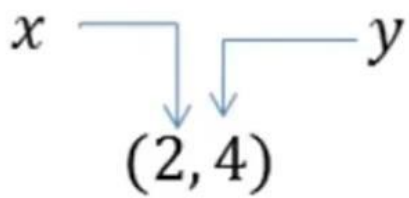
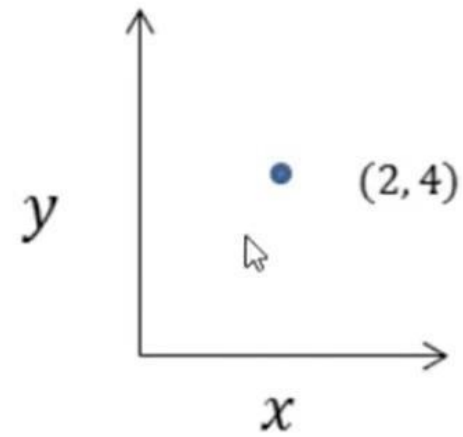
What is a Regression?

Regression: is a statistical approach for modeling the relationship between some variables x (features) and some real valued outcome y (target).

What is a Regression?

Regression: is a statistical approach for modeling the relationship between some variables x (features) and some real valued outcome y (target).





What is a Regression?

Regression: is a statistical approach for modeling the relationship between some variables x (features) and some real valued outcome y (target).

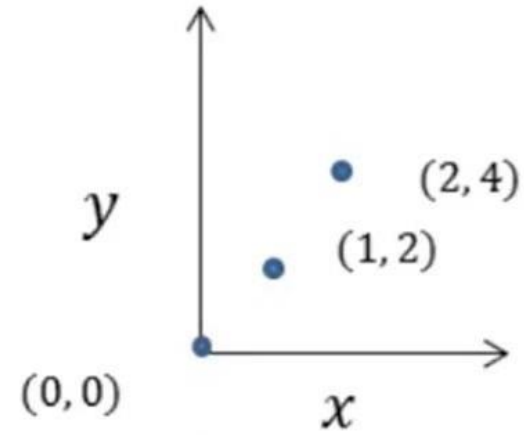
x y

$(2, 4)$

$(1, 2)$

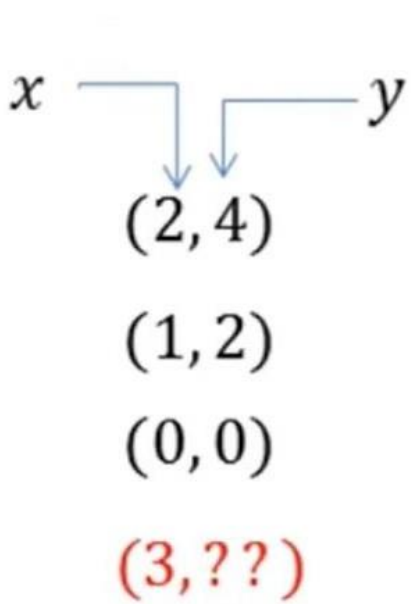
$(0, 0)$

$(3, ??)$



What is a Regression?

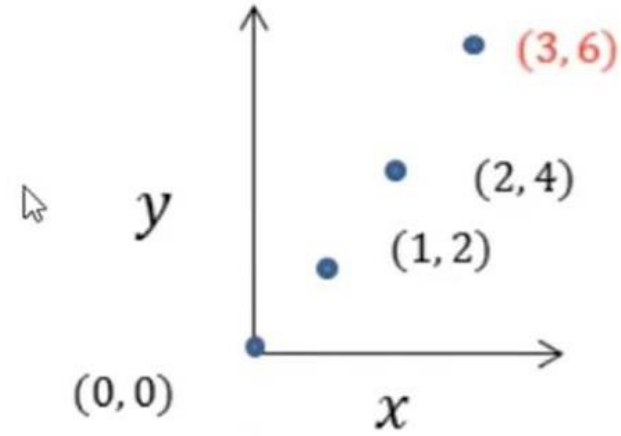
Regression: is a statistical approach for modeling the relationship between some variables x (features) and some real valued outcome y (target).



Model

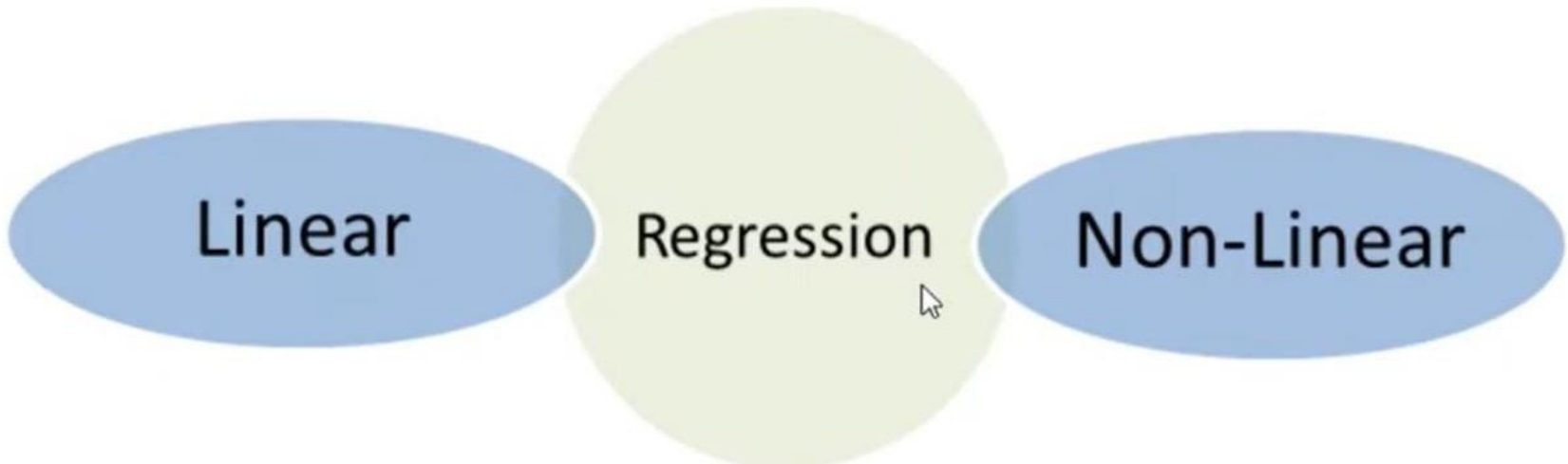
$f(x) = 2x$

$f(x) = 2 * 3$



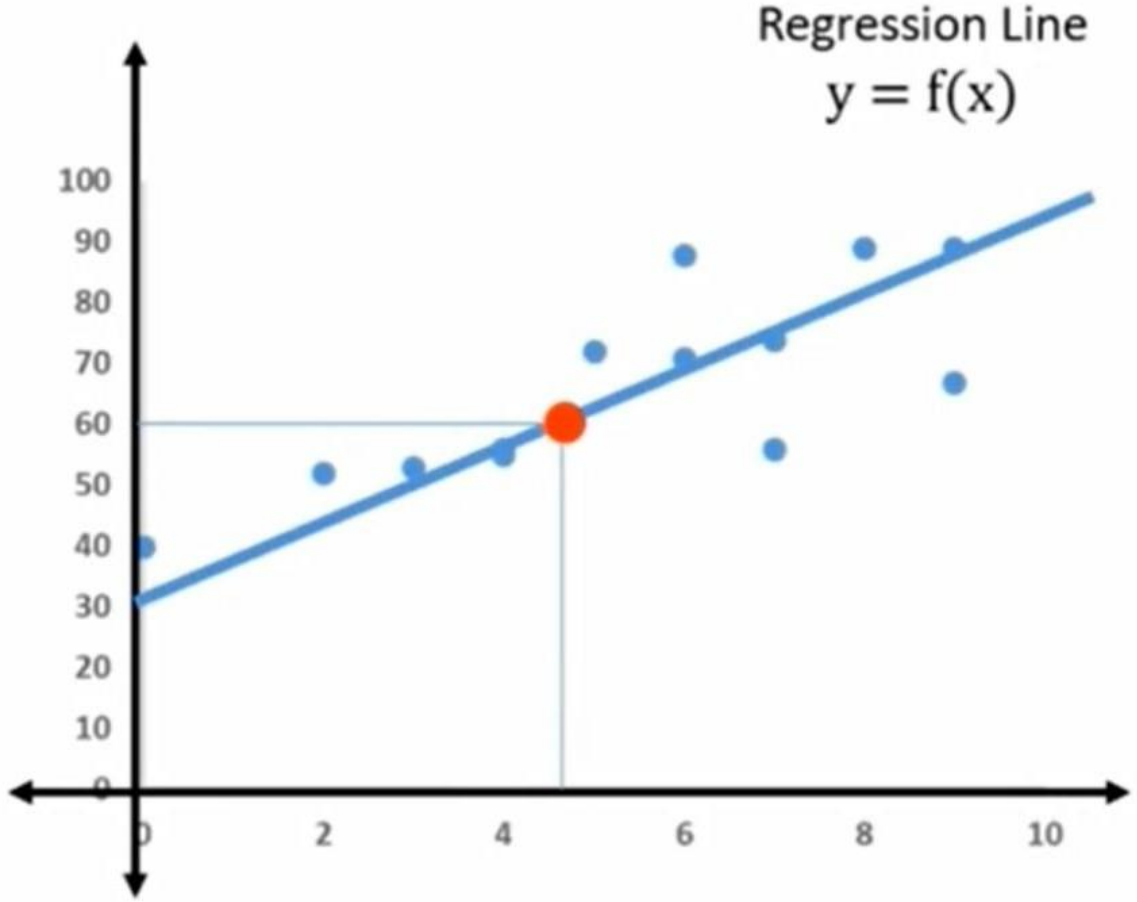
Regression

- Regression problem: the output (y) that you want is continues value (numeric).



