Gift This Course

A-Z Machine Learning using Azure Machine Learning (AzureML)

Hands on AzureML: From Azure Machine Learning Introduction to Advance Machine Learning Algorithms. No Coding Required.

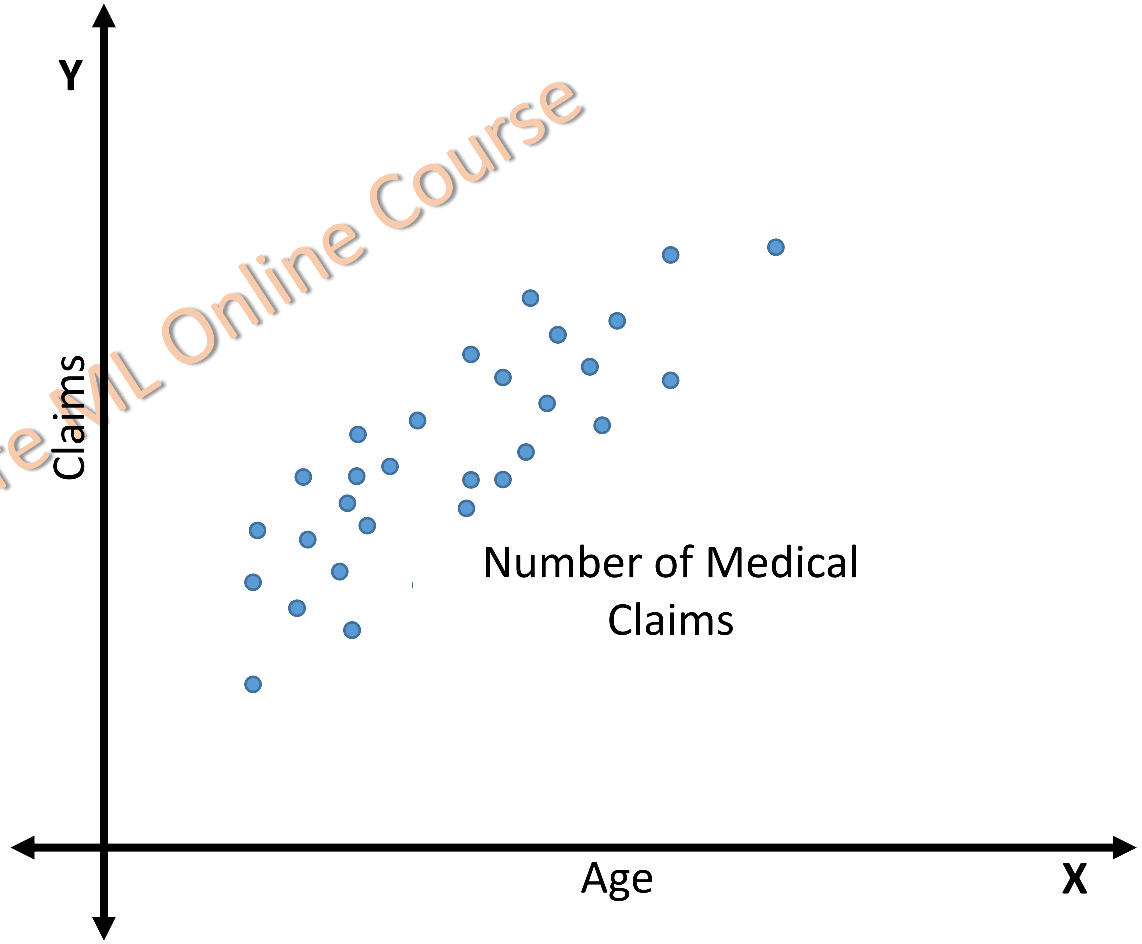
BEST SELLER 4.3 (215 ratings) 1,597 students enrolled

Created by Jitesh Khurkhuriya Last updated 3/2018 English English

Section 7
Regression

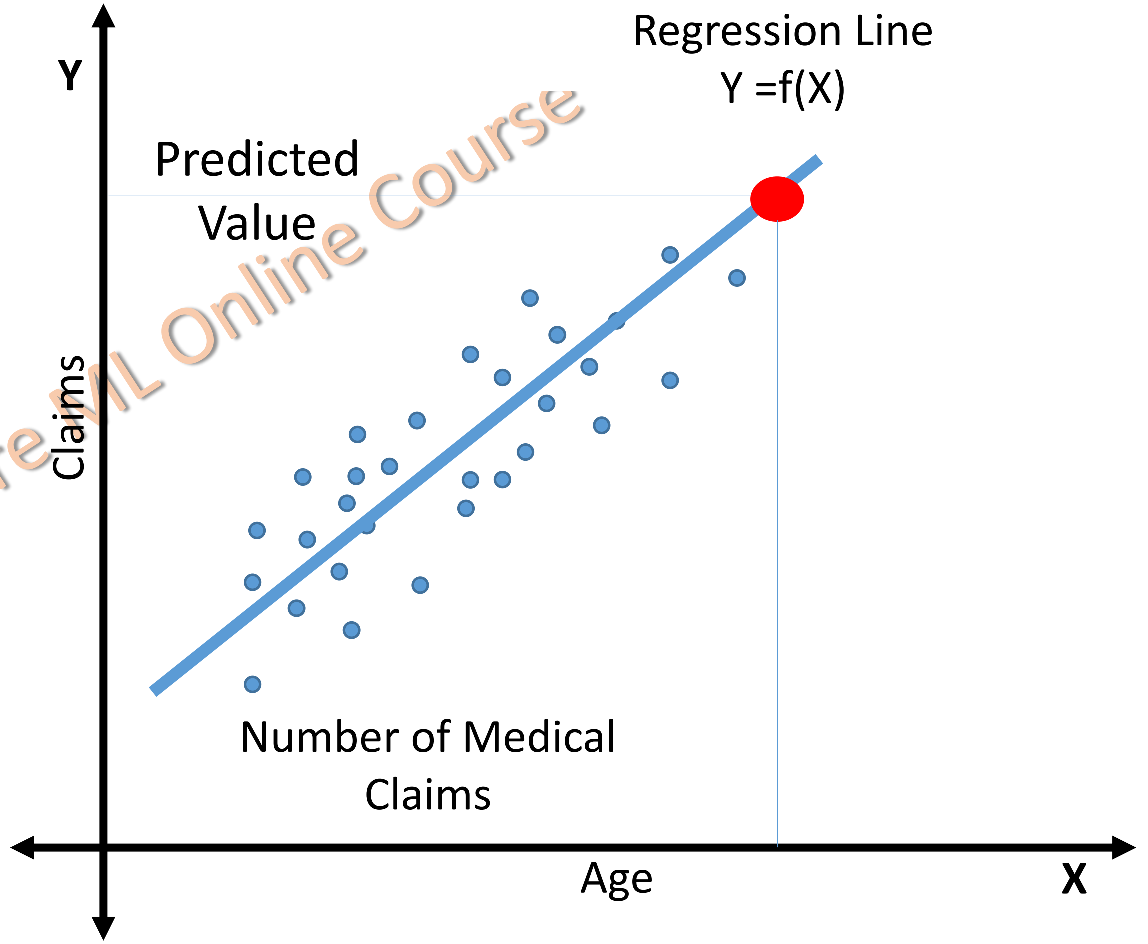
Regression Analysis

- Statistical process for estimating the relationships among variables
- Relationship between a dependent variable and one or more independent variables (or 'predictors')
- The predictor is a continuous variable
- Can also be used to infer causal relationships between dependent and independent variables.

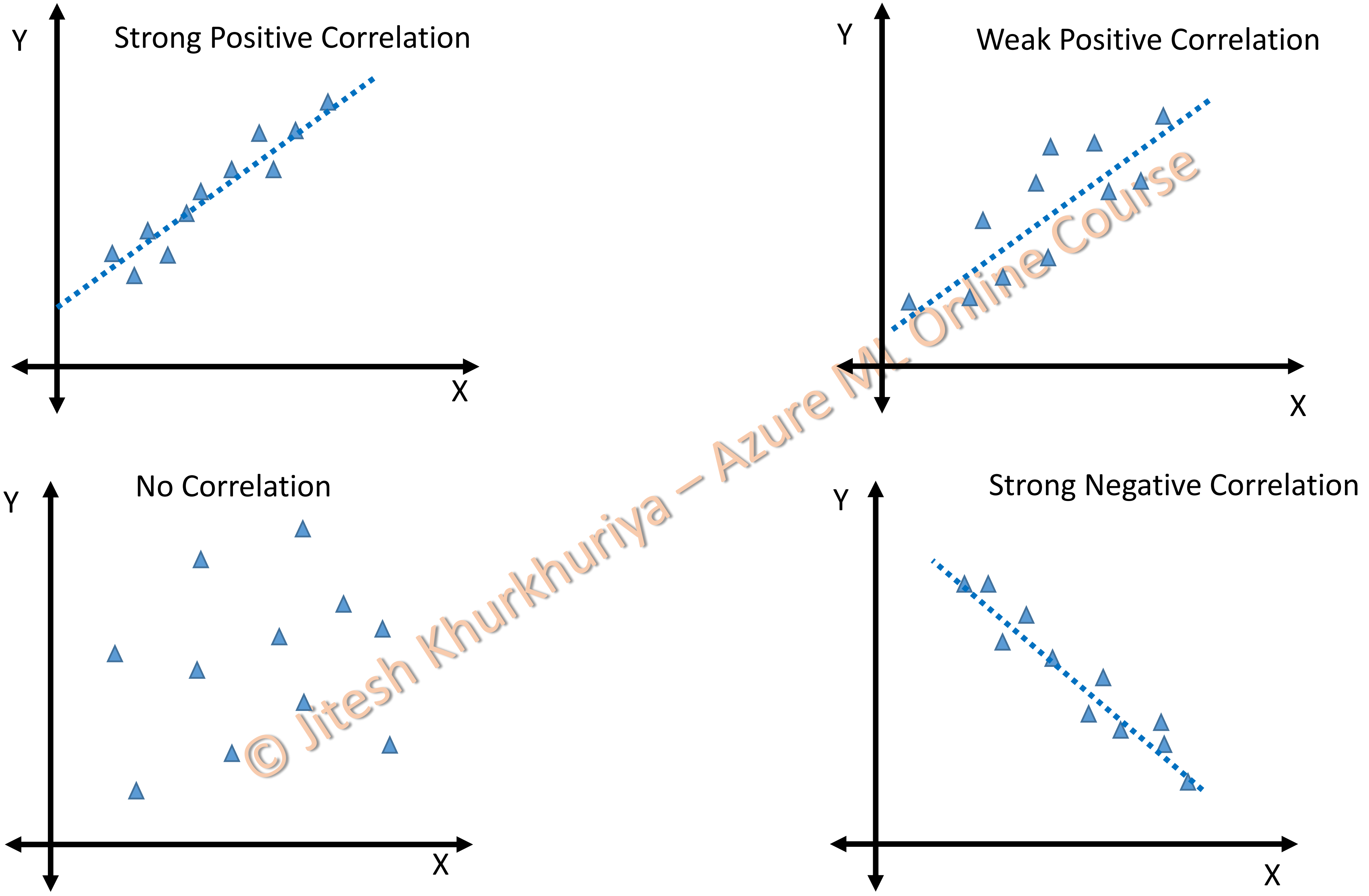


Regression Analysis

- Statistical process for estimating the relationships among variables
- Relationship between a dependent variable and one or more independent variables (or 'predictors')
- The predictor is a continuous variable
- Can also be used to infer causal relationships between dependent and independent variables.



