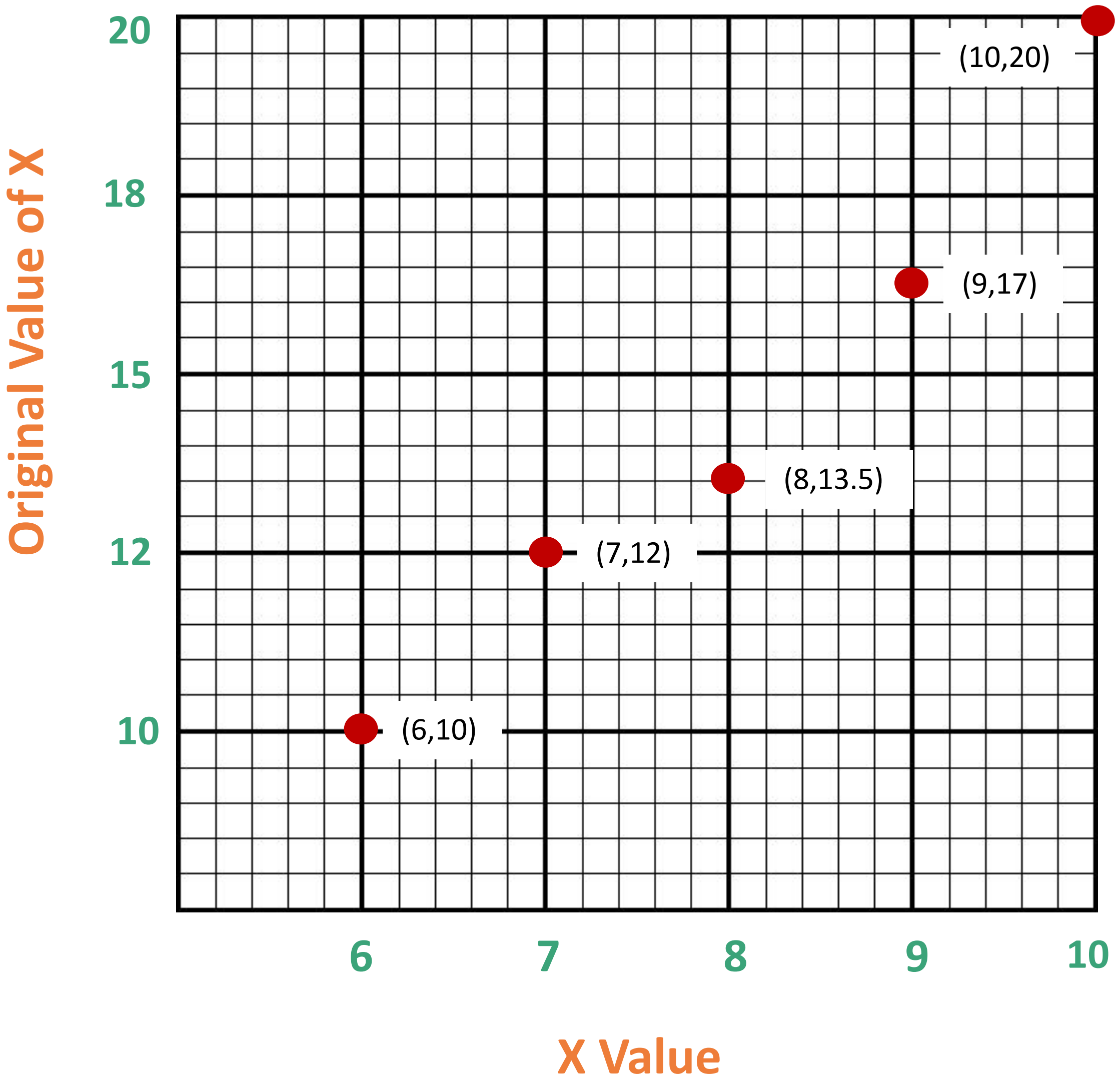
X	y		
6	10	24	36
7	12	84	49
8	15	120	64
9	18	162	81
10	20	200	100