Regression Analysis

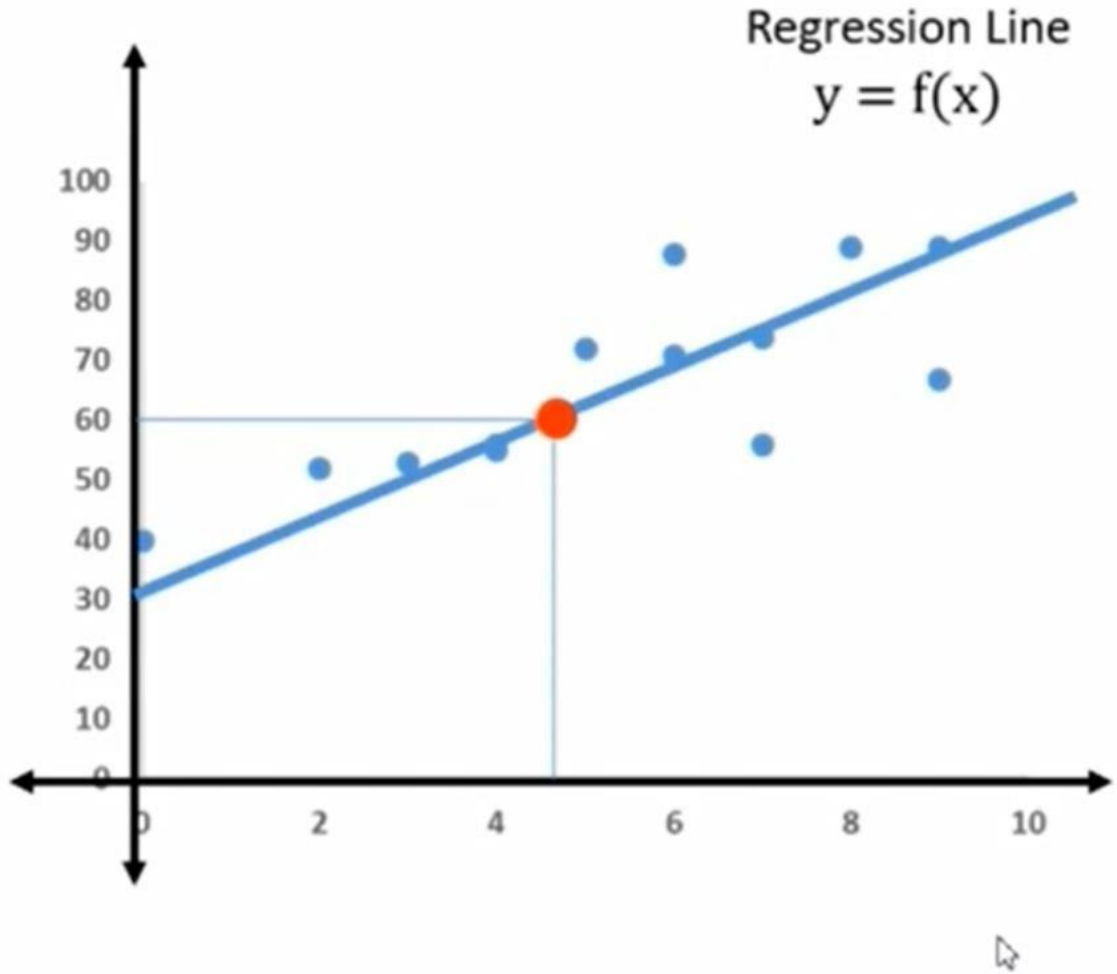
- Statistical process for estimating the relationships among variables



Regression Analysis

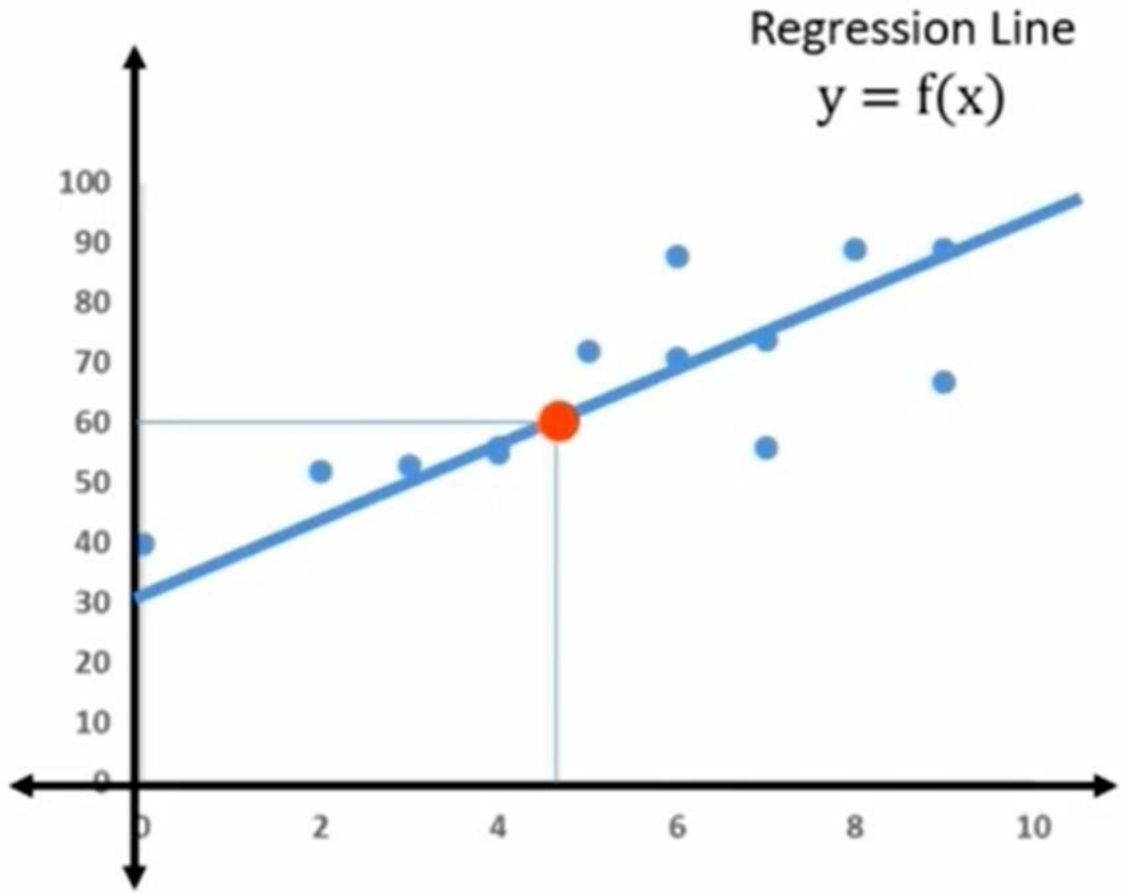
- Statistical process for estimating the relationships among variables

- The predictor is a continuous variable



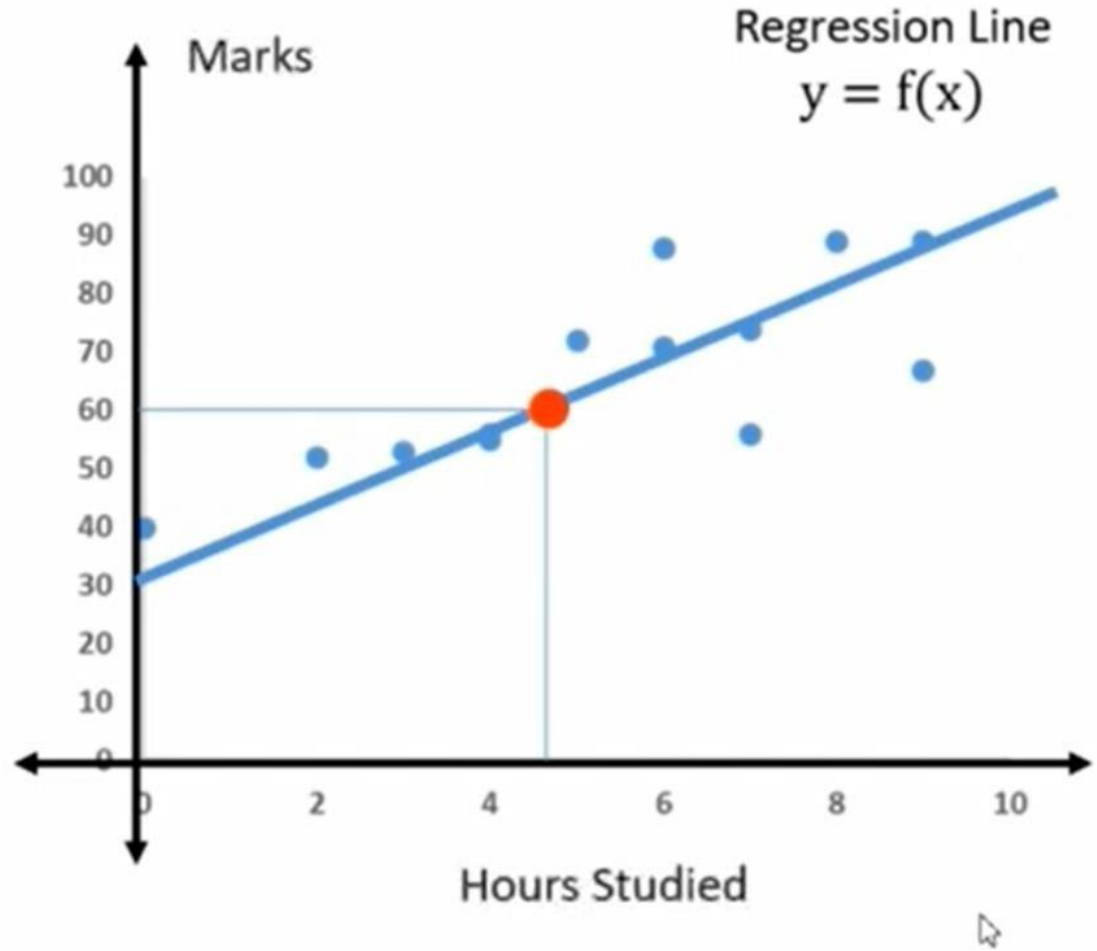
Regression Analysis

- Statistical process for estimating the relationships among variables
- The predictor is a continuous variable
- Relationship between a dependent variable and one or more independent variables (or 'predictors')



Regression Analysis

- Statistical process for estimating the relationships among variables
- The predictor is a continuous variable
- Relationship between a dependent variable and one or more independent variables (or 'predictors')

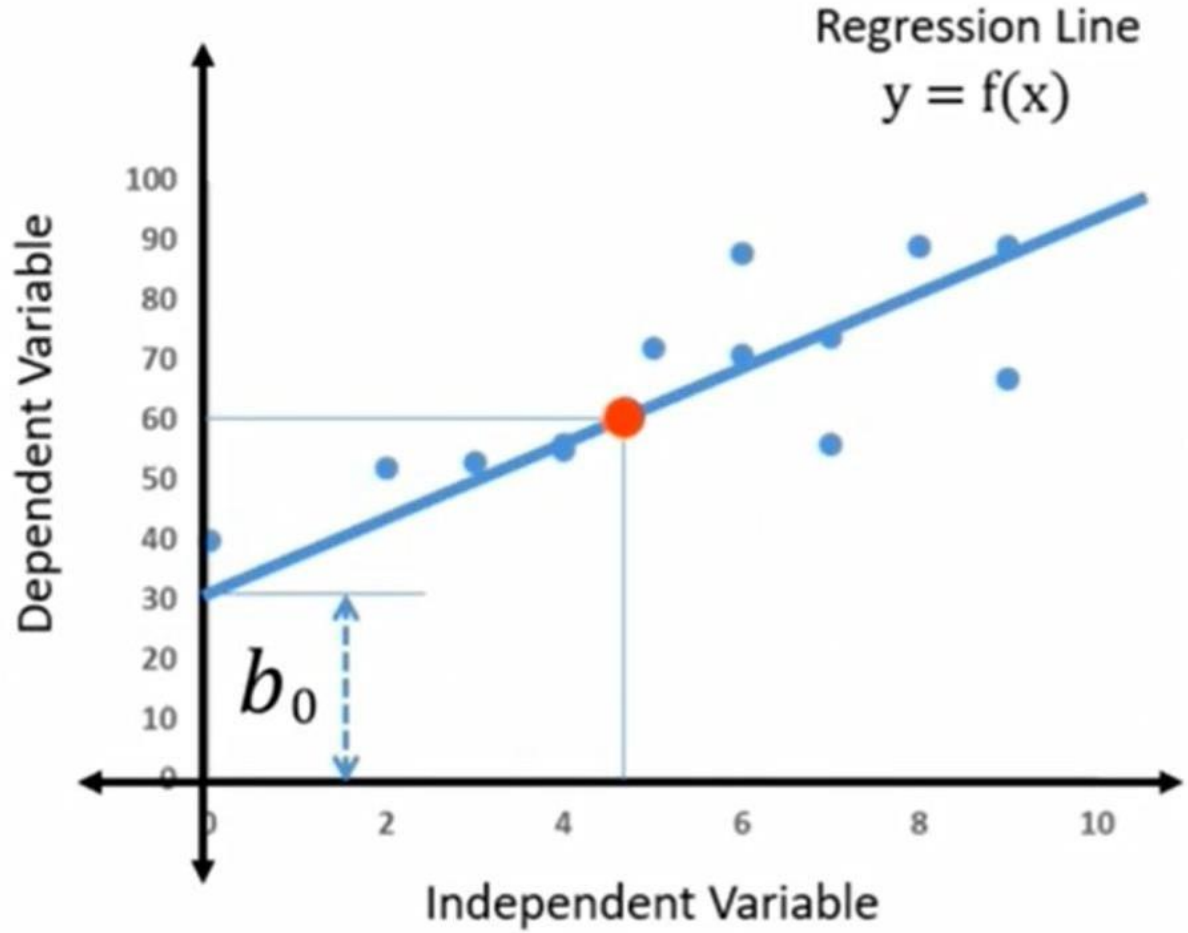


Simple Linear Regression

Simple Regression :