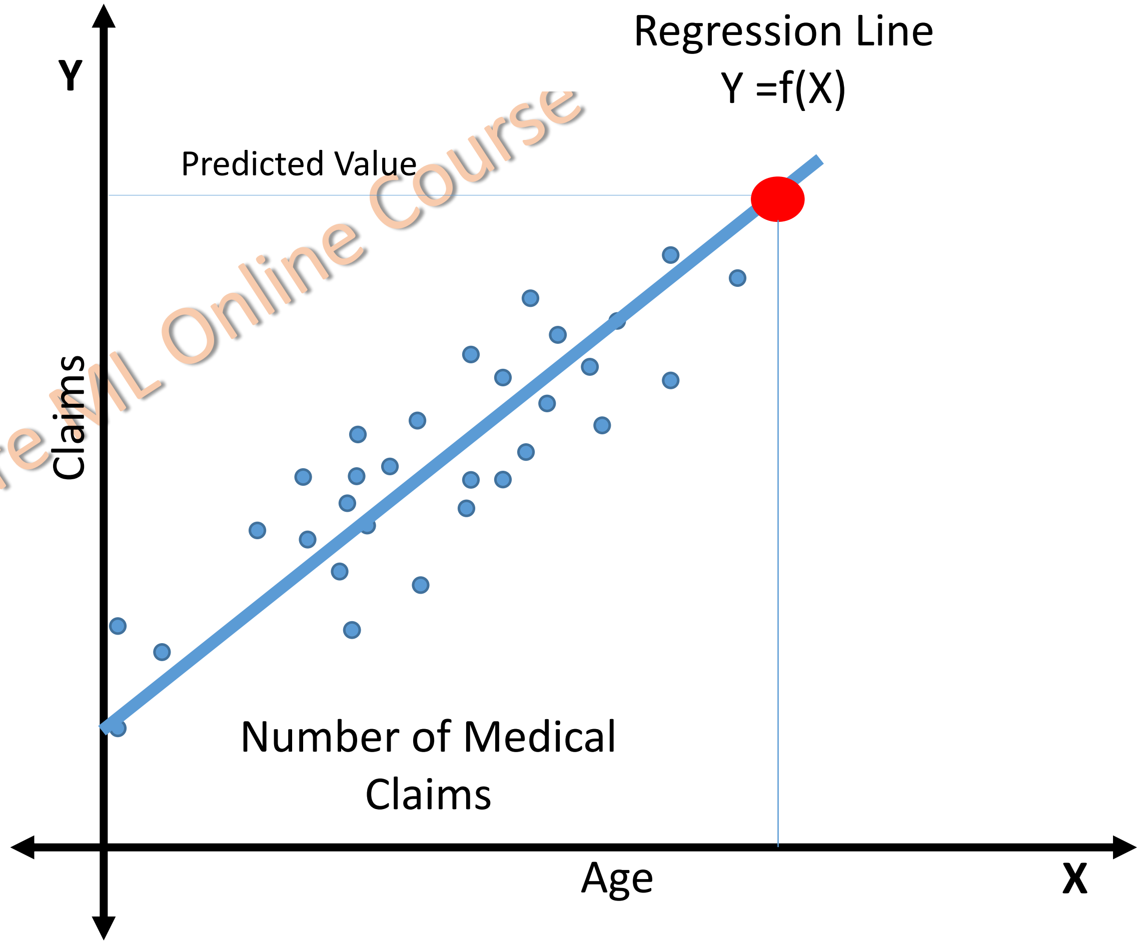
Causal Relationship?


Strong Positive Correlation
Y
X
Weak Positive Correlation
Y
X
No Correlation
Y
X
Strong Negative Correlation
Y
X

Linear Regression

Simple Regression

$$Y = \beta_0 + \beta_1 X$$

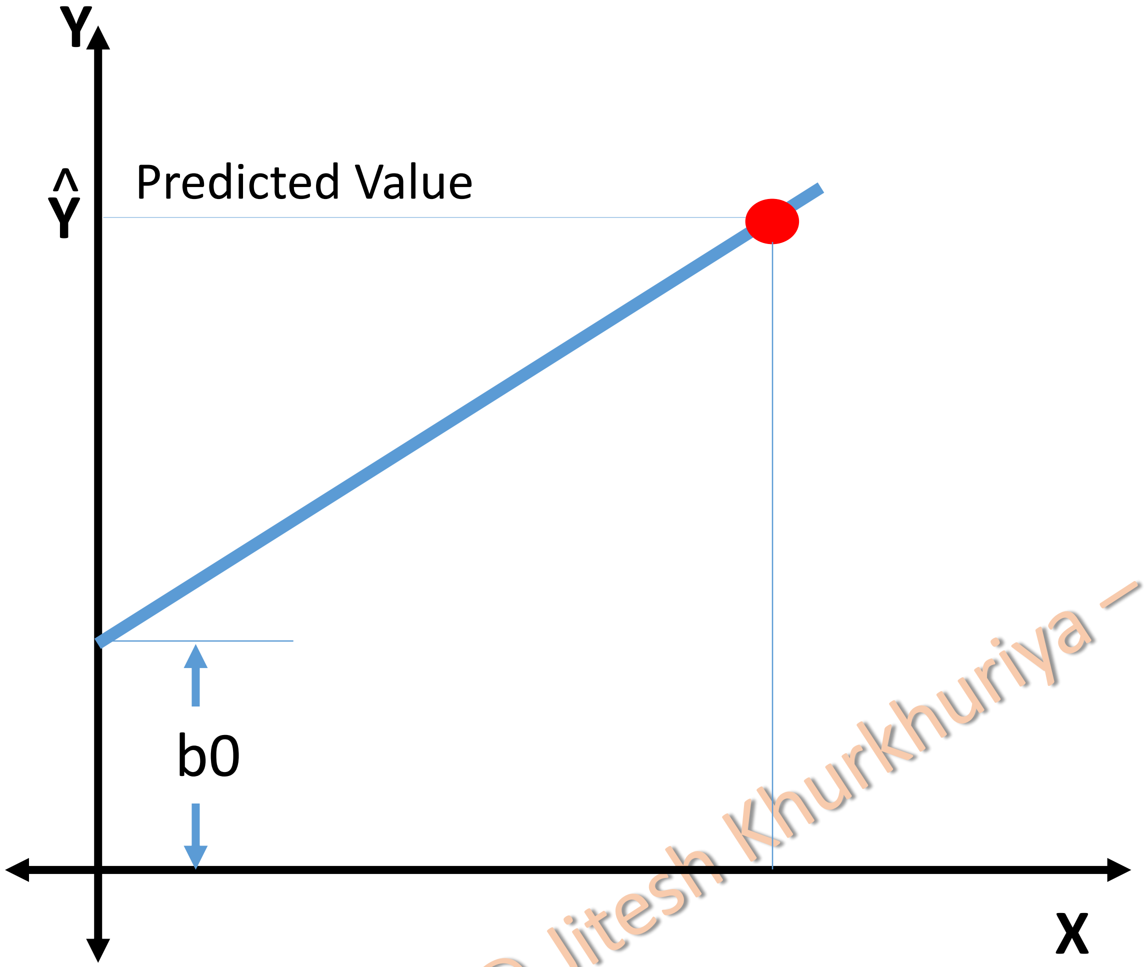


multivariate linear regression.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots\ldots + \beta_n X_n$$

Simple Linear Regression

Simple Linear Regression



Simple Regression

$$Y = b0 + b1X$$

Dependent Variable

Independent Variable

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
		Sum	

$$Y = b0 + b1X$$

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	

$$Y = b0 + b1X$$

$$b1 = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sum (X - \overline{X})^2}$$

$$= 405.46 / 89.08$$

$$= 4.55$$

Simple Linear Regression

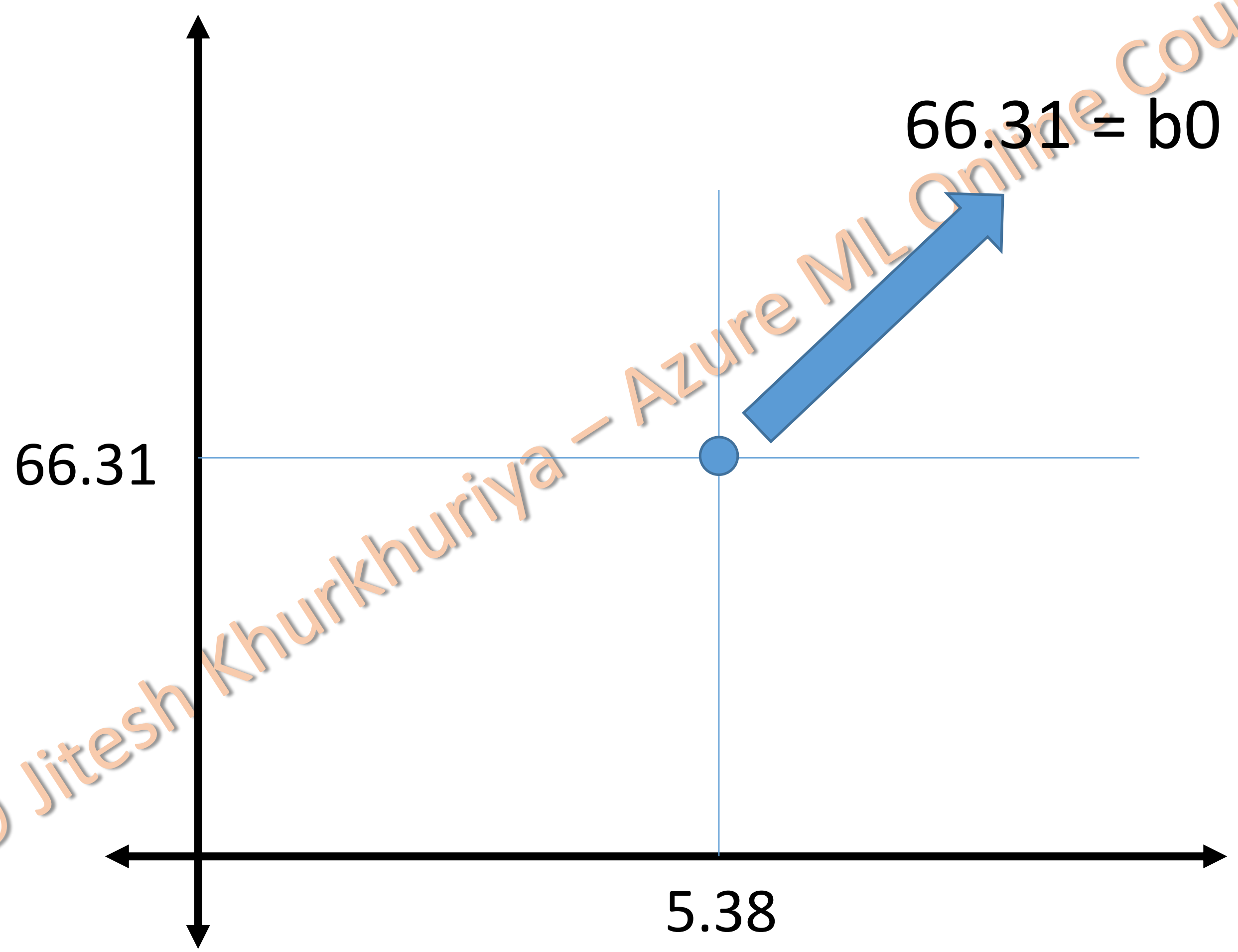
Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