Understanding Linear Regression Algorithm



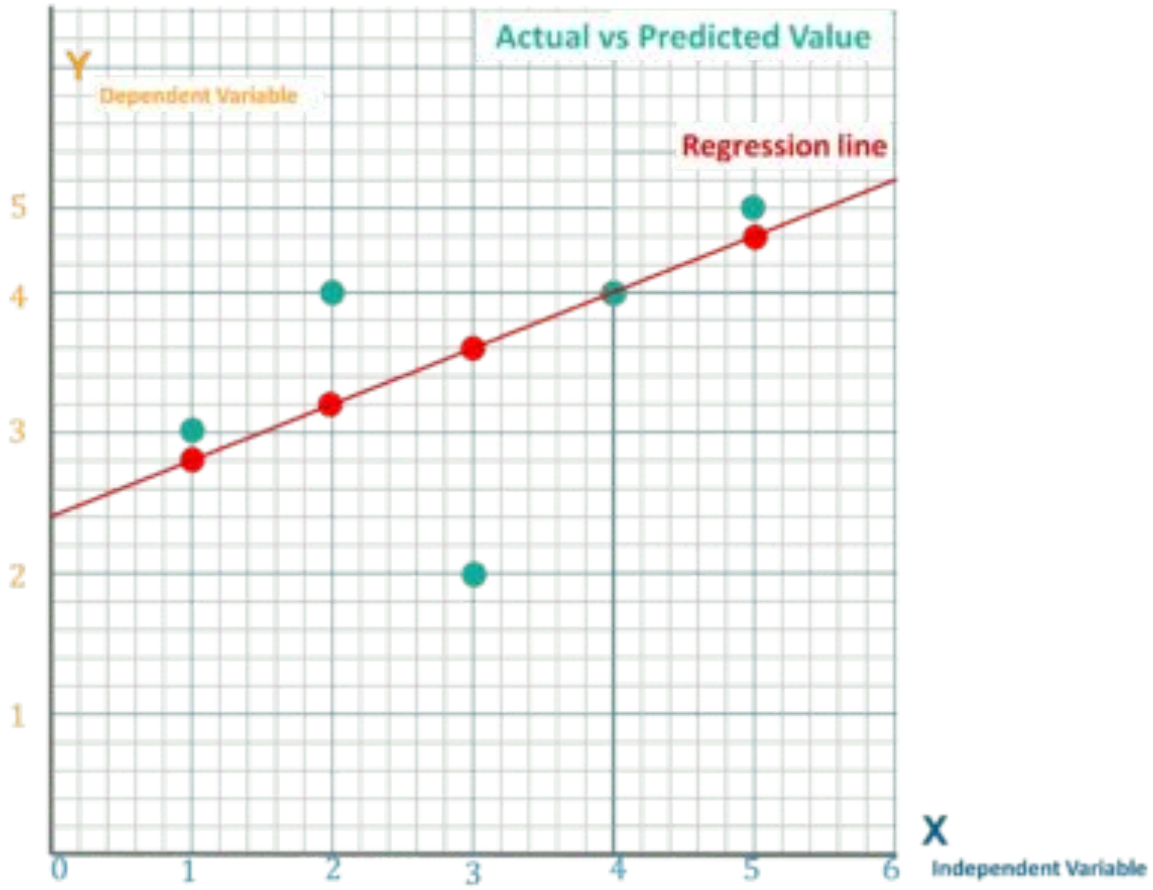
X	y		
6	10	24	36
7	12	84	49
8	15	120	64
9	18	162	81
10	20	200	100