$$y = b_0 + b_1 x$$

Only one Dependent
Only one Independent

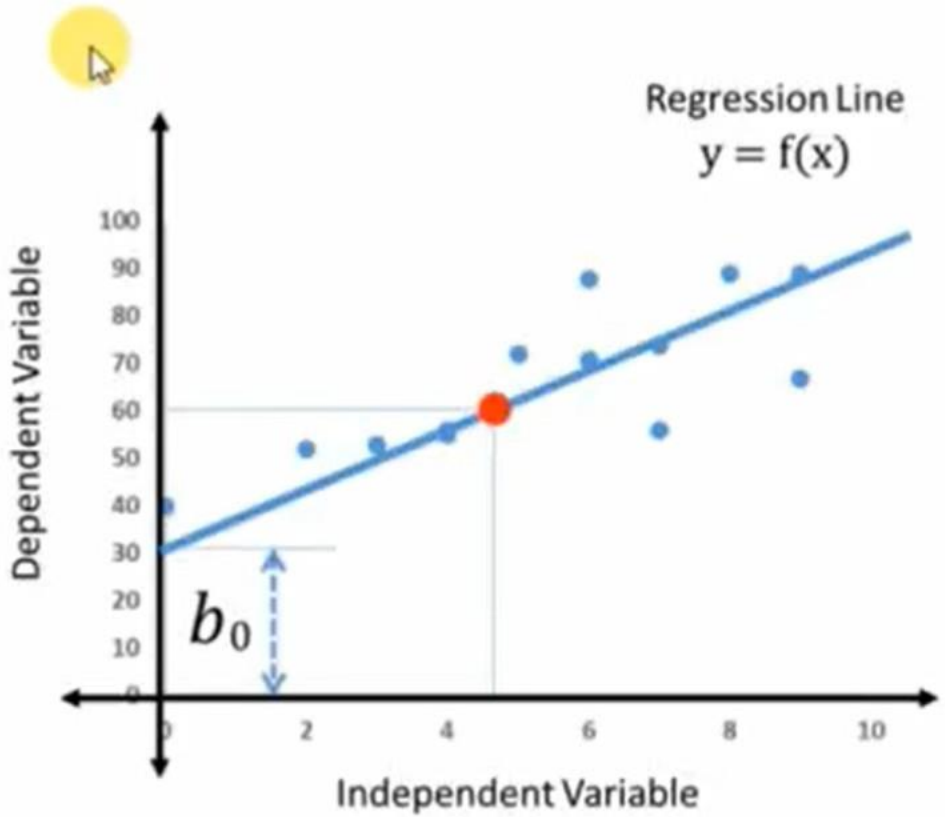


Simple Linear Regression

Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

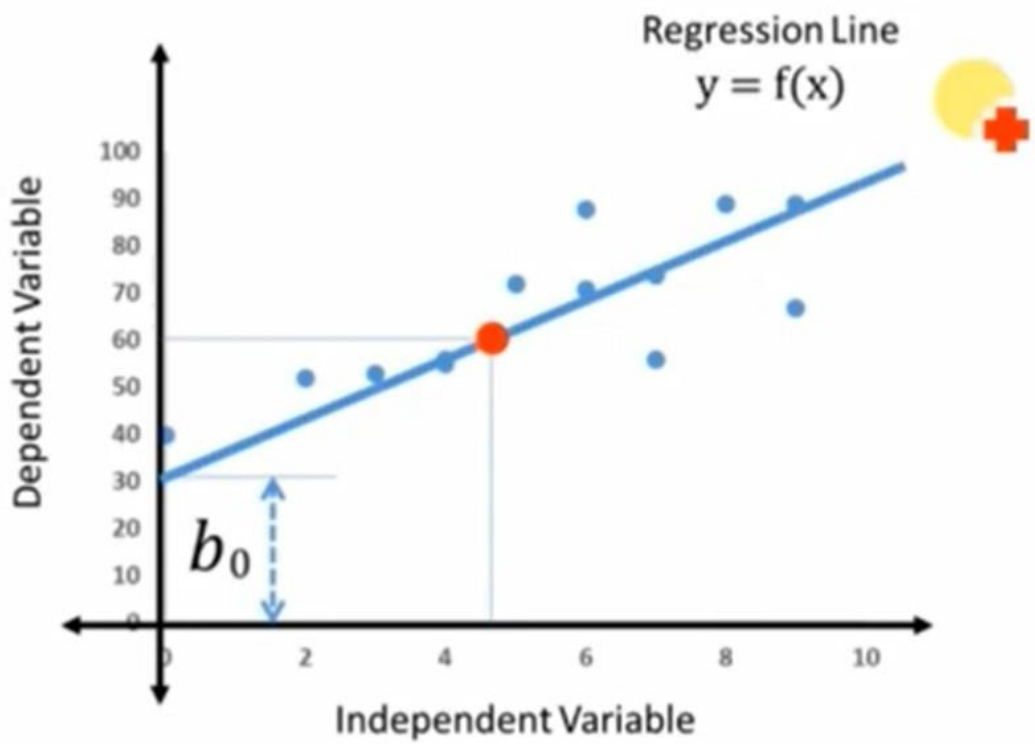
Simple Linear Regression

Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	



Simple Linear Regression

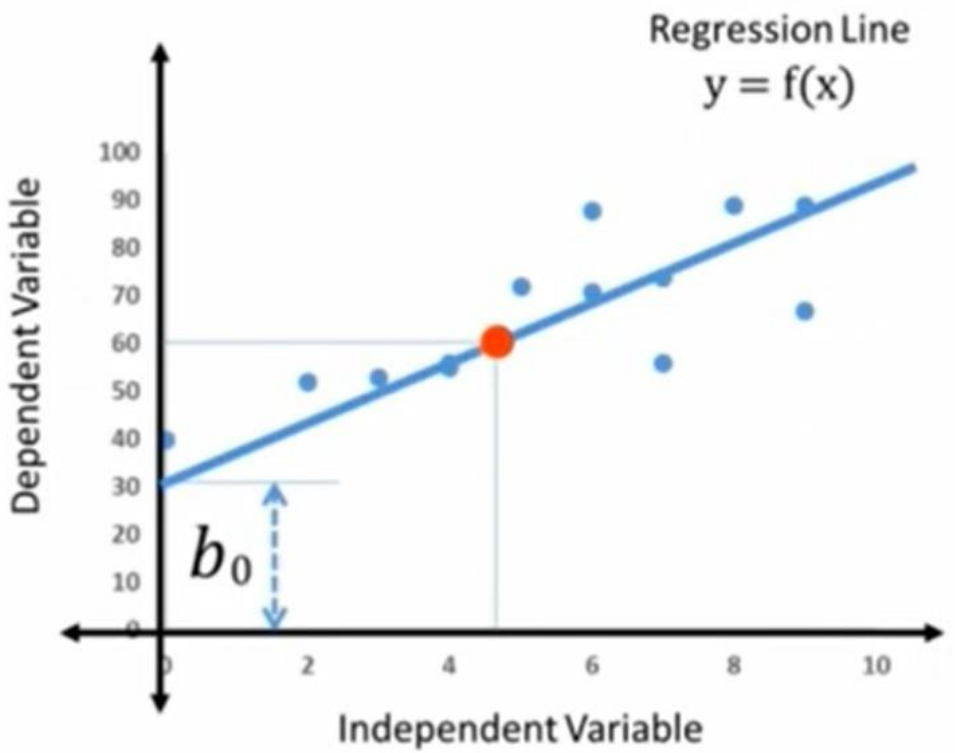
Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	



$$y = b_0 + b_1 x$$

Simple Linear Regression

Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	



$$y = b_0 + b_1 x$$

$$b_1 = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

Simple Linear Regression

Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

X – Mean (A)	Y – Mean (B)
-5.38	

$$y = b_0 + b_1 x$$

$$b_1 = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sum (X - \overline{X})^2}$$

Simple Linear Regression

Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

X – Mean (A)	Y – Mean (B)
-5.38	-26.31
-3.38	-14.31
-2.38	-13.31
-1.38	-11.31
-1.38	-10.31
-0.38	5.69
0.62	4.69
0.62	21.69
1.62	-10.31
1.62	7.69
2.62	22.69
3.62	0.69
3.62	22.69

$$y = b_0 + b_1 x$$

$$b_1 = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

Simple Linear Regression

Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

X – Mean (A)	Y – Mean (B)	A^2	A*B
-5.38	-26.31	28.99	141.66
-3.38	-14.31		
-2.38	-13.31		
-1.38	-11.31		
-1.38	-10.31		
-0.38	5.69		
0.62	4.69		
0.62	21.69		
1.62	-10.31		
1.62	7.69		
2.62	22.69		
3.62	0.69		
3.62	22.69		

$$y = b_0 + b_1 x$$

$$b_1 = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

Simple Linear Regression

Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

X – Mean (A)	Y – Mean (B)	A^2	A*B
-5.38	-26.31	28.99	141.66
-3.38	-14.31	11.46	48.43
-2.38	-13.31	5.69	31.73
-1.38	-11.31	1.92	15.66
-1.38	-10.31	1.92	14.27
-0.38	5.69	0.15	-2.19
0.62	4.69	0.38	2.89
0.62	21.69	0.38	13.35
1.62	-10.31	2.61	-16.65
1.62	7.69	2.61	12.43
2.62	22.69	6.84	59.35
3.62	0.69	13.07	2.50
3.62	22.69	13.07	82.04
		89.08	405.46
		Sum	