$Y = b0 + b1X$

$b1 = 4.55$

$b0 = 41.8$

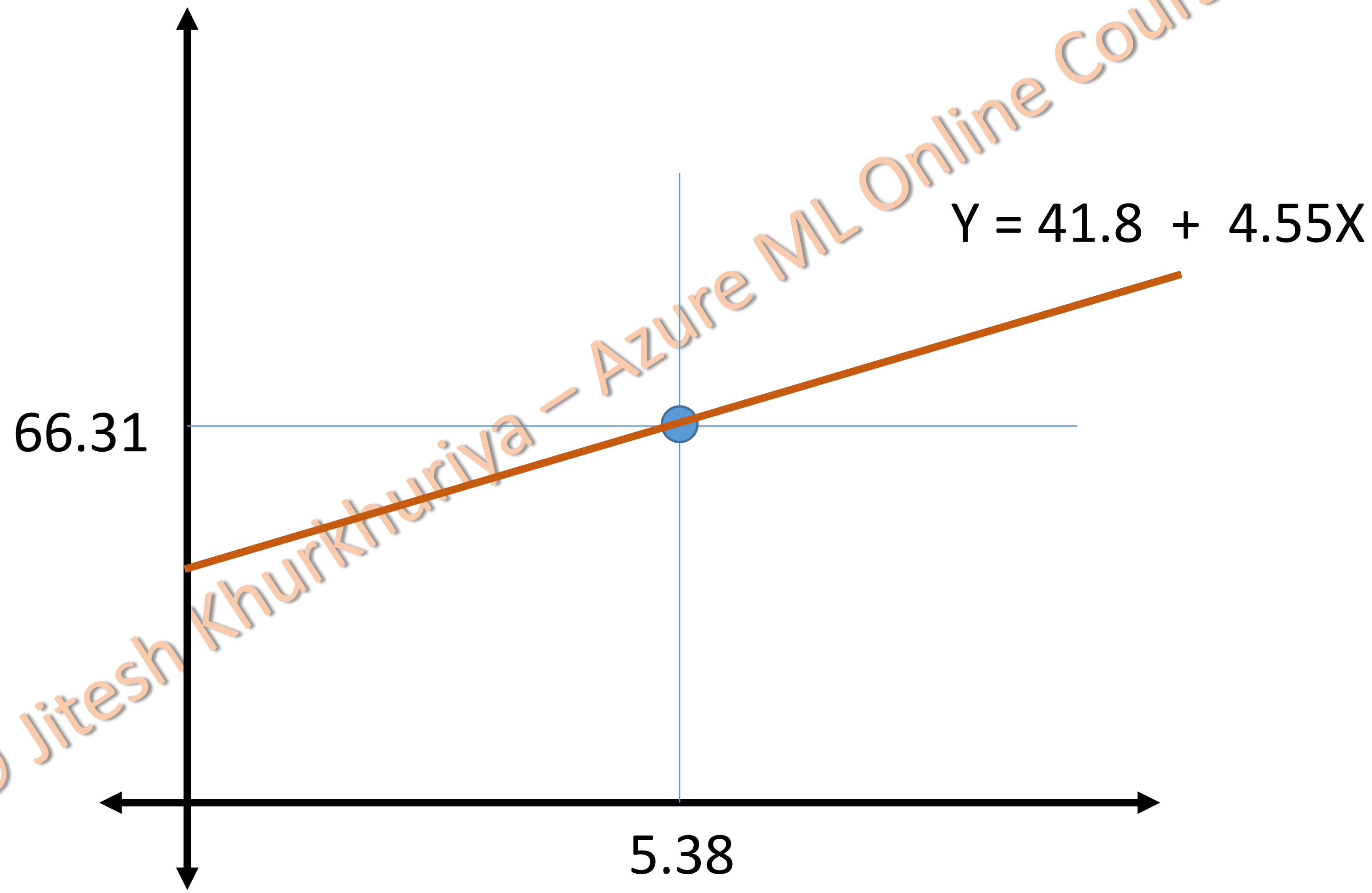
$66.31 = b0 + 4.55\ (5.38)$



Simple Linear Regression

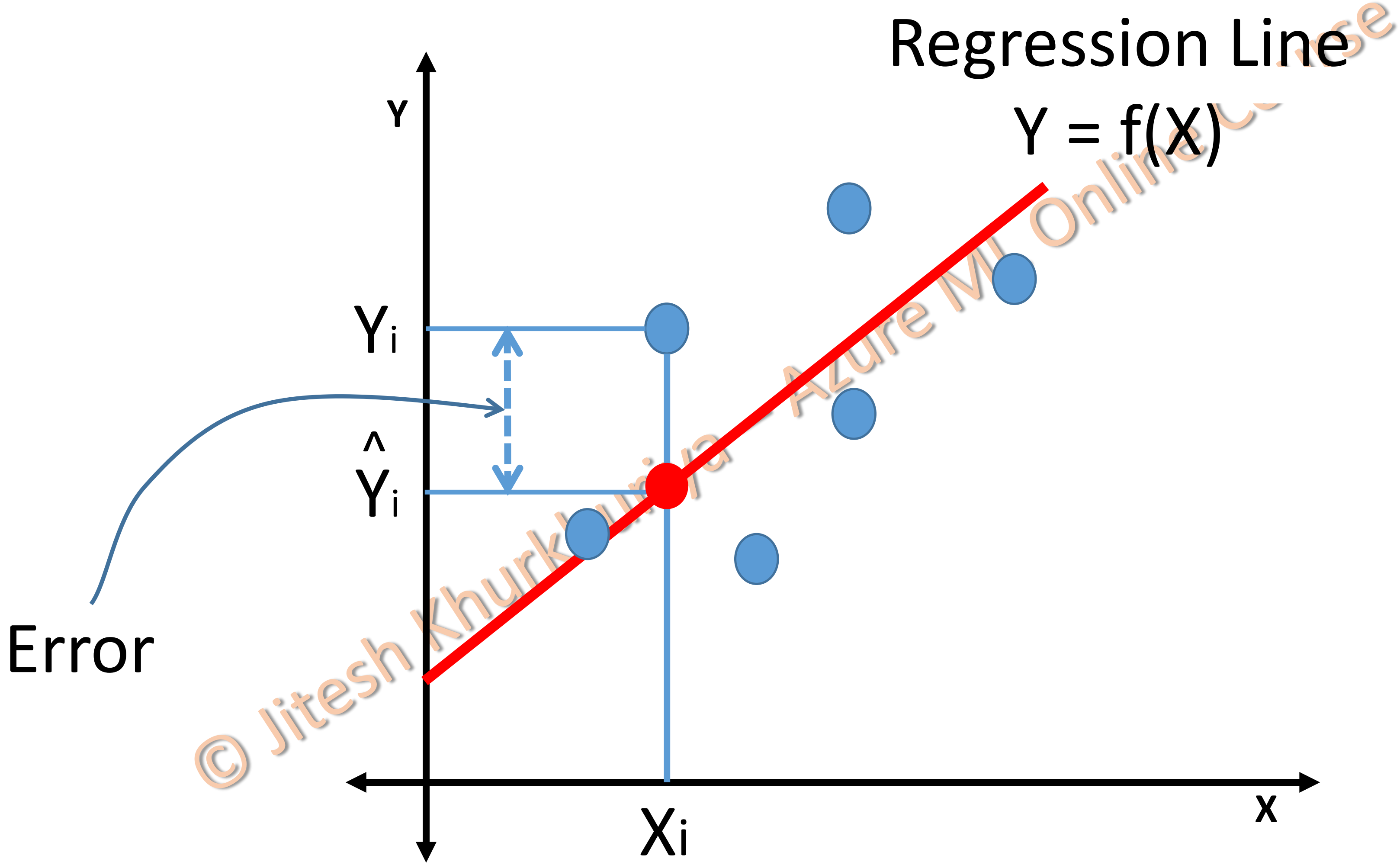
Hrs Studied (X)	Marks (Y)
0	40
2	52
3	53
4	55
4	56
5	72
6	71
6	88
7	56
7	74
8	89
9	67
9	89
5.38	66.31
Mean	