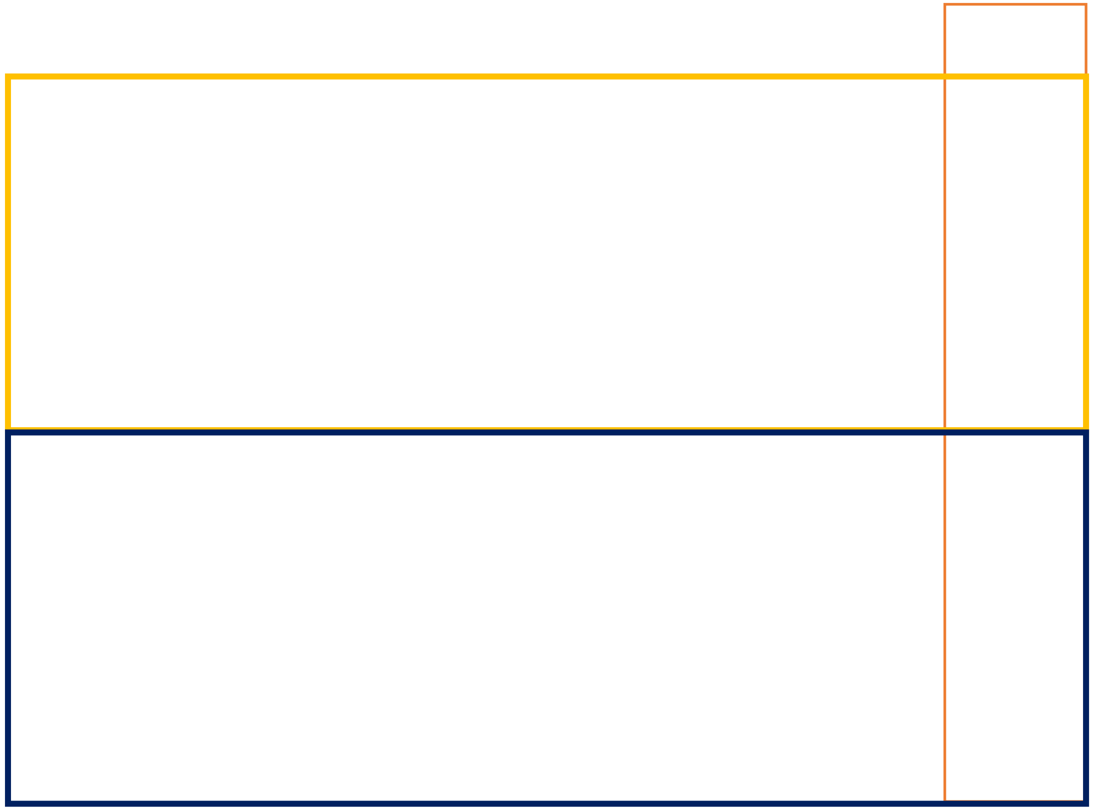
Understanding Linear Regression Algorithm