$$y = b_0 + b_1 x$$

$$b_1 = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

Simple Linear Regression

Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

X – Mean (A)	Y – Mean (B)	A^2	A*B
-5.38	-26.31	28.99	141.66
-3.38	-14.31	11.46	48.43
-2.38	-13.31	5.69	31.73
-1.38	-11.31	1.92	15.66
-1.38	-10.31	1.92	14.27
-0.38	5.69	0.15	-2.19
0.62	4.69	0.38	2.89
0.62	21.69	0.38	13.35
1.62	-10.31	2.61	-16.65
1.62	7.69	2.61	12.43
2.62	22.69	6.84	59.35
3.62	0.69	13.07	2.50
3.62	22.69	13.07	82.04
		89.08	405.46
		Sum	

$$y = b_0 + b_1 x$$

$$b_1 = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

$$= 405.46 / 89.08$$

$$= 4.55$$

Simple Linear Regression

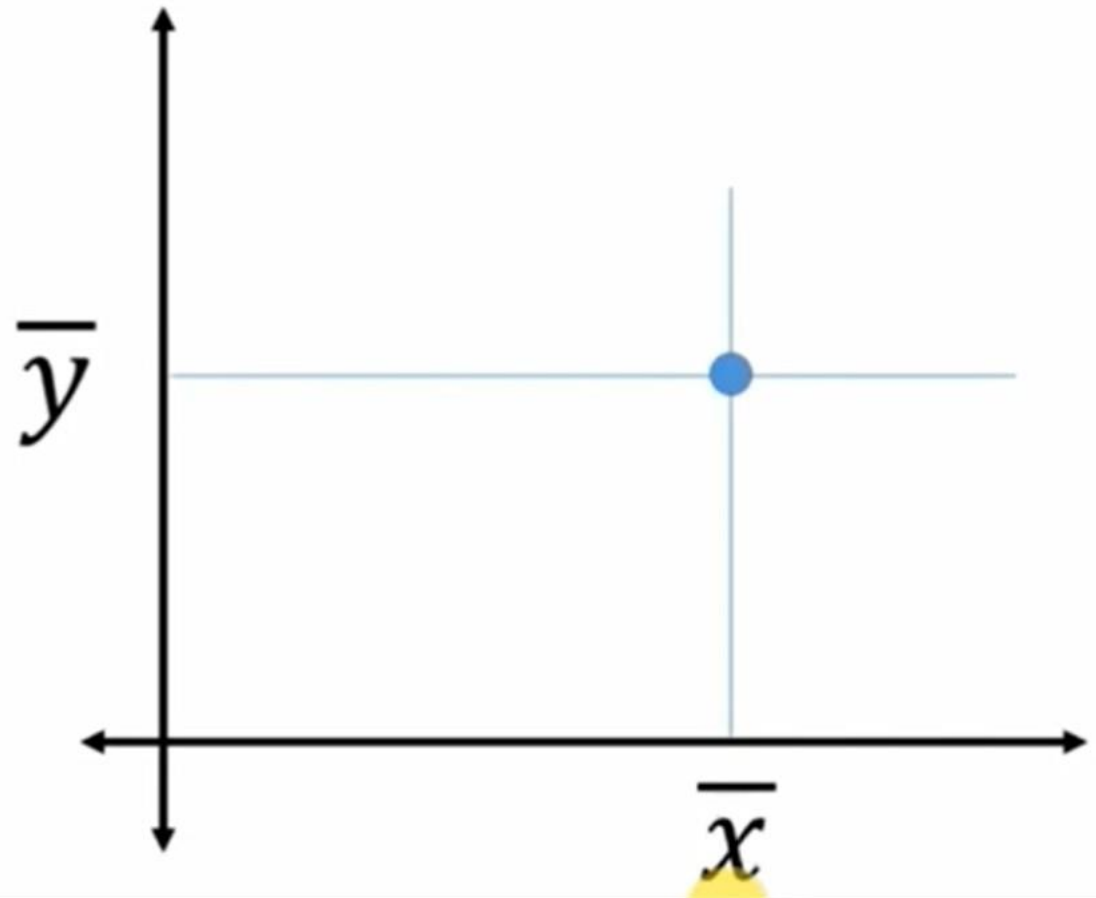
Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

$y = b_0 + b_1 x$ $\quad b_1 = 4.55$ $\quad b_0 = ?$

Simple Linear Regression

Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

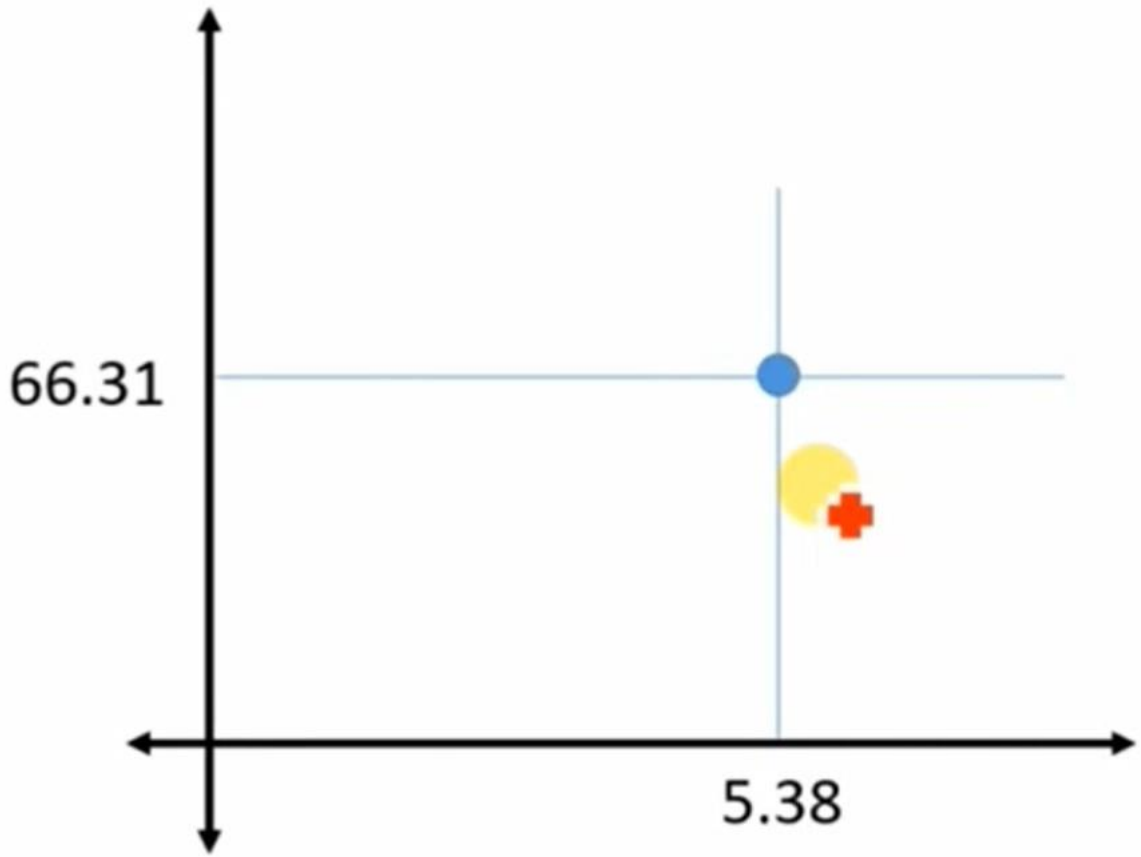
$y = b_0 + b_1 x$ $b_1 = 4.55$ $b_0 = ?$



Simple Linear Regression

Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

$y = b_0 + b_1 x$ $b_1 = 4.55$ $b_0 = ?$



Simple Linear Regression

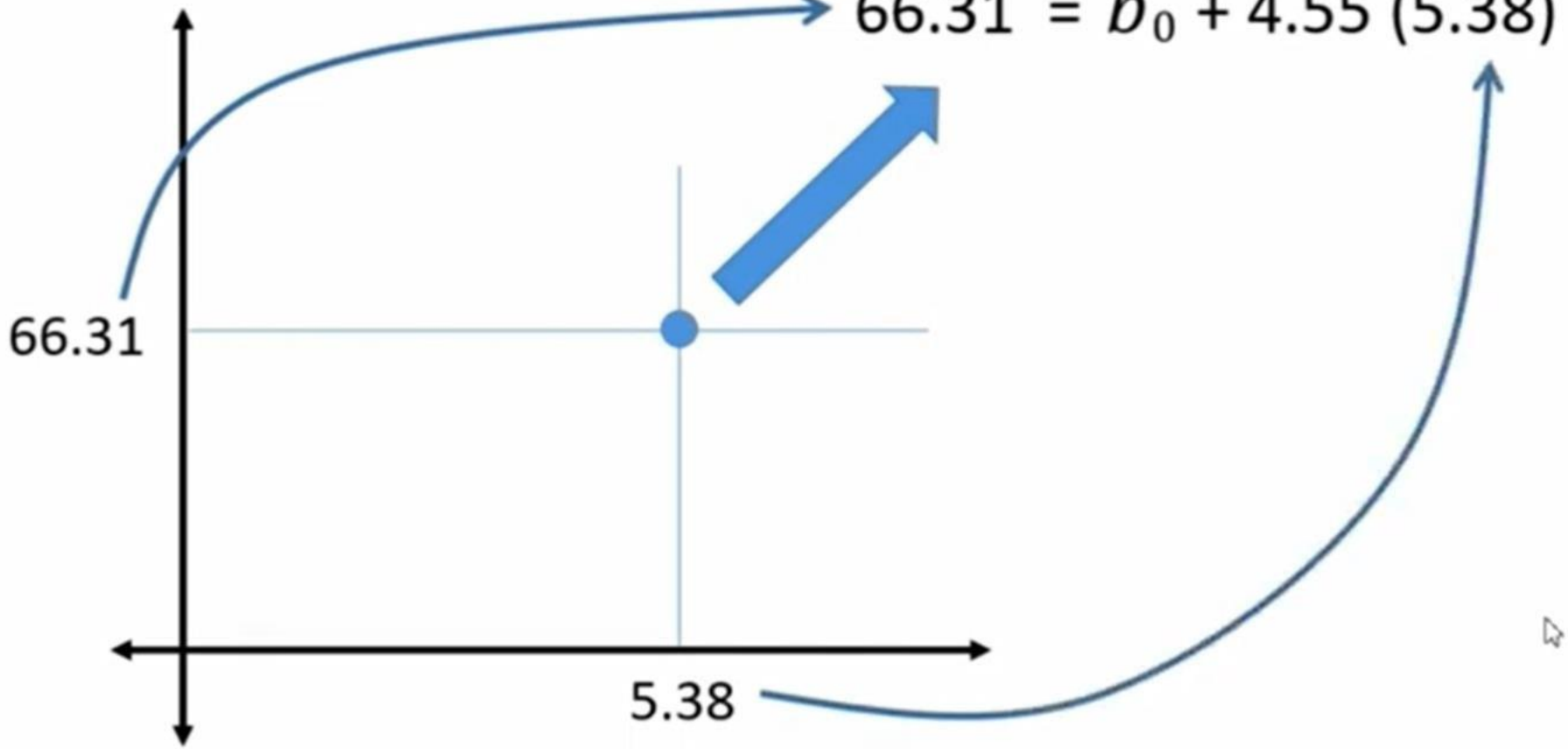
Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