$Y = b0 + b1X$ $\quad b1 = 4.55 \quad b0 = 41.8$

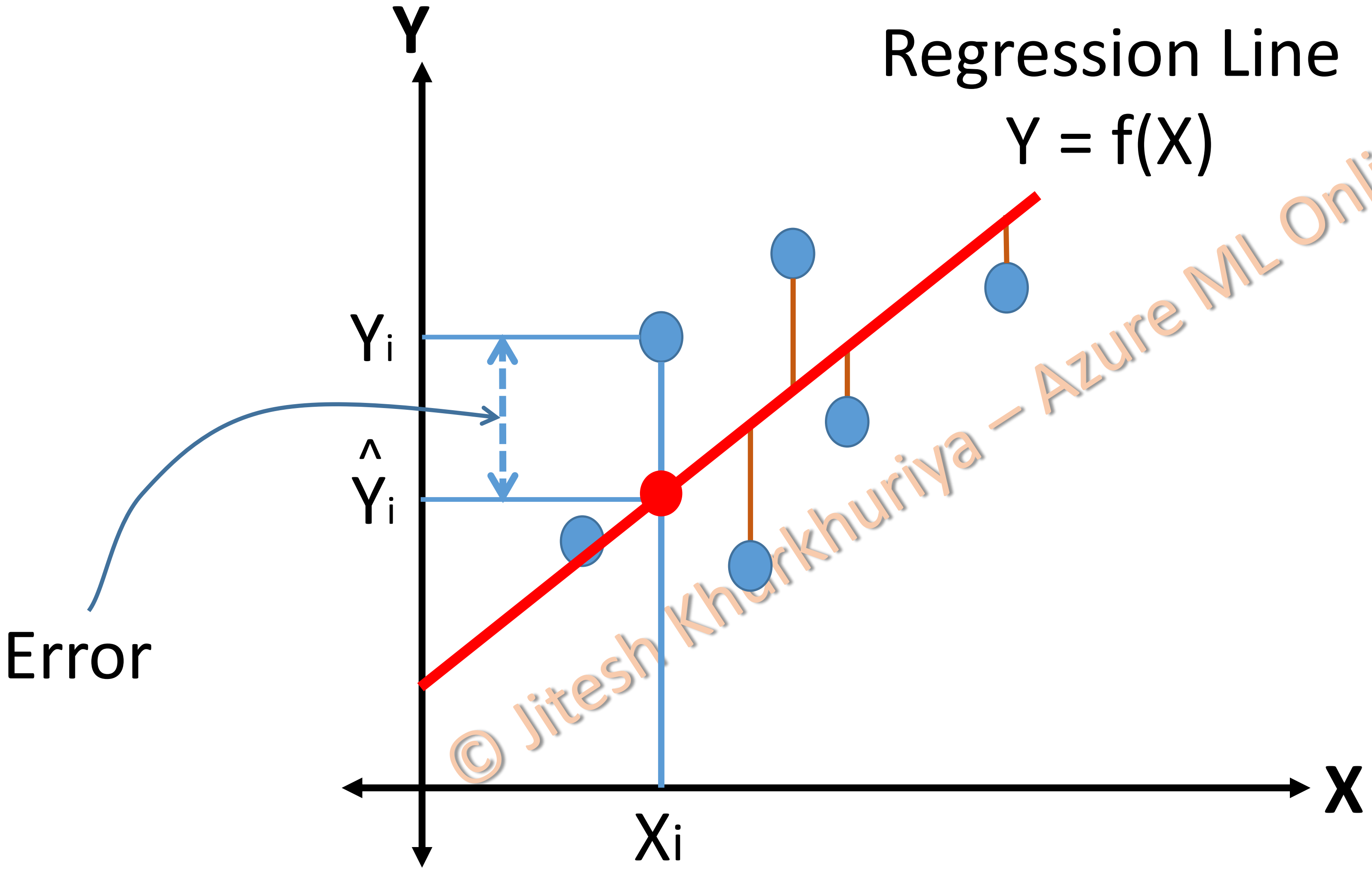


Common Regression Terms

Ordinary Least Square



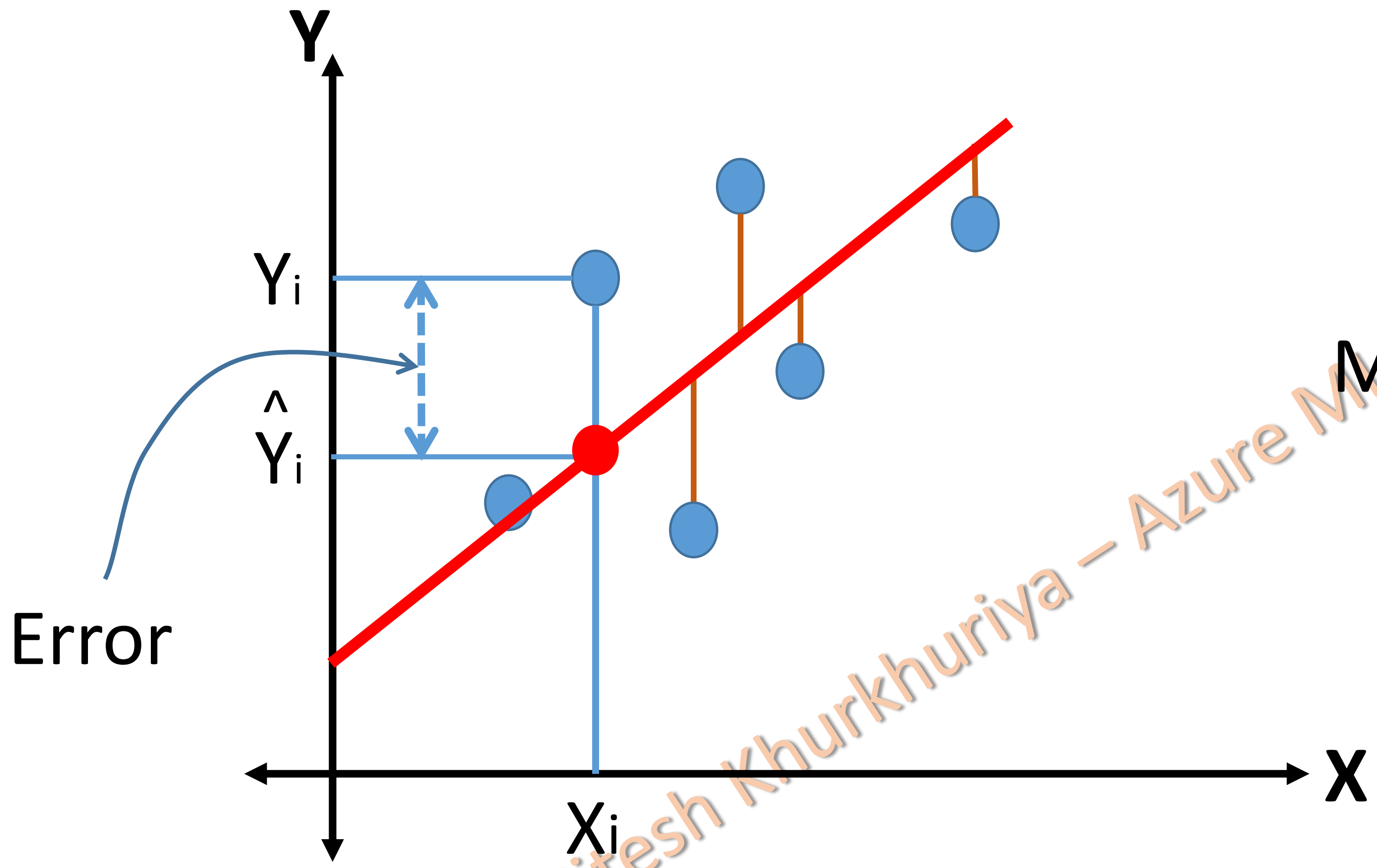
Ordinary Least Square



Minimum

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

Mean Absolute Error



$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|yi - \hat{y}i\right|$$

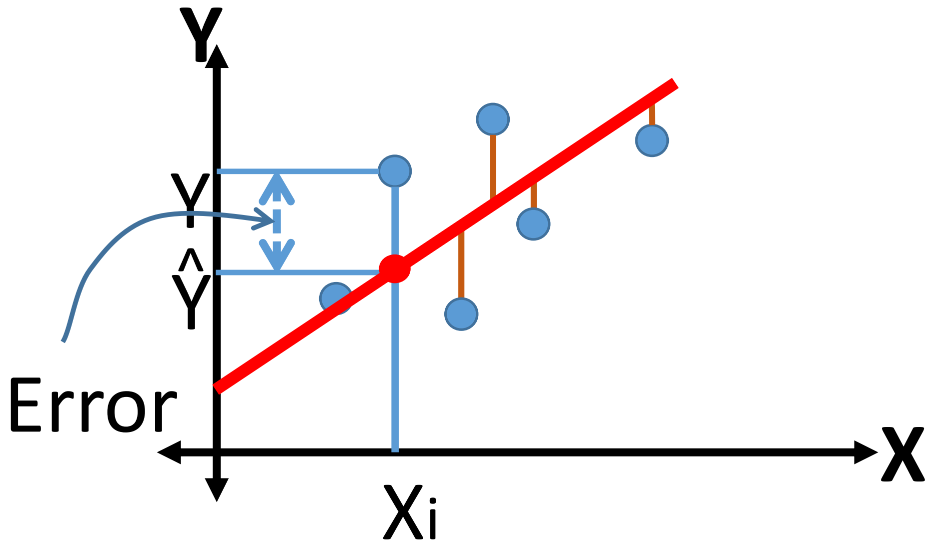
Mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes.

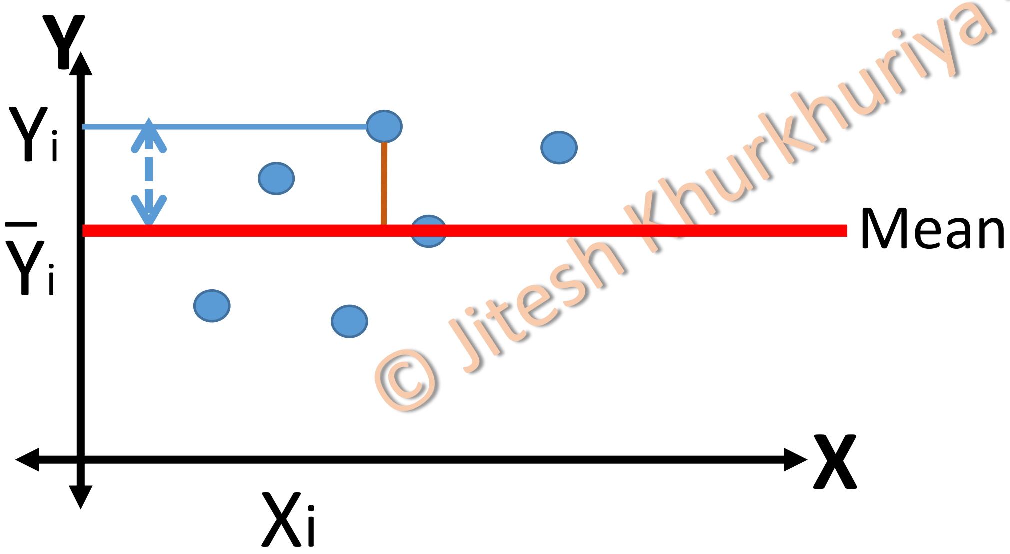
Root Mean Square Error

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

- Very commonly used and makes for an excellent general purpose error metric for numerical predictions.
- Compared to the similar Mean Absolute Error, RMSE amplifies and severely punishes large errors.