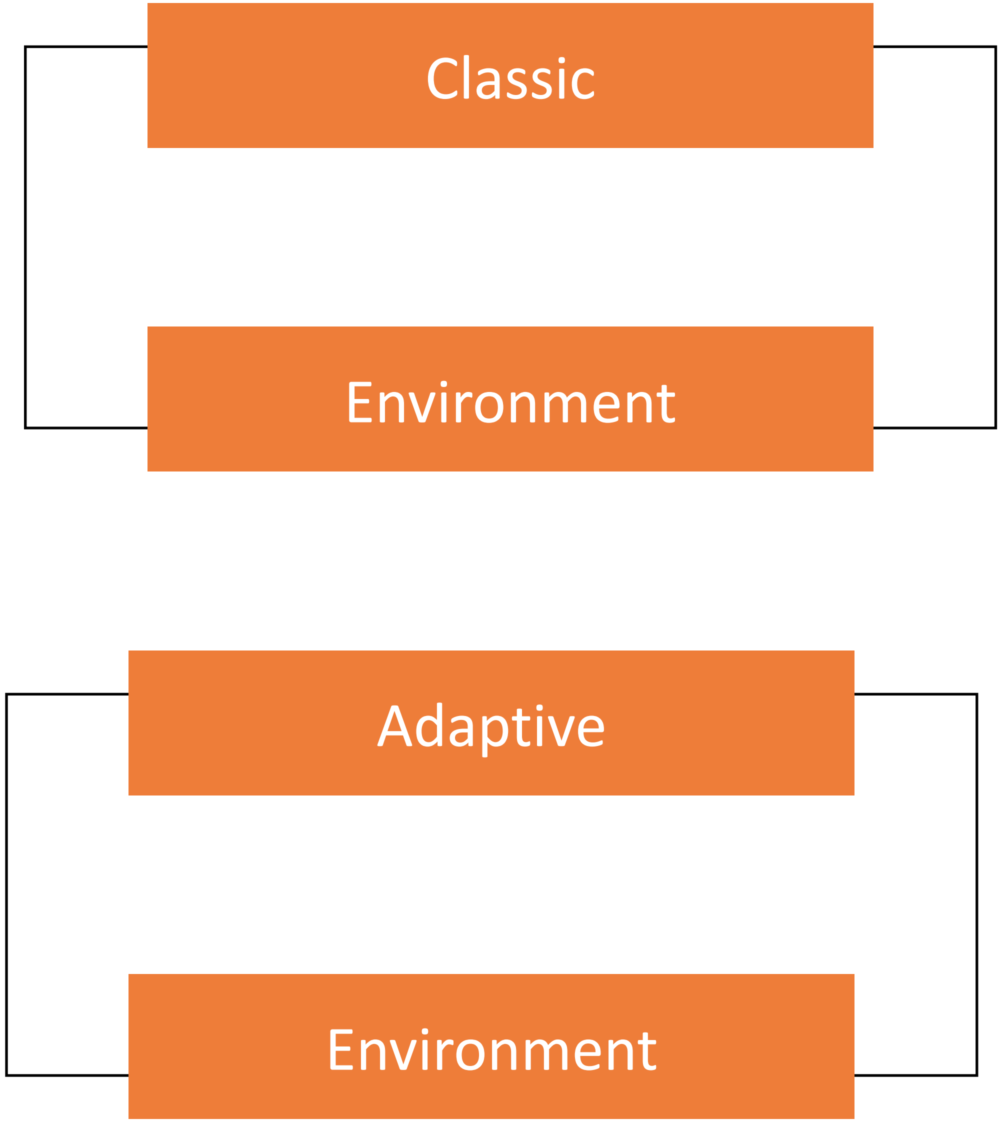


X	y		
6	10	24	36
7	12	84	49
8	13.5	120	64
9	17	162	81
10	20	200	100

Dictate Understanding Linear Regression Algorithm






Classic
Environment
Adaptive
Environment

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPay	TotalPayBenefits	Year	Notes	Agency	Status
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.00	400184.25	NaN	567595.43	567595.43	2011	NaN	San Francisco	NaN
1	2	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN	538909.28	538909.28	2011	NaN	San Francisco	NaN
2	3	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.60	NaN	335279.91	335279.91	2011	NaN	San Francisco	NaN
3	4	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC	77916.00	56120.71	198306.90	NaN	332343.61	332343.61	2011	NaN	San Francisco	NaN
4	5	PATRICK GARDNER	DEPUTY CHIEF OF DEPARTMENT,(FIRE DEPARTMENT)	134401.60	9737.00	182234.59	NaN	326373.19	326373.19	2011	NaN	San Francisco	NaN
5	6	DAVID SULLIVAN	ASSISTANT DEPUTY CHIEF II	118602.00	8601.00	189082.74	NaN	316285.74	316285.74	2011	NaN	San Francisco	NaN
6	7	ALSON LEE	BATTALION CHIEF, (FIRE DEPARTMENT)	92492.01	89062.90	134426.14	NaN	315981.05	315981.05	2011	NaN	San Francisco	NaN
7	8	DAVID KUSHNER	DEPUTY DIRECTOR OF INVESTMENTS	256576.96	0.00	51322.50	NaN	307899.46	307899.46	2011	NaN	San Francisco	NaN
8	9	MICHAEL MORRIS	BATTALION CHIEF, (FIRE DEPARTMENT)	176932.64	86362.68	40132.23	NaN	303427.55	303427.55	2011	NaN	San Francisco	NaN
9	10	JOANNE HAYES-WHITE	CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)	285262.00	0.00	17115.73	NaN	302377.73	302377.73	2011	NaN	San Francisco	NaN
10	11	ARTHUR KENNEY	ASSISTANT CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)	194999.39	71344.88	33149.90	NaN	299494.17	299494.17	2011	NaN	San Francisco	NaN
11	12	PATRICIA JACKSON	CAPTAIN III (POLICE DEPARTMENT)	99722.00	87082.62	110804.30	NaN	297608.92	297608.92	2011	NaN	San Francisco	NaN
12	13	EDWARD HARRINGTON	EXECUTIVE CONTRACT EMPLOYEE	294580.02	0.00	0.00	NaN	294580.02	294580.02	2011	NaN	San Francisco	NaN
13	14	JOHN MARTIN	DEPARTMENT HEAD V	271329.03	0.00	21342.59	NaN	292671.62	292671.62	2011	NaN	San Francisco	NaN
14	15	DAVID FRANKLIN	BATTALION CHIEF, (FIRE DEPARTMENT)	174872.64	74050.30	37424.11	NaN	286347.05	286347.05	2011	NaN	San Francisco	NaN