$y = b_0 + b_1 x$

$b_1 = 4.55$

$b_0 = ?$

$66.31 = b_0 + 4.55\ (5.38)$

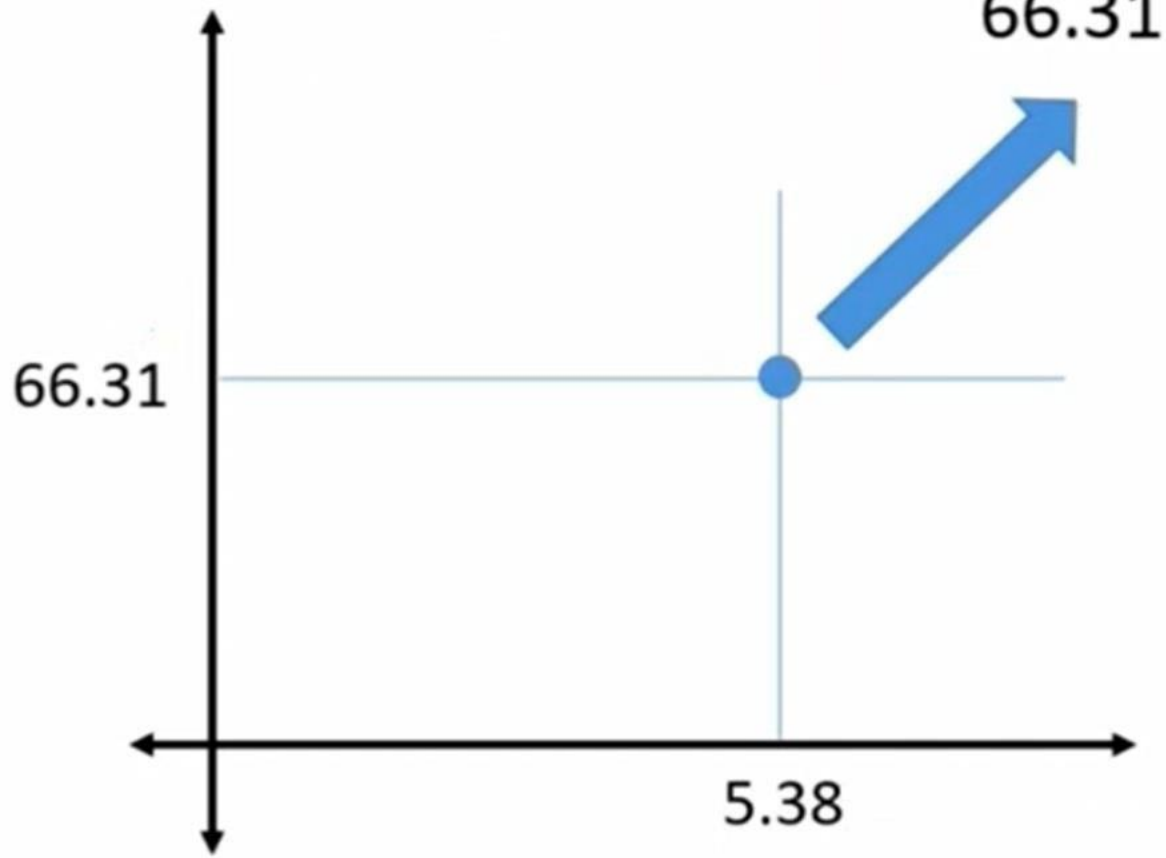


Simple Linear Regression

Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

$y = b_0 + b_1 x$ $\quad b_1 = 4.55$ $\quad b_0 = 41.8$

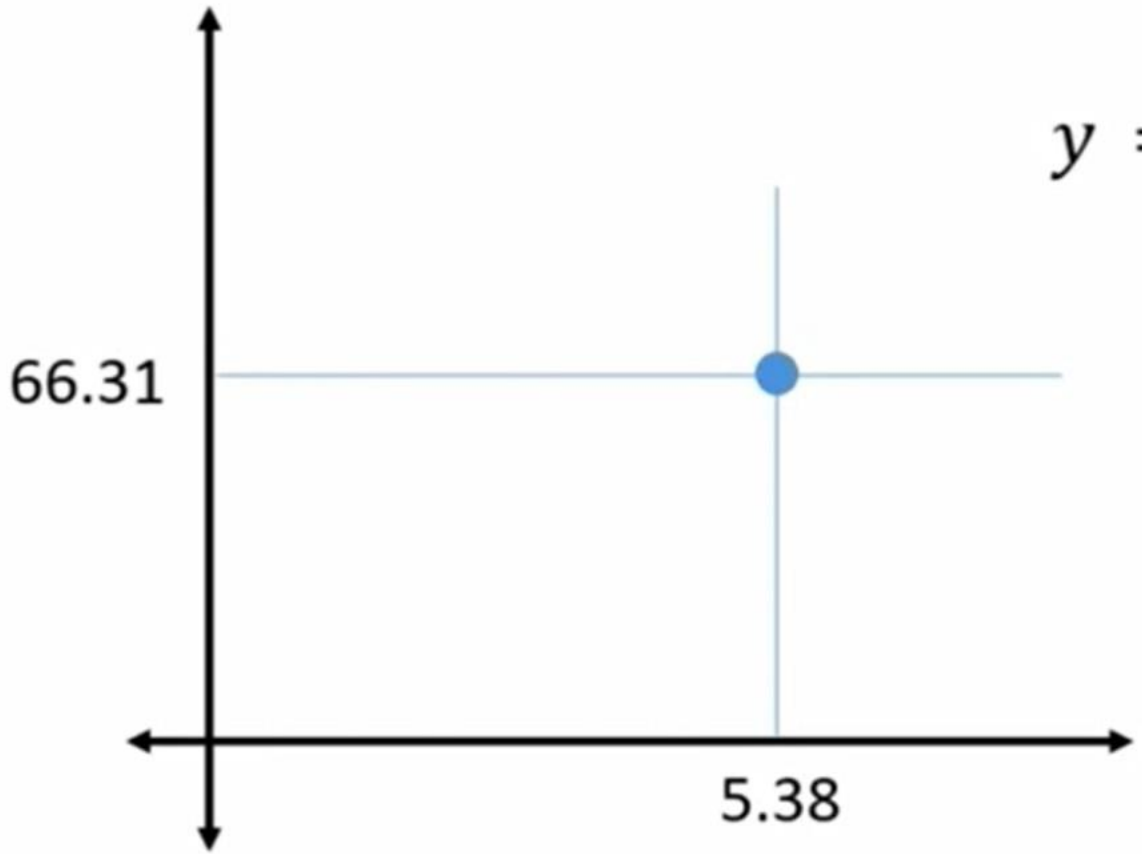
$66.31 = b_0 + 4.55\ (5.38)$



Simple Linear Regression

Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

$y = b_0 + b_1 x$ $b_1 = 4.55$ $b_0 = 41.8$

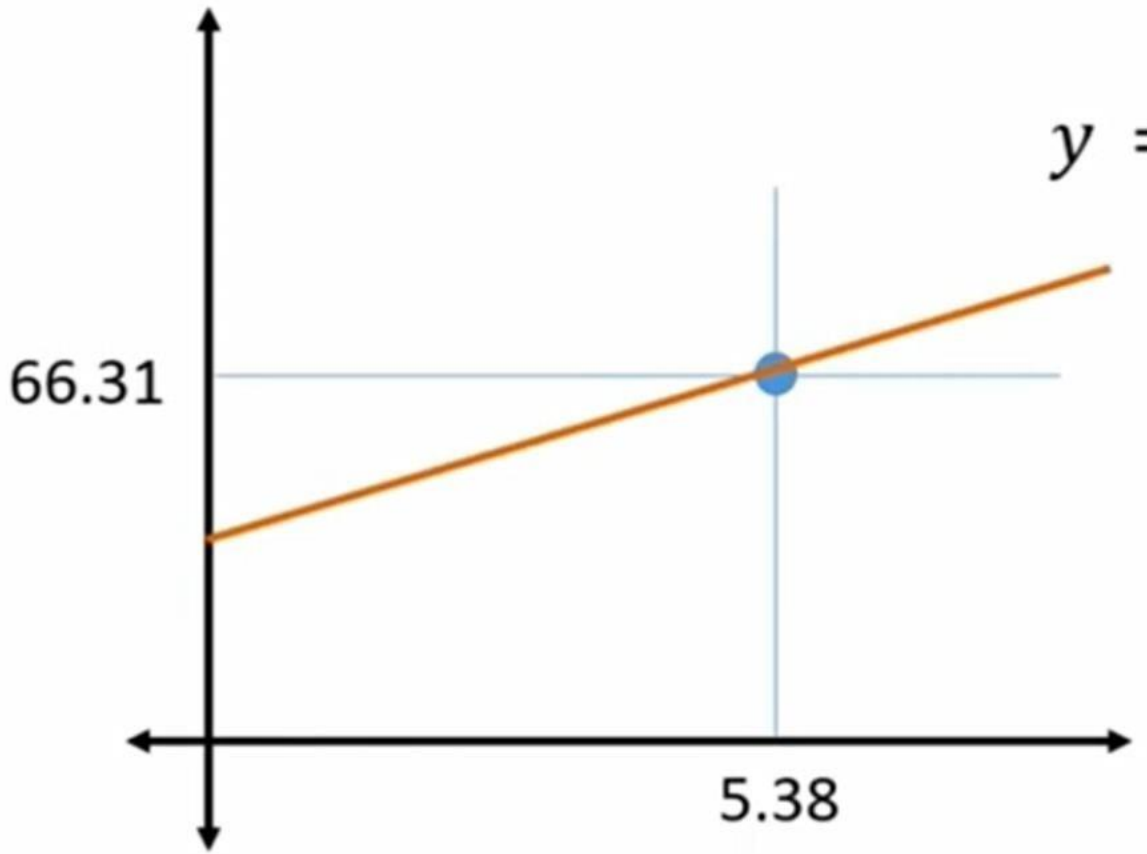


$y = 41.8 + 4.55\,x$

Simple Linear Regression

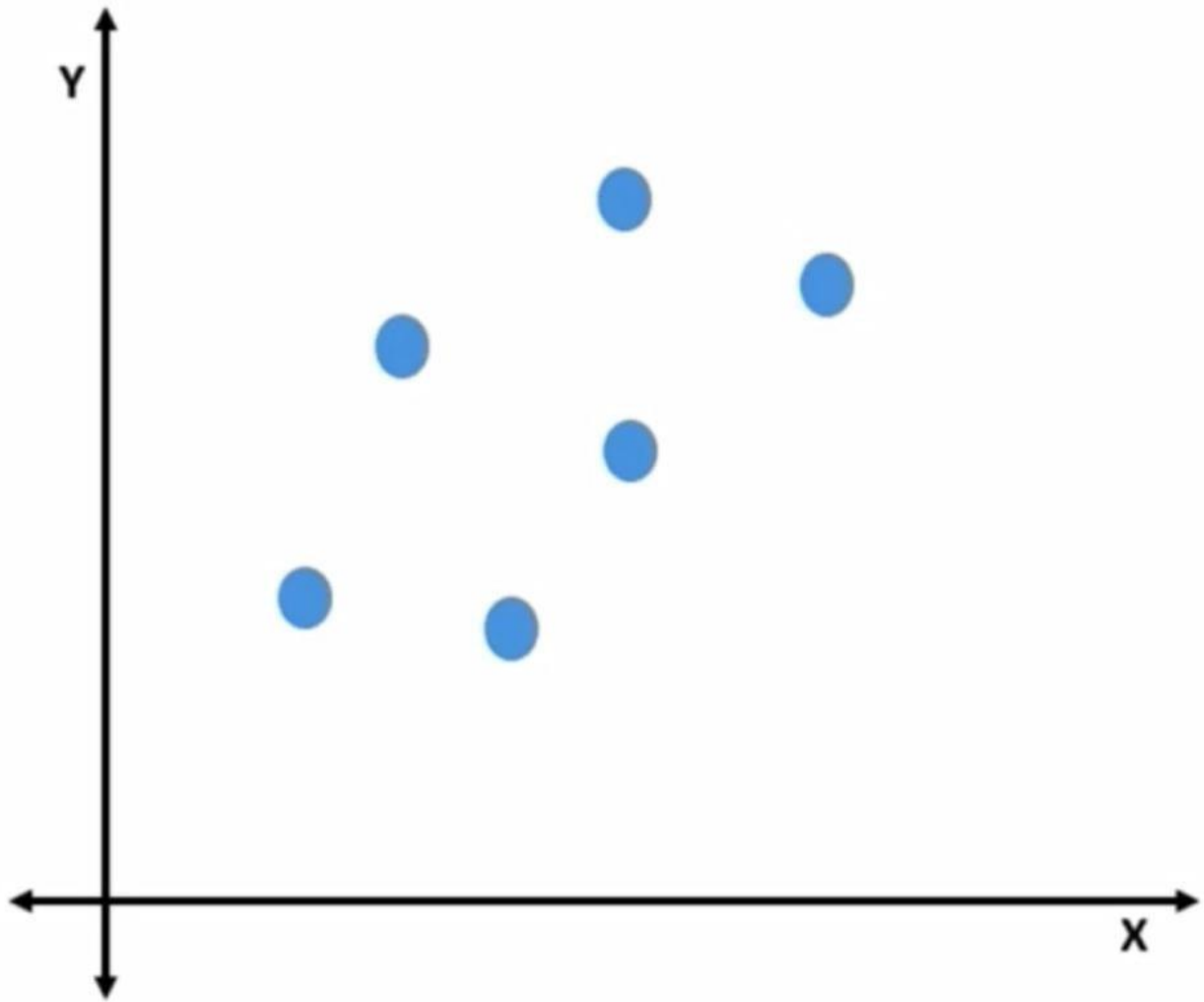
Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