Relative Absolute Error



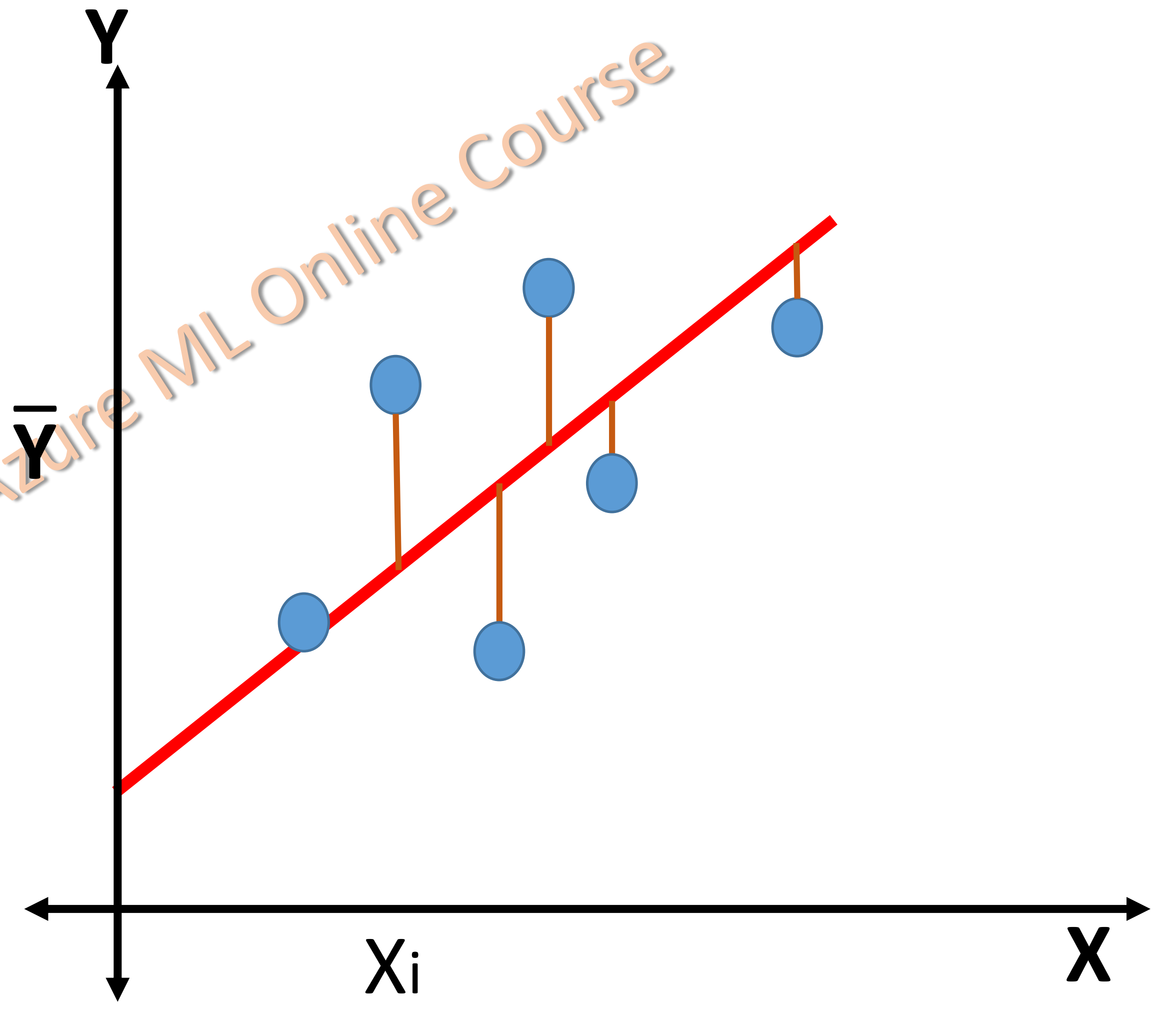


$$RAE = \frac{\sum_{i=1}^{n} |yi - \hat{y}i|}{\sum_{i=1}^{n} |yi - \bar{y}i|}$$

R Squared or Coefficient of Determination

Coefficient of Determination

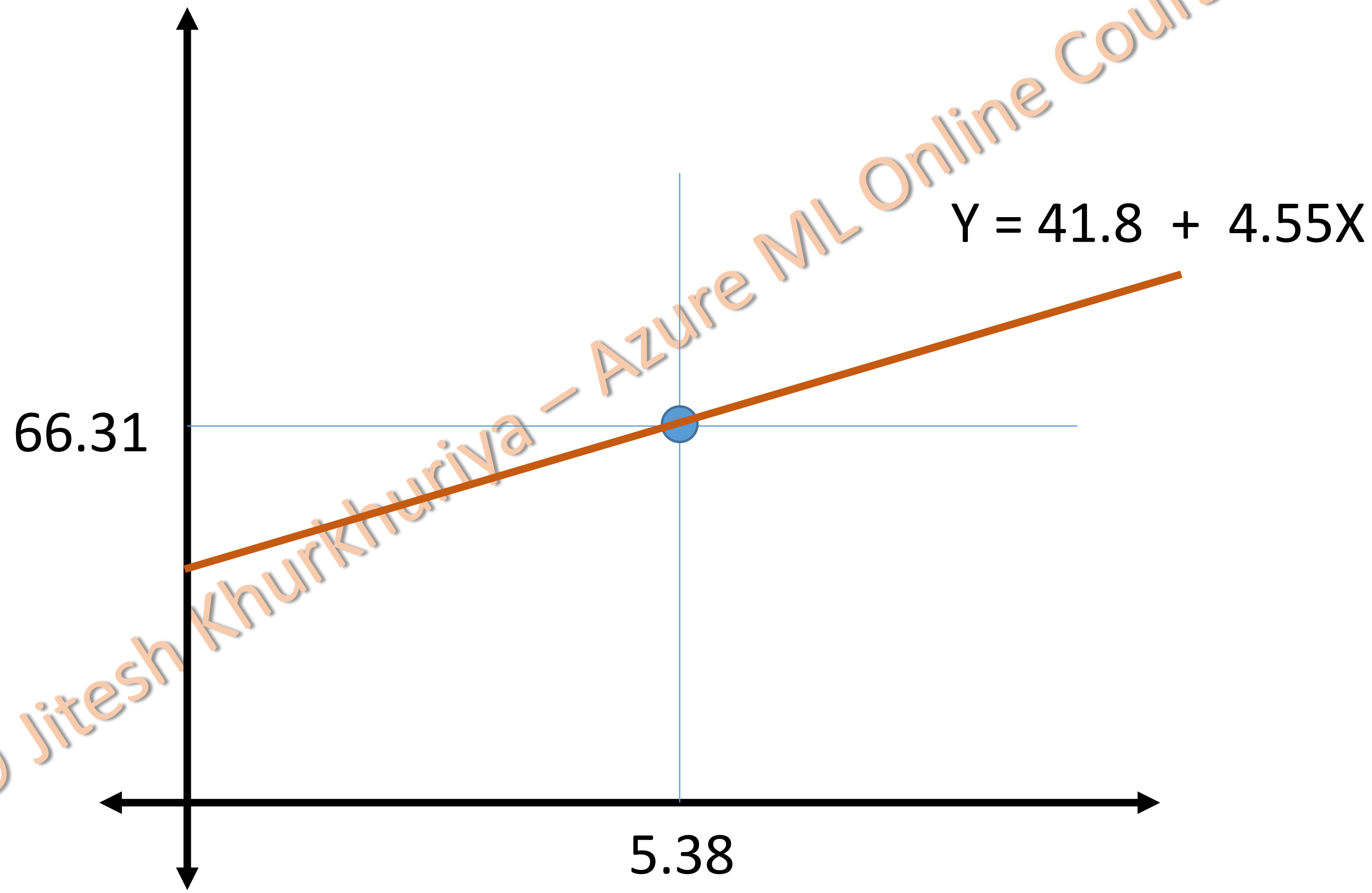
How much (what %) of variation in Y is described by the variation in X?



R-Square With an Example

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	

$Y = b0 + b1X$ $b1 = 4.55$ $b0 = 41.8$



R-Square With an Example

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	

$$Y = 41.8 + 4.55X$$

Predicted Marks $\hat{Y}$
41.80

R-Square With an Example

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	

$Y = 41.8 + 4.55X$

Predicted Marks $\hat{Y}$
41.80
50.90
55.45
60.00
60.00
64.55
69.10
69.10
73.65
73.65
78.20
82.75
82.75

$(Y - \bar{Y})^2$	$(\hat{Y} - \bar{Y})^2$
$(40 - 66.31)^2$	$(41.8 - 66.31)^2$

© Jitesh Khurkhuriya – Azure ML Online Course

R-Square With an Example

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	

$Y = 41.8 + 4.55X$

Predicted Marks $\hat{Y}$
41.80
50.90
55.45
60.00
60.00
64.55
69.10
69.10
73.65
73.65
78.20
82.75
82.75

$(Y - \bar{Y})^2$	$(\hat{Y} - \bar{Y})^2$
692.22	600.74
204.78	237.47
177.16	117.94
127.92	39.82
106.30	39.82
32.38	3.10
22.00	7.78
470.46	7.78
106.30	53.88
59.14	53.88
514.84	141.37
0.48	270.27
514.84	270.27
3028.77	1844.12
SST	SSR

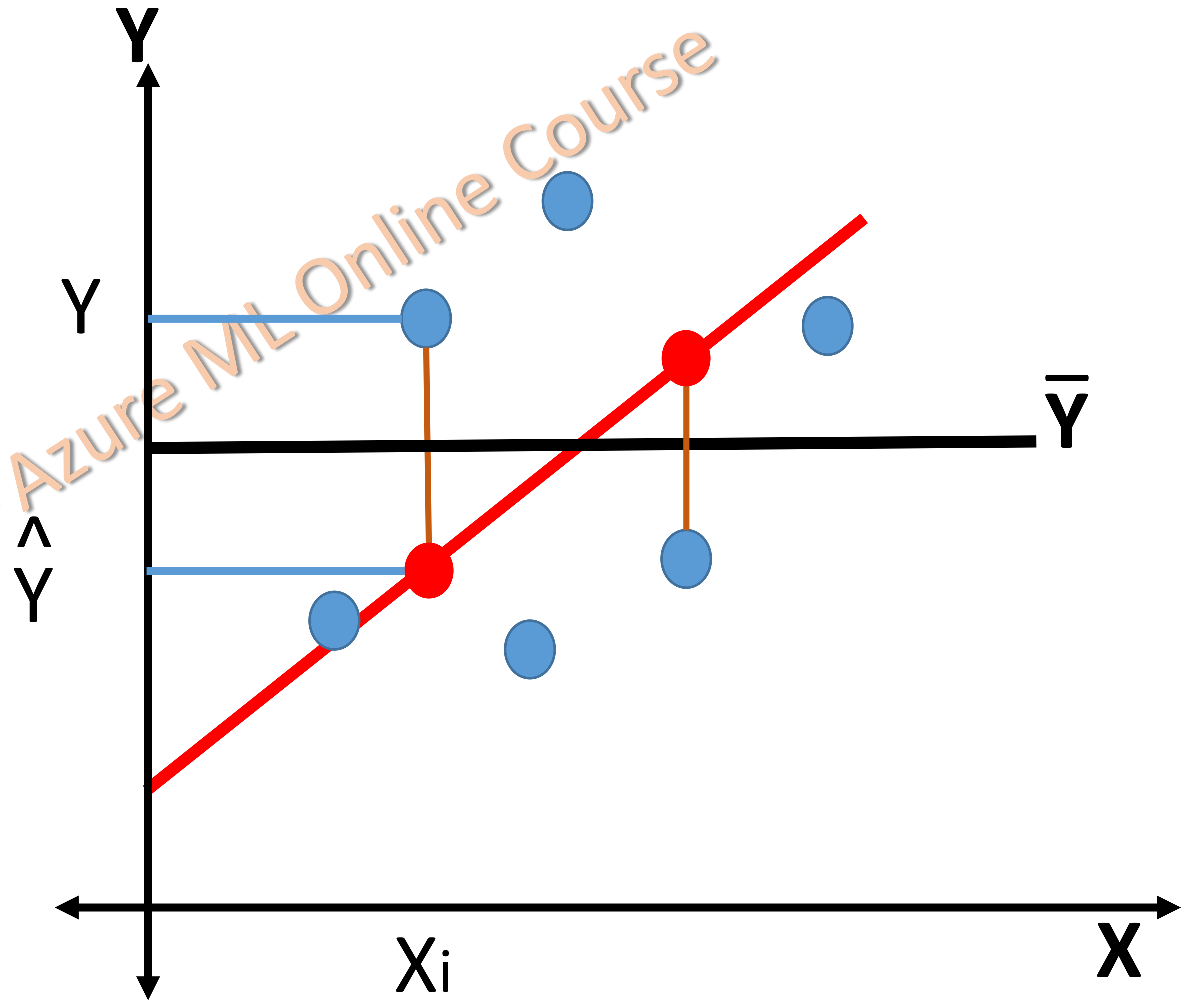
Coefficient of Determination

Sum of Squares Due to Regression

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

Total Sum of Squares

$$SST = \sum_{i=1}^{n} (\boxed{yi} - \bar{yi})^2$$



Coefficient of Determination

$$R^2 = SSR/SST \quad = 1844.12/3028.77$$
$$= 0.60886$$

Higher the value → Variation in Y is explained by variation in X.

Gradient Descent

Hypothesis

"Proposed explanation made on the basis of limited evidence as a starting point for further investigation"

$$h(x) = b0 + b1x$$

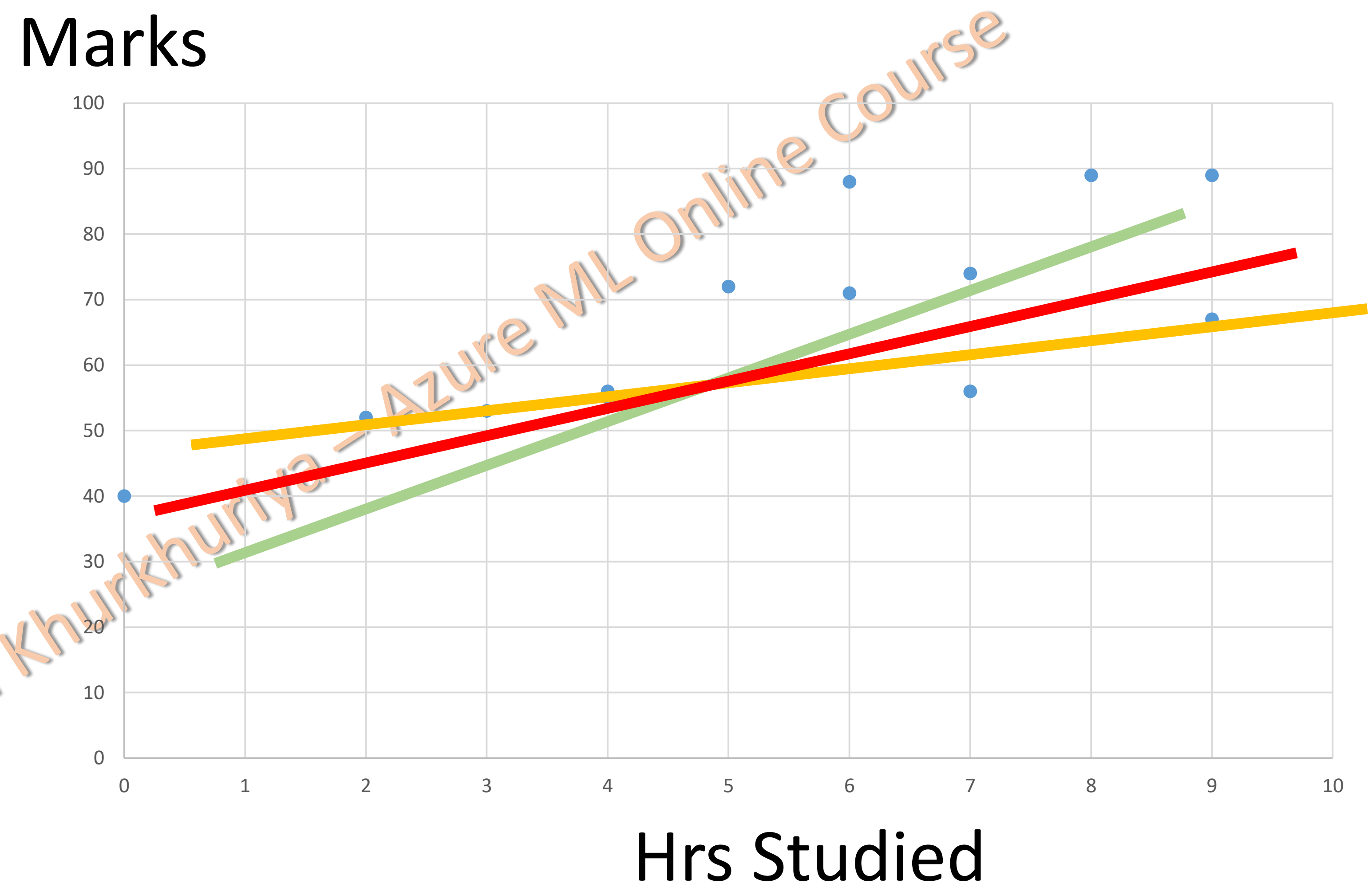
Find out value of b0 and b1 such that

$$Y \sim h(x)$$

for the given observations

Example of Linear Regression

Hrs Studied	Marks
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



Cost Function

Hypothesis: h(x) = b0 + b1x

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

Cost Function

Hypothesis: $h(x) = b0 + b1x$

Hrs Studied	Marks	b0 = 0; b1 = 1 Marks Predicted
0	40	0
2	52	2
3	53	
4	55	
4	56	
5	72	
6	71	
6	88	
7	56	
7	74	
8	89	
9	67	
9	89	

Cost Function

Hypothesis: h(x) = b0 + b1x

Hrs Studied	Marks	b0 = 0; b1 = 1 Marks Predicted
0	40	0
2	52	2
3	53	3
4	55	4
4	56	4
5	72	5
6	71	6
6	88	6
7	56	7
7	74	7
8	89	8
9	67	9
9	89	9

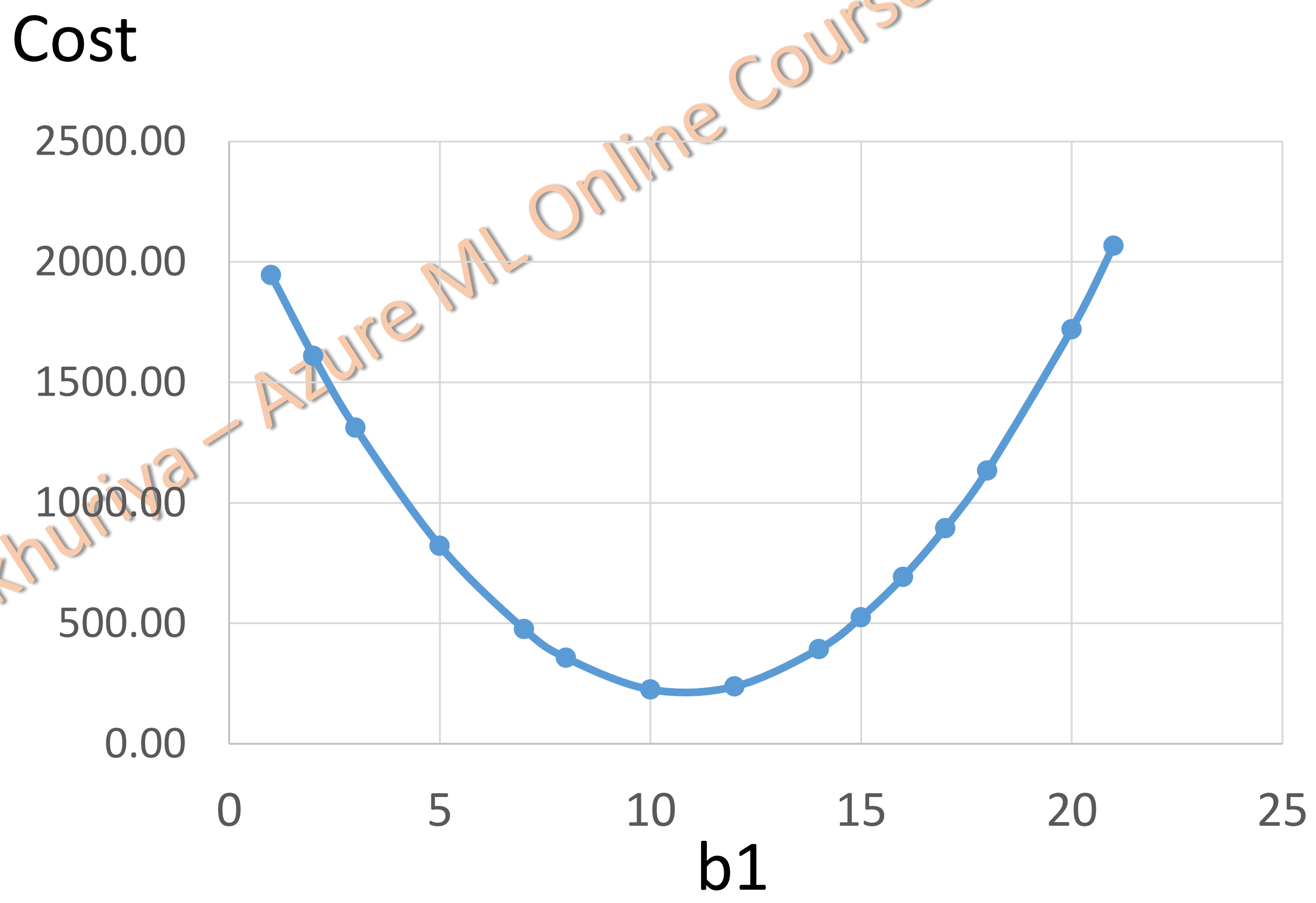
(Yi – Yi)^2
1600
2500
2500
2601
2704
4489
4225
6724
2401
4489
6561
3364
6400

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

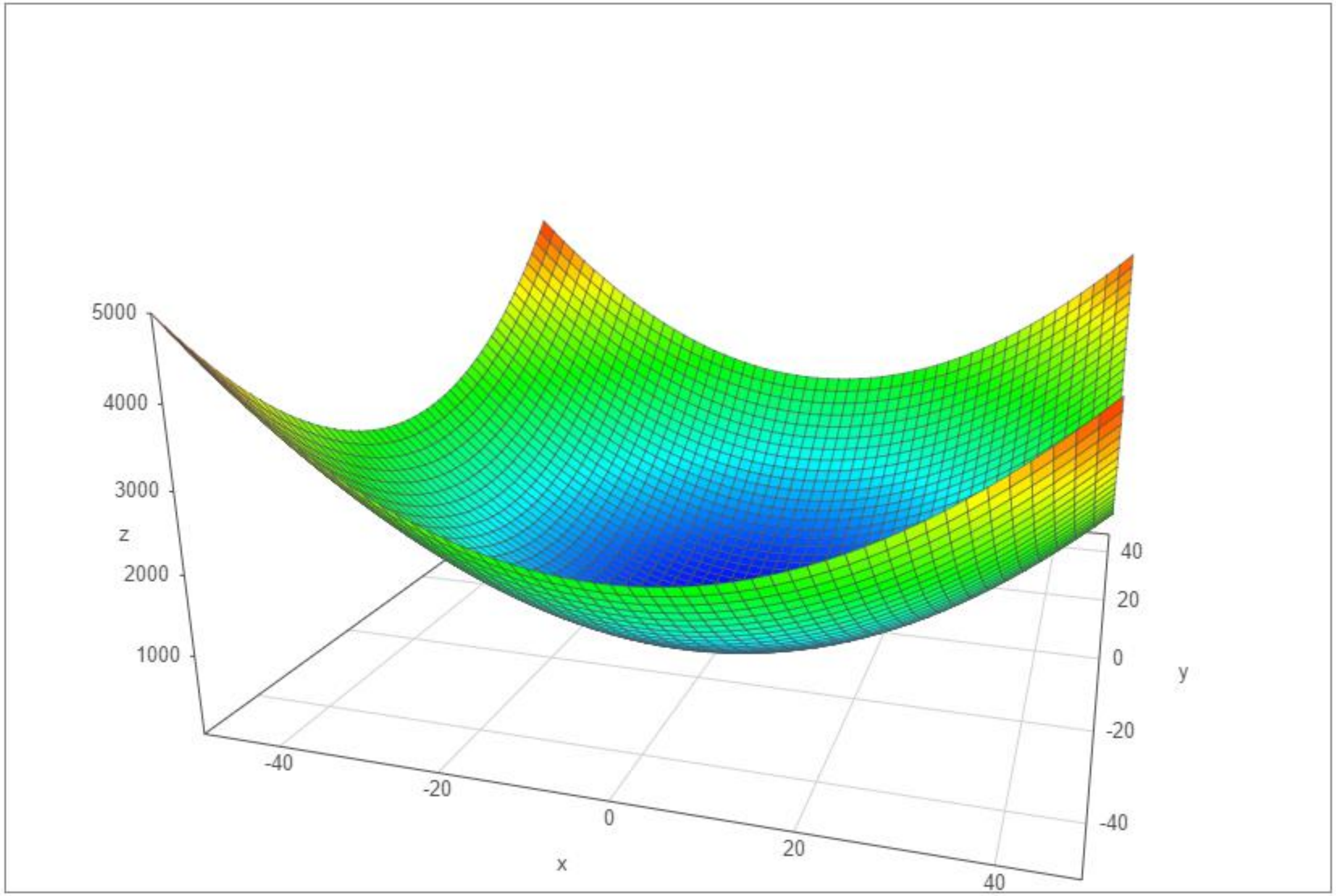
b0	b1	Cost
0	1	1944.538

Cost Function Plot

b0	b1	cost
0	1	1944.54
0	2	1610.08
0	3	1311.46
0	5	821.77
0	7	475.46
0	8	356.08
0	10	224.85
0	12	237.00
0	14	392.54
0	15	524.08
0	16	691.46
0	17	894.69
0	18	1133.77
0	20	1719.46
0	21	2066.08