$y = b_0 + b_1 x$ $b_1 = 4.55$ $b_0 = 41.8$

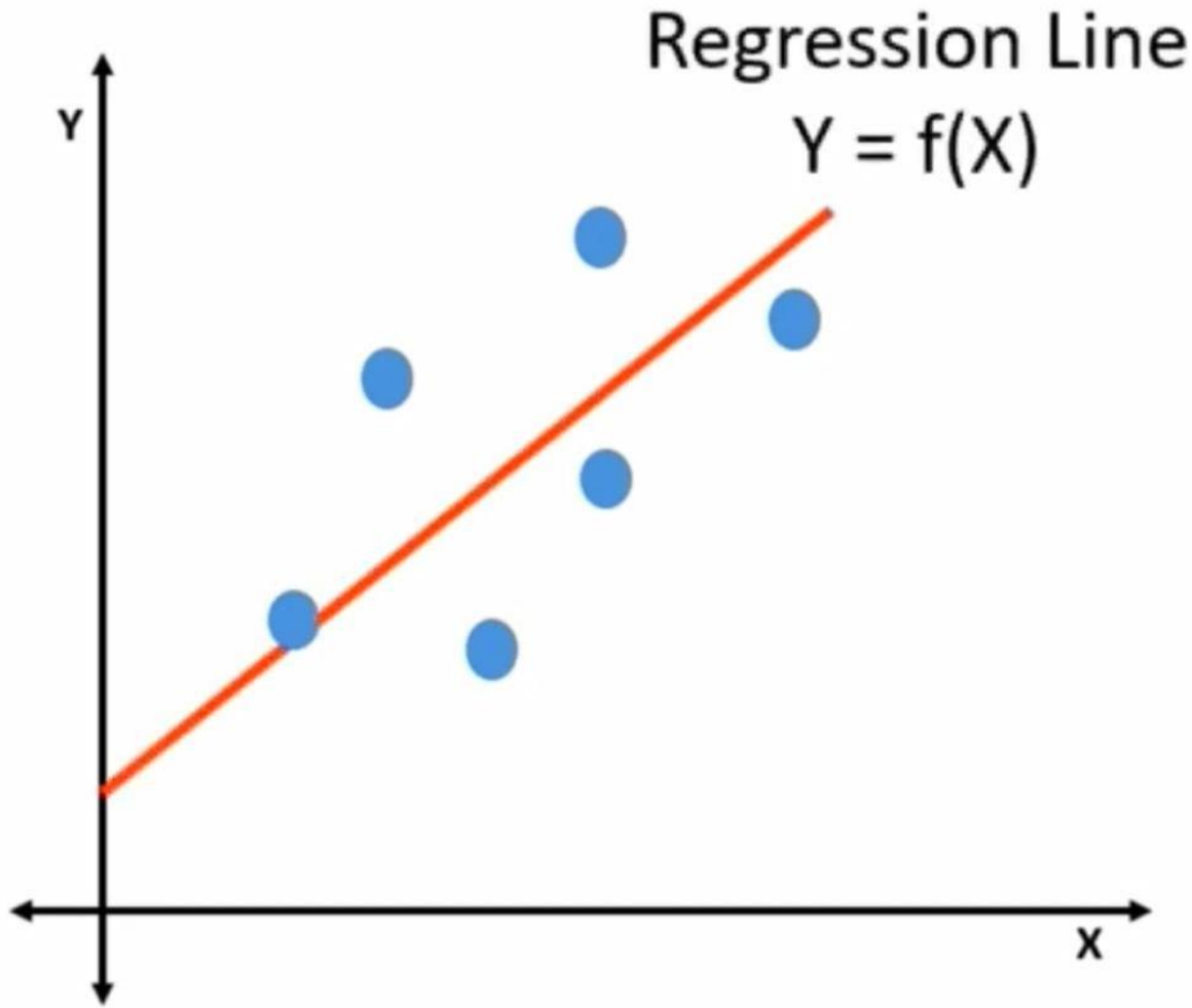


$y = 41.8 + 4.55\,x$

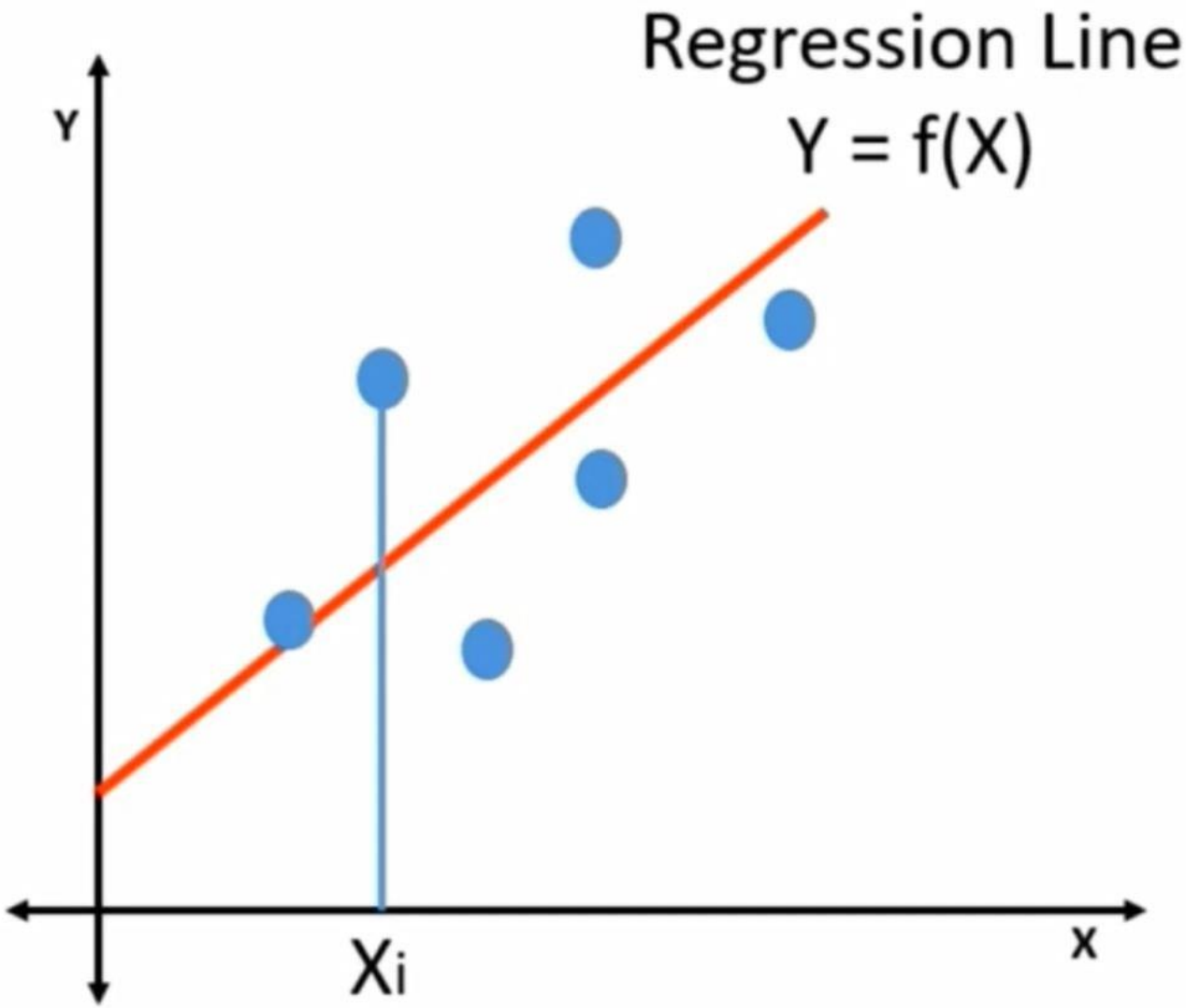
Ordinary Least Square



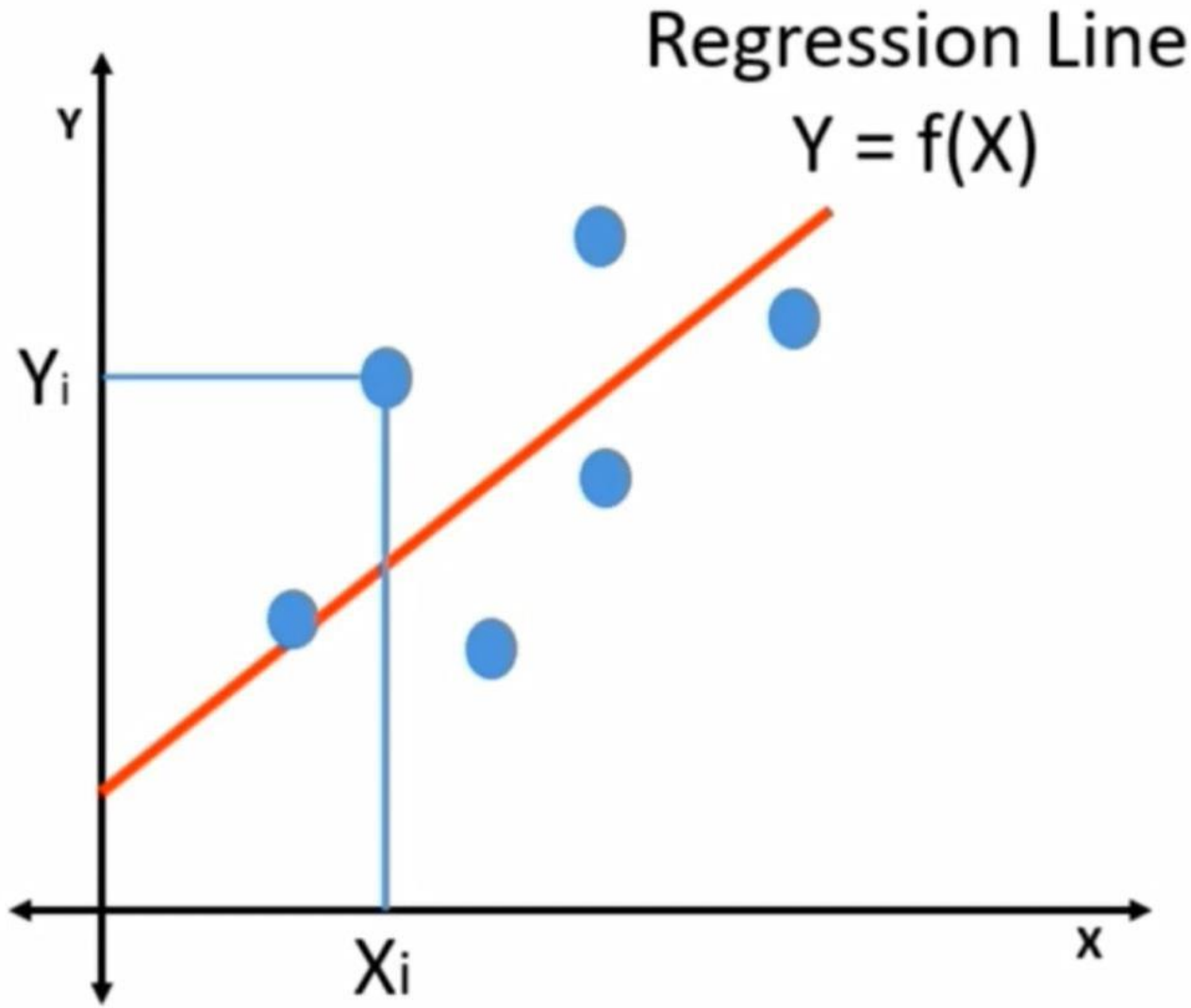