Cost Function with b0 and b1



X axis – b1

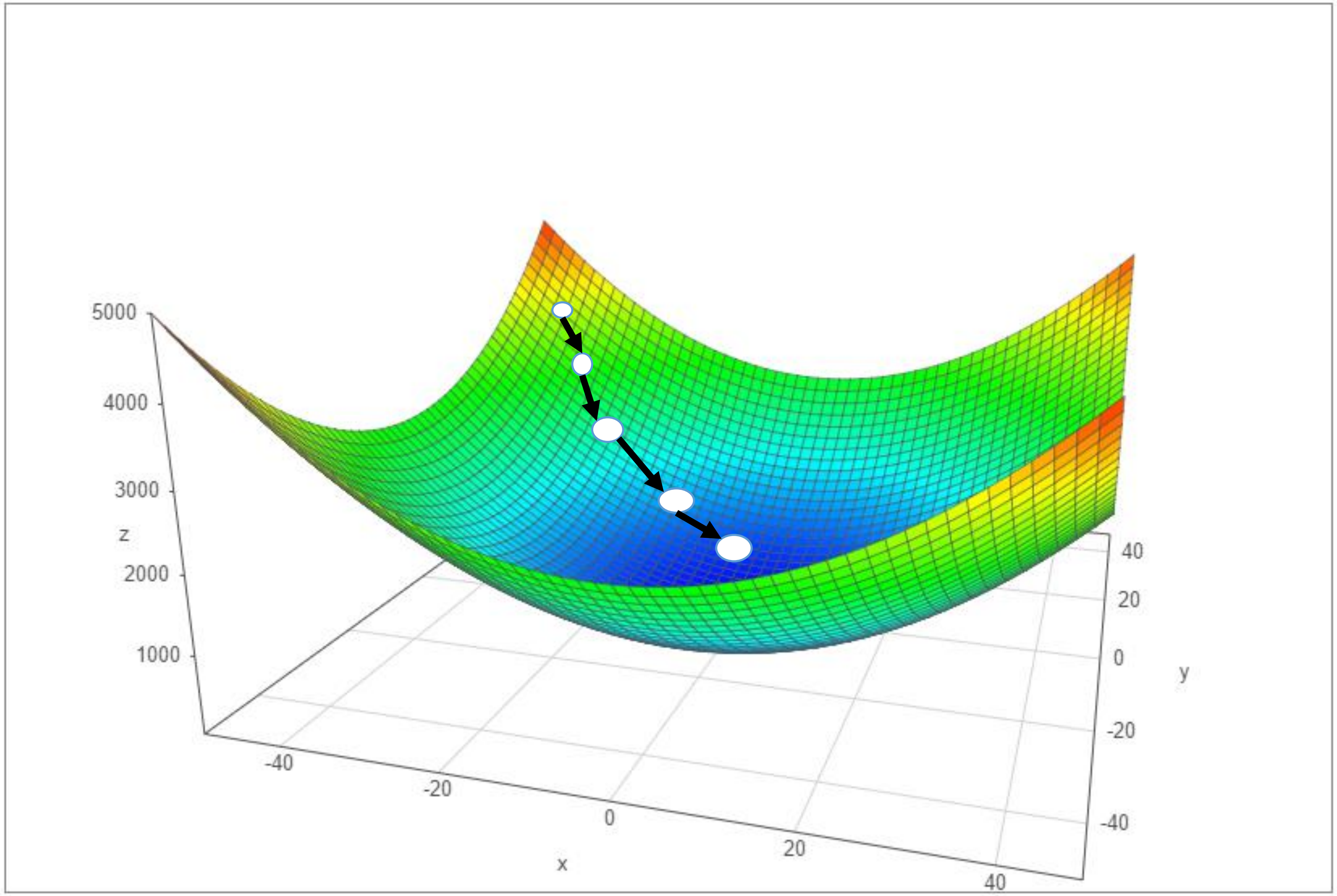
Y axis – b0

Z = C(b0,b1)

https://academo.org/

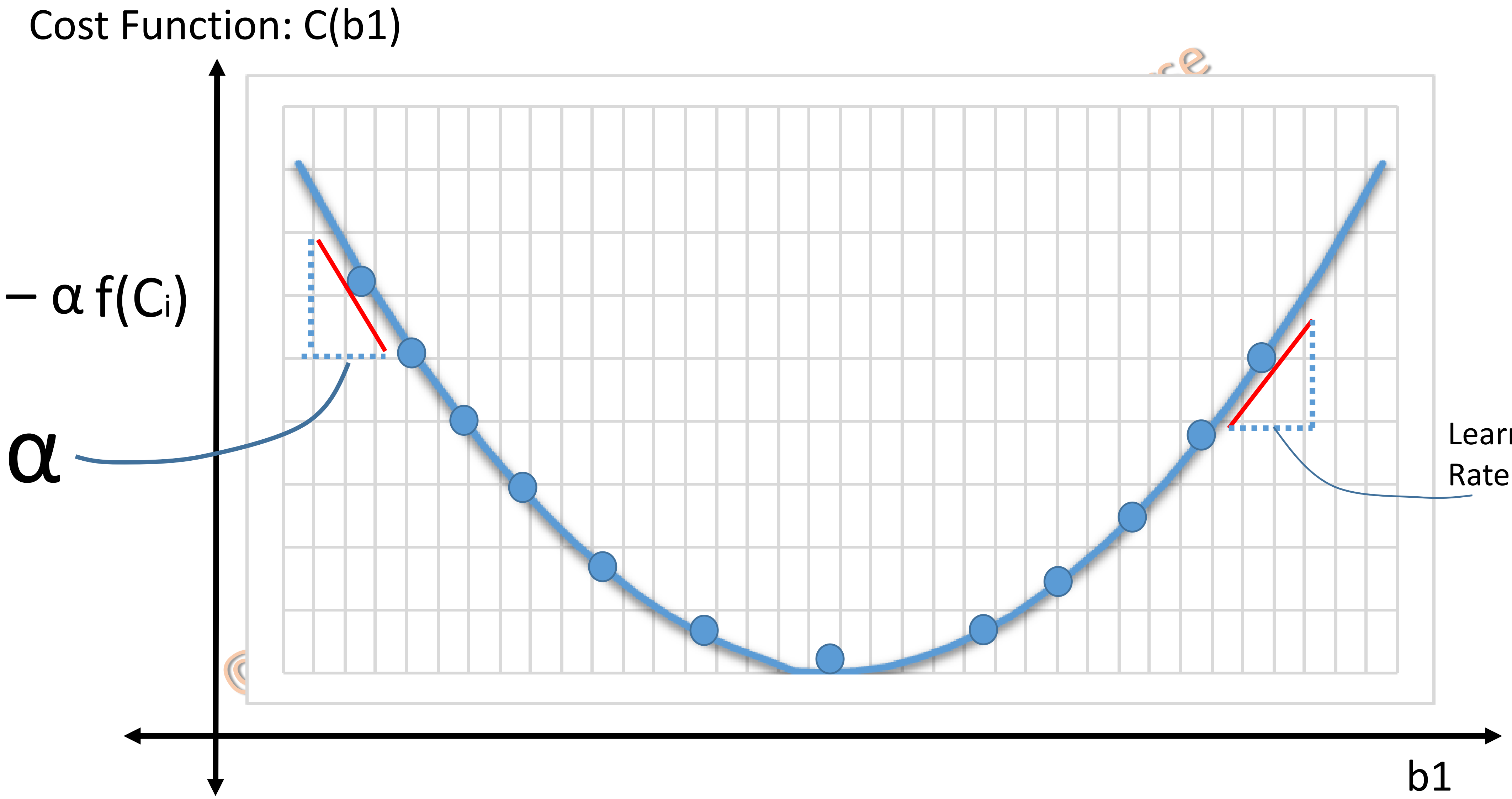
Gradient Descent

Gradient Descent



Gradient Descent

$$b_j := b_j - \alpha f(C_i)$$



Gradient Descent?