Ordinary Least Square



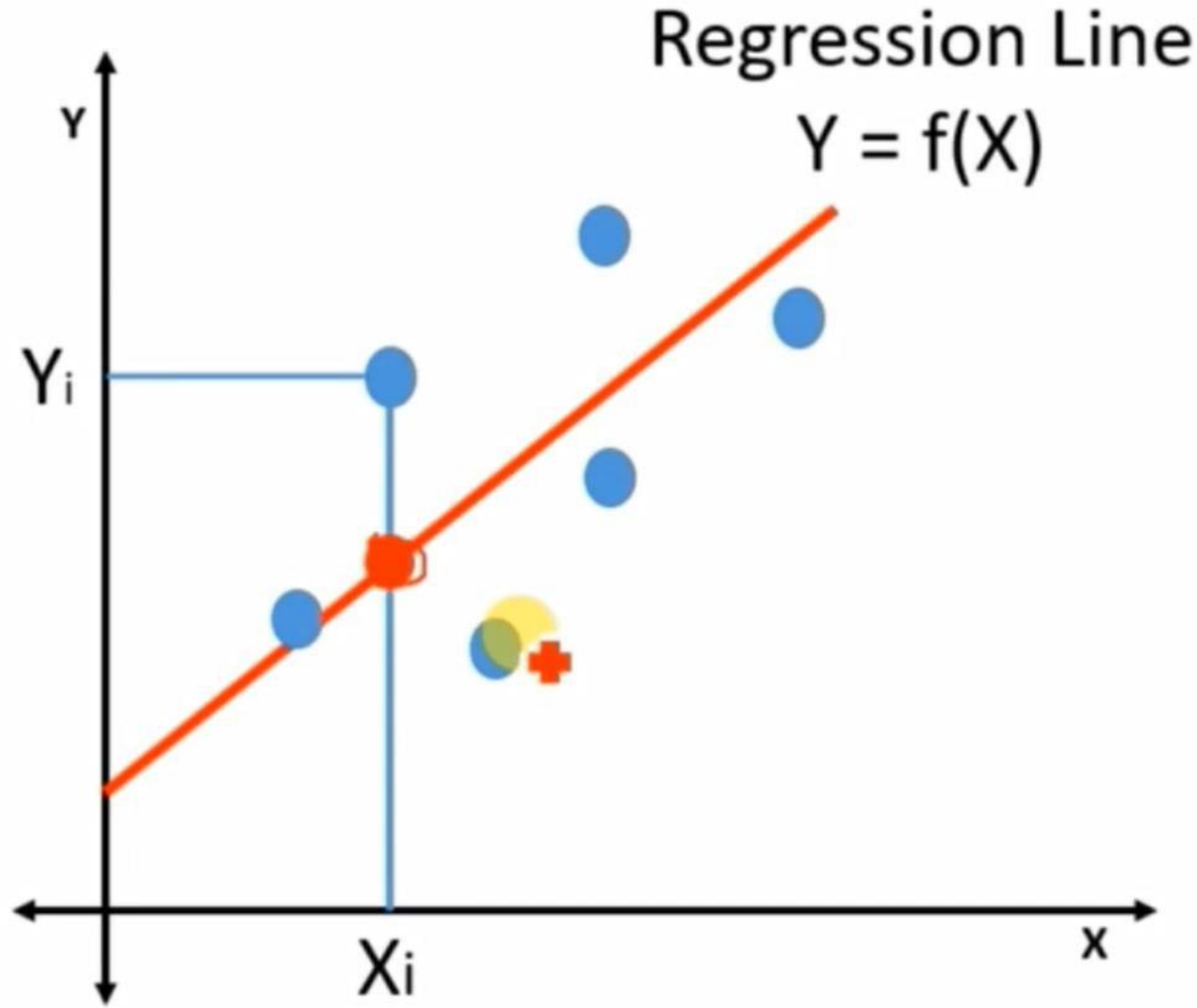
Ordinary Least Square



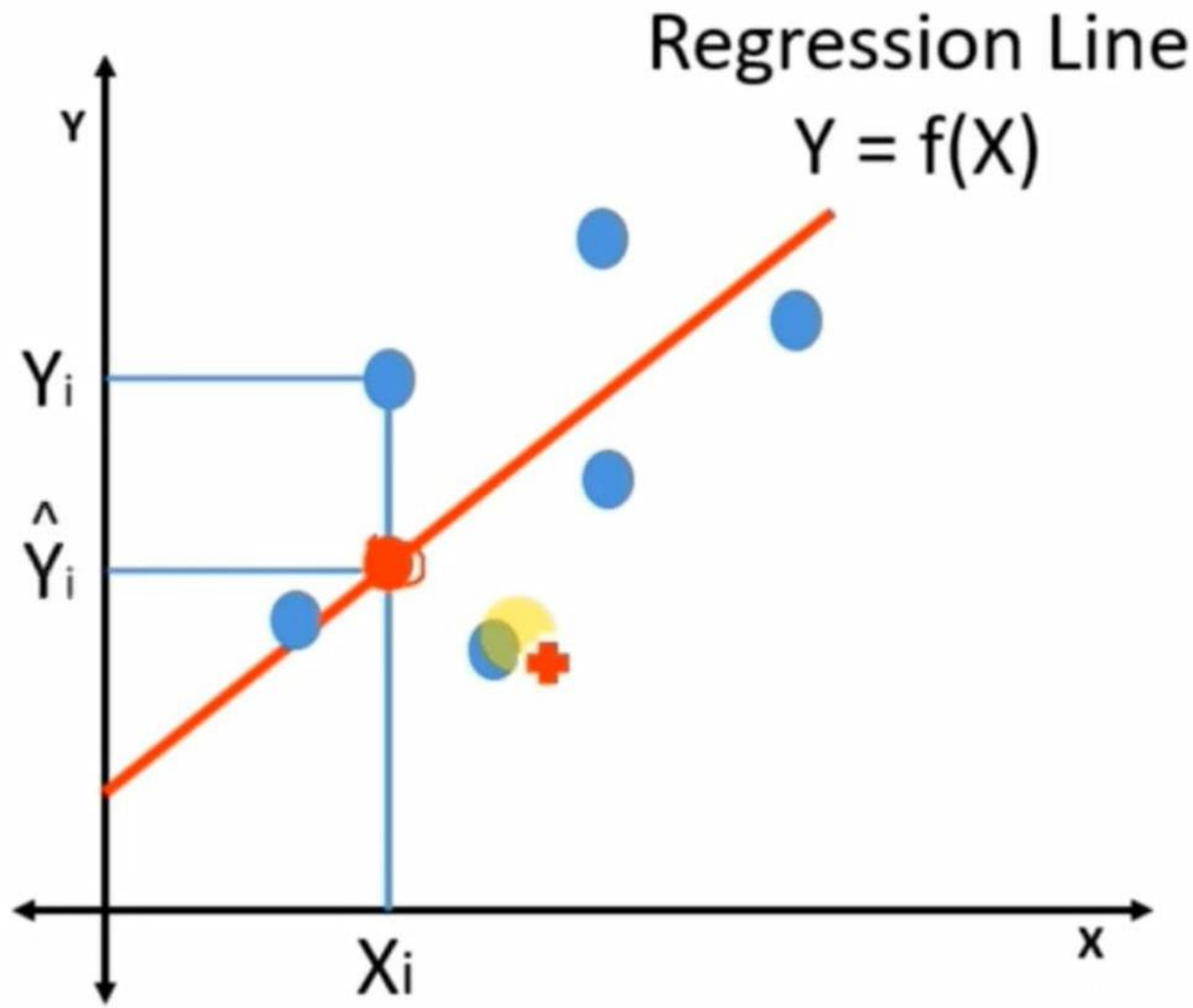
Ordinary Least Square



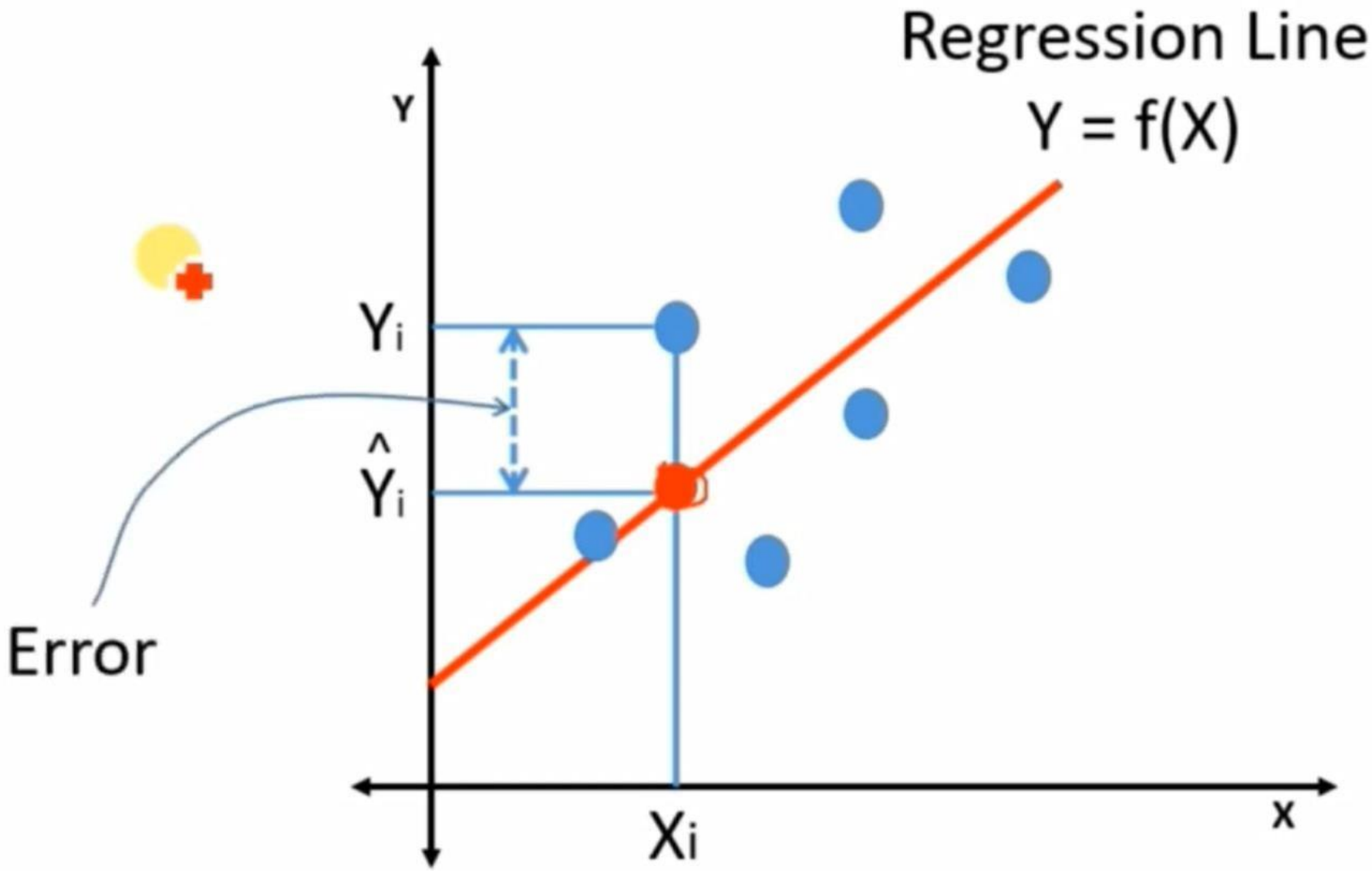
Ordinary Least Square



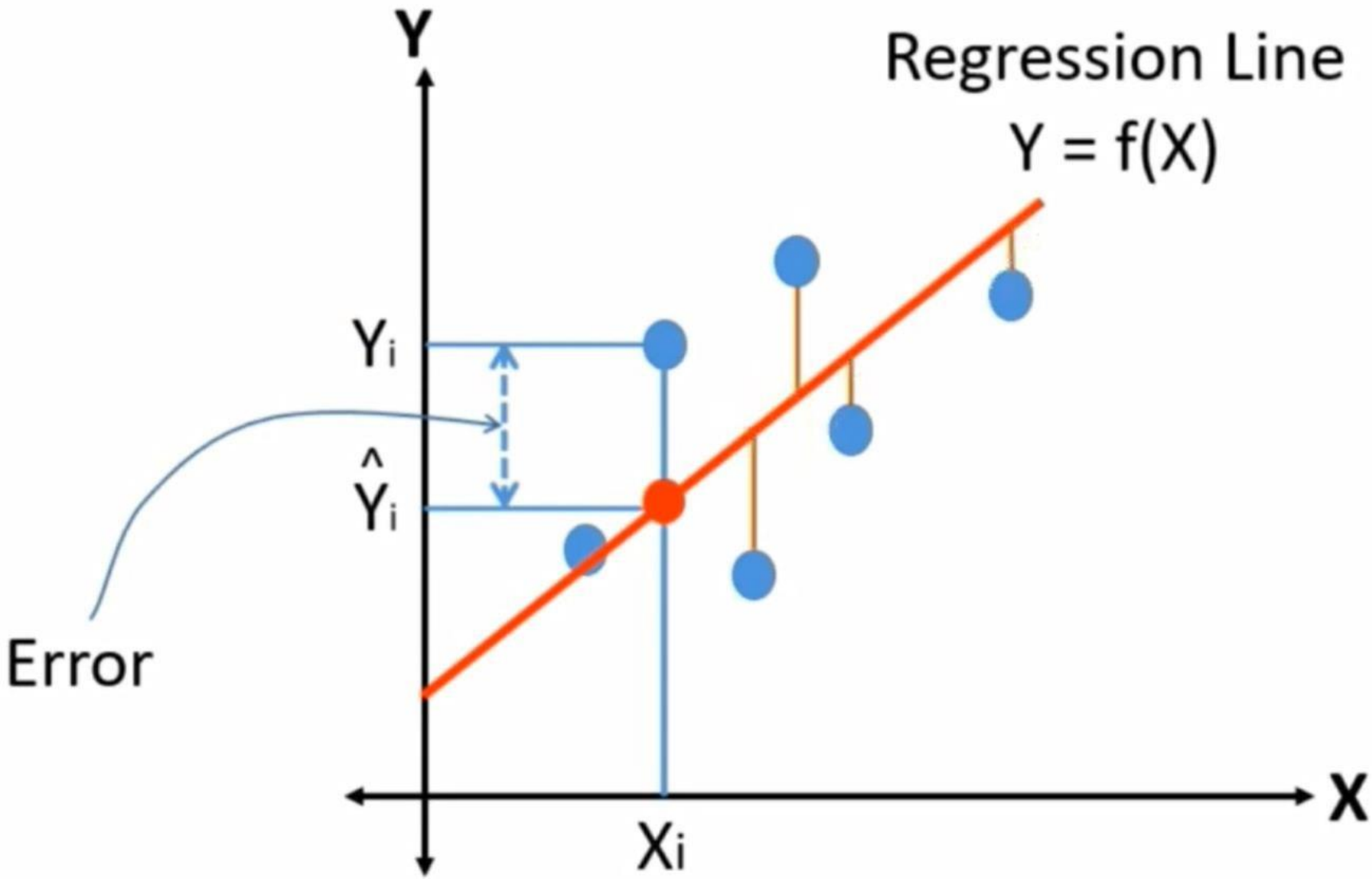
Ordinary Least Square



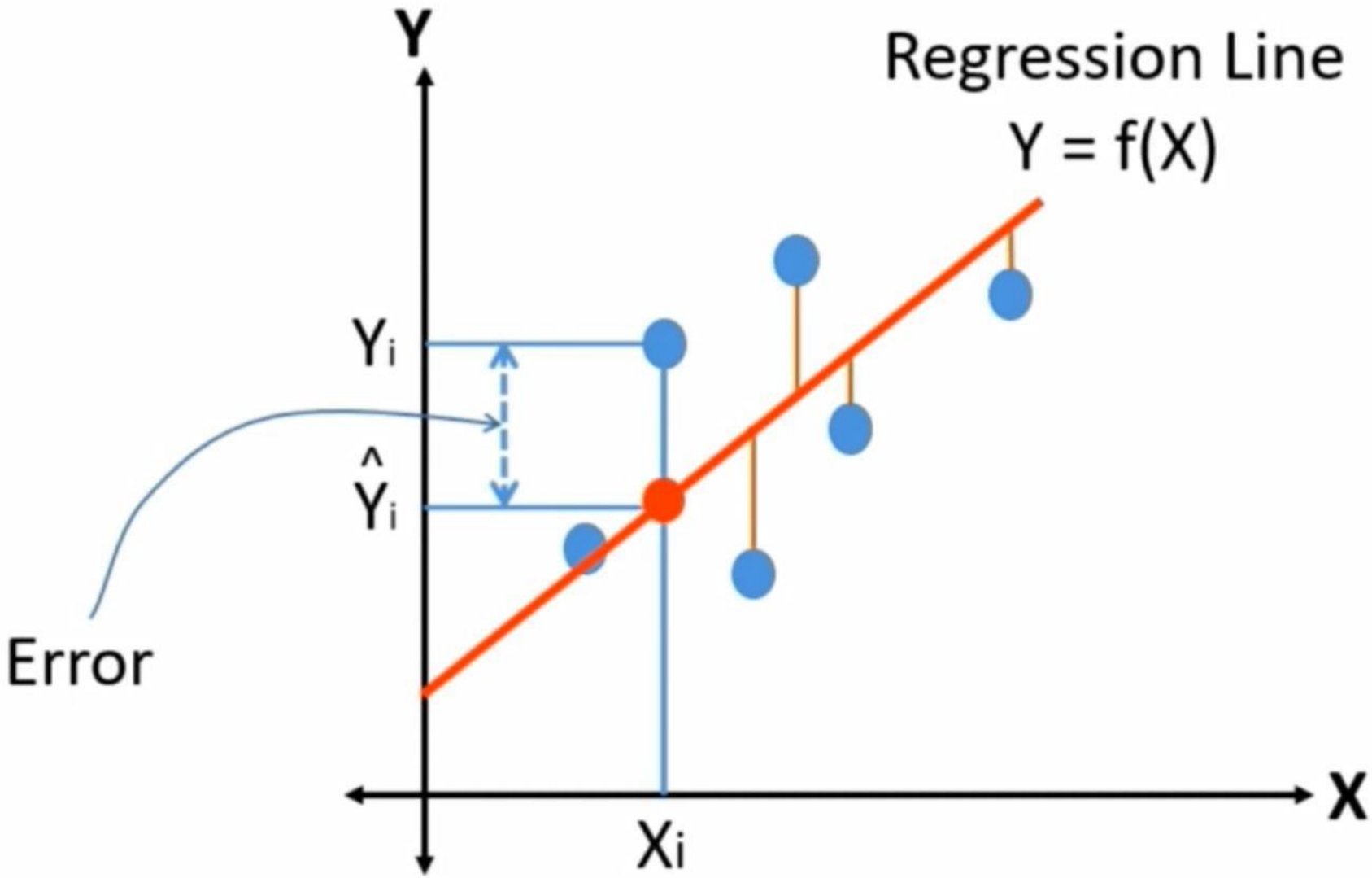
Ordinary Least Square



Ordinary Least Square



Ordinary Least Square



Minimum

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