Batch Gradient Descent

X1	X2	….	Xn

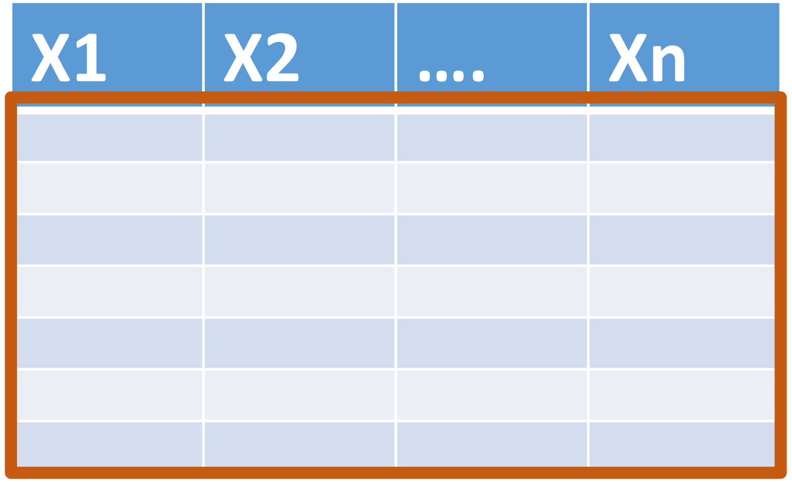
$b_j := b_j - \alpha\ f(C_i)$

Sum of All before taking one step (epoch)
Long time to reach the bottom

Does it for
number of examples
number of features
learning rate

Batch Gradient Descent

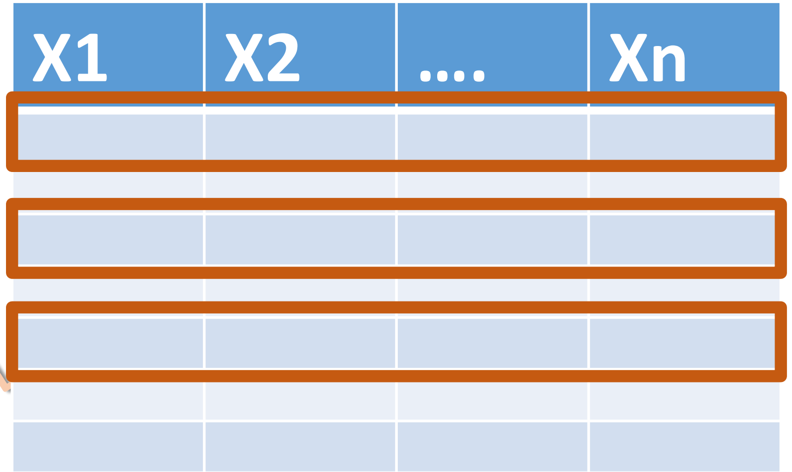
Batch Vs Stochastic Gradient Descent



$$b_j := b_j - \alpha f(C_i)$$

Does it for
- number of examples
- number of features
- learning rate

Batch Gradient Descent

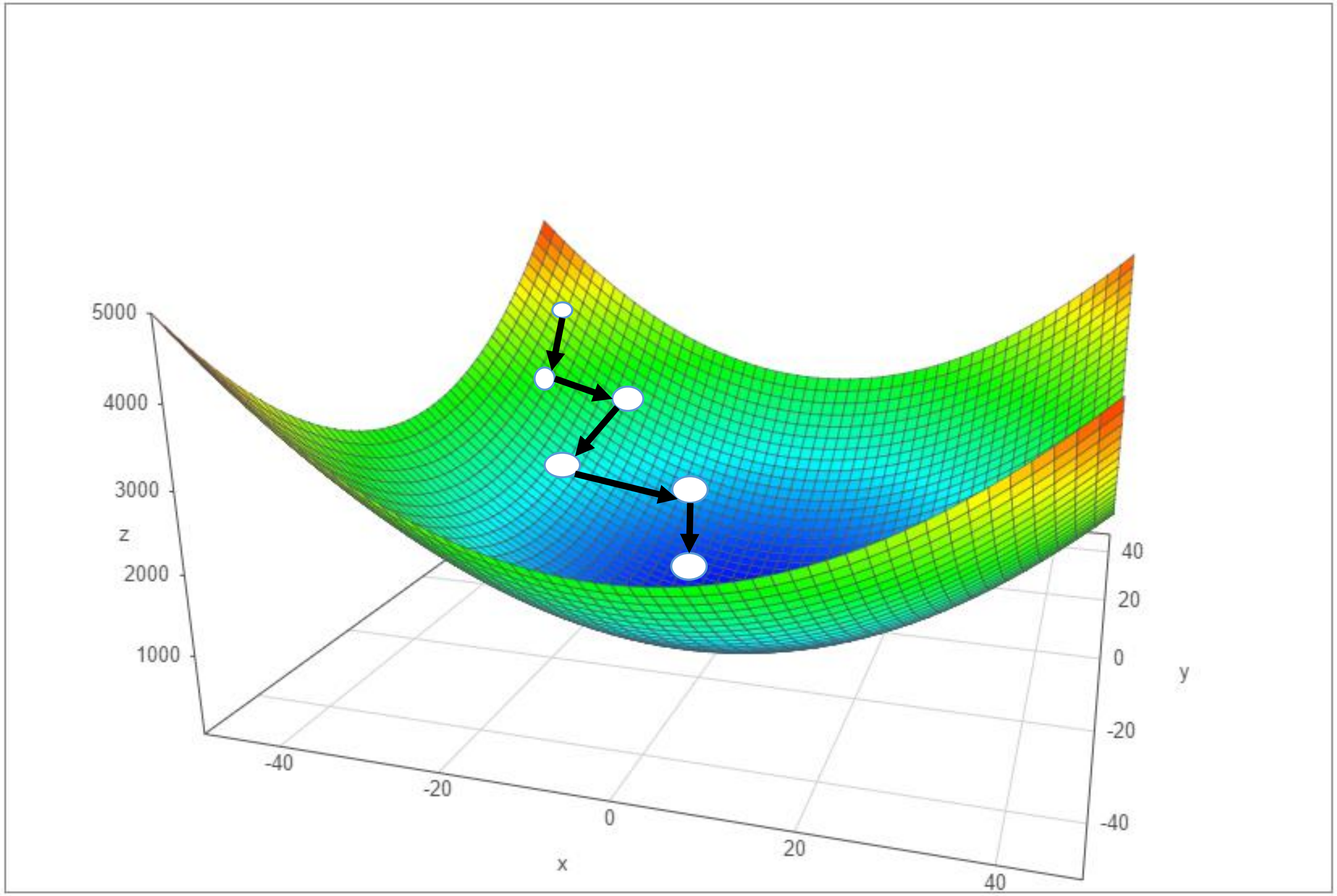


Randomly shuffle the dataset

Repeat the steps for every example

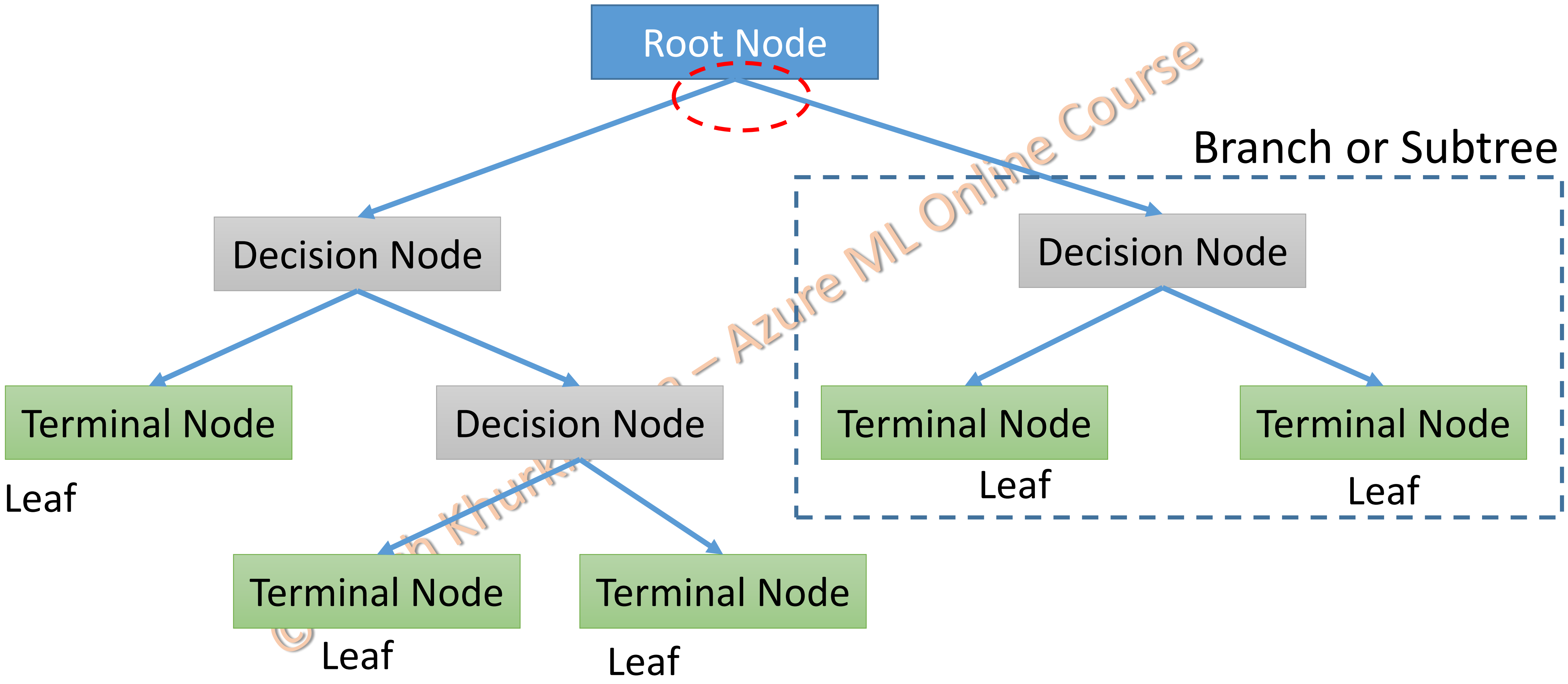
Modify the coefficient at every step

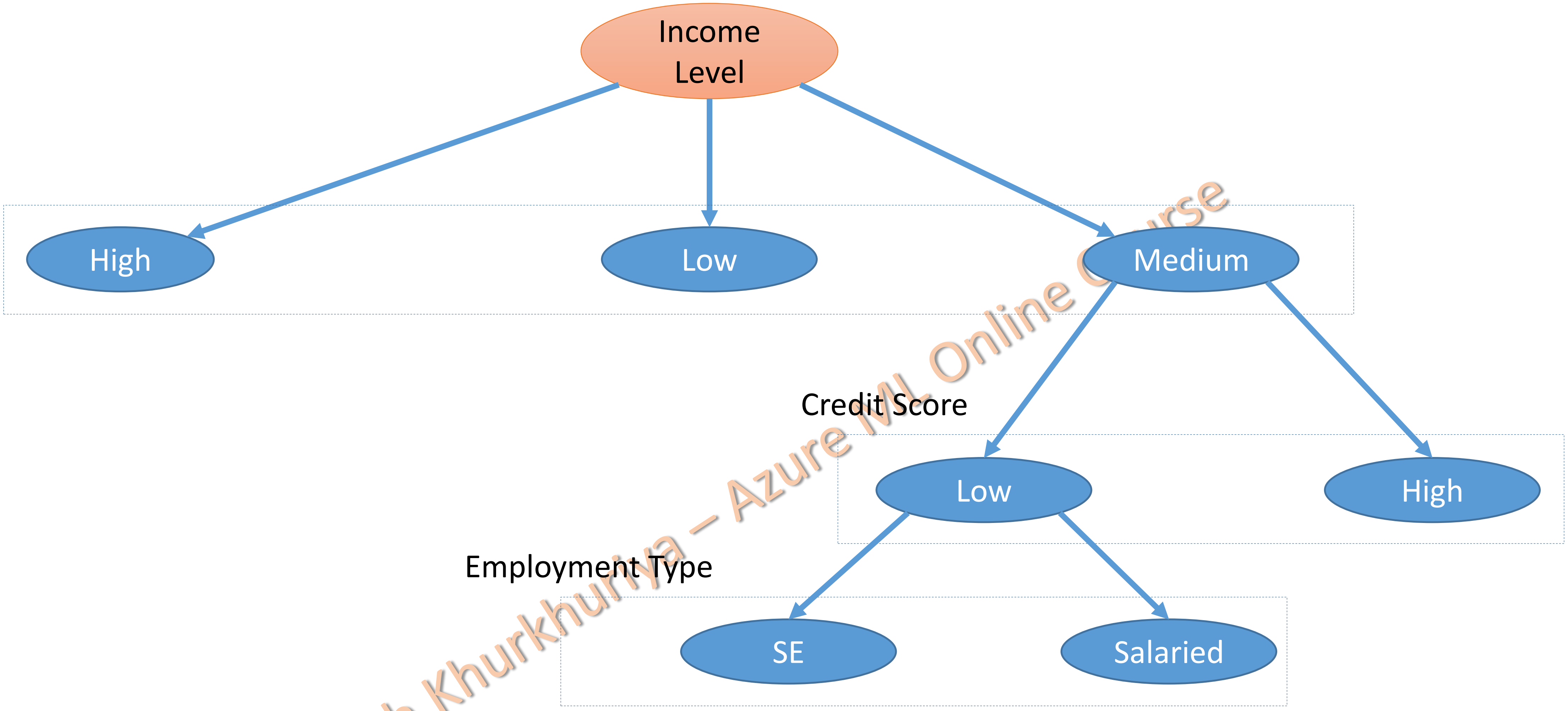
Stochastic Gradient Descent



Decision Tree Regression

Decision Tree Terms





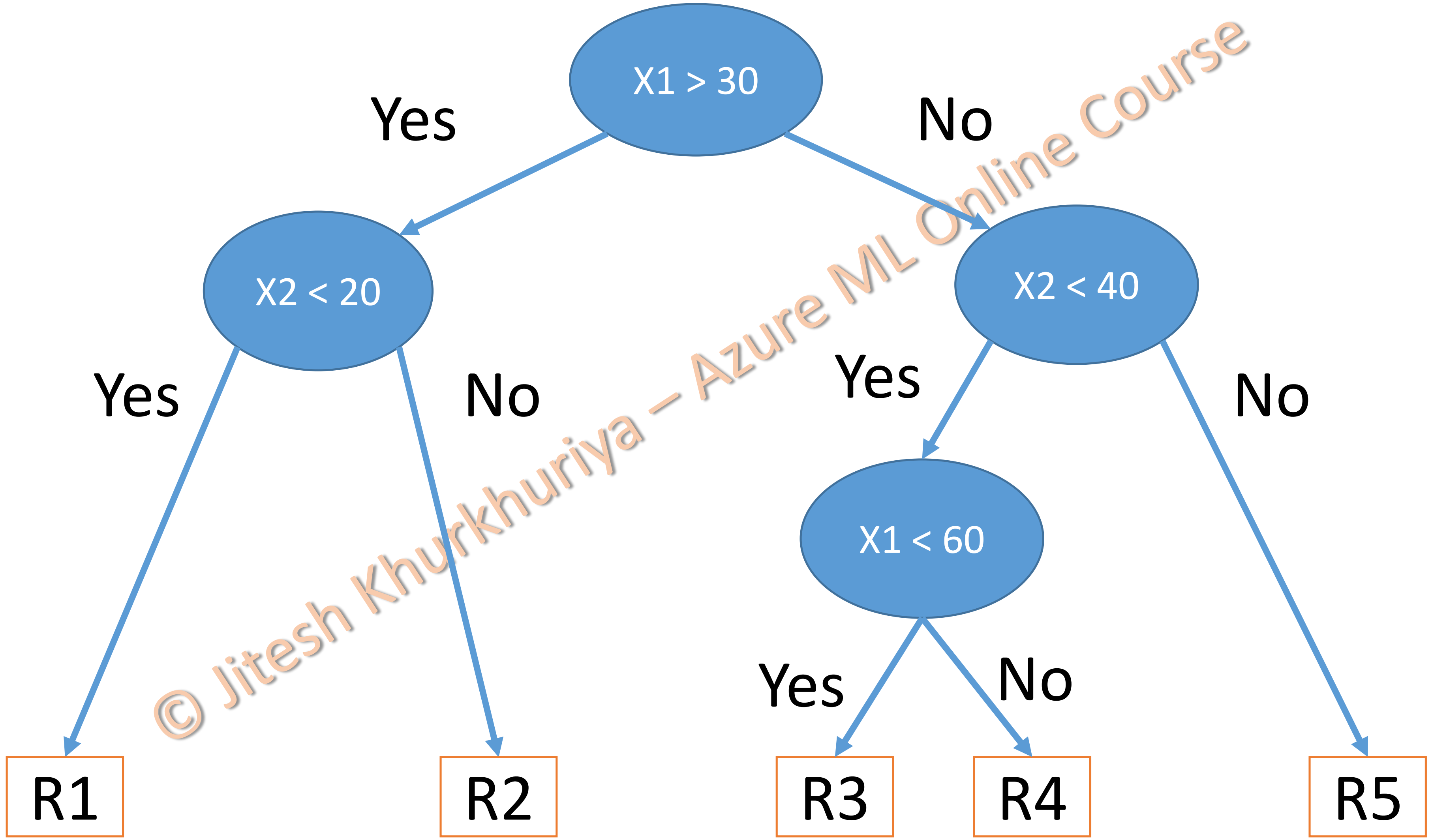
LID	IL	CS	ET	Status
L1	Medium	Low	SE	No
L8	Medium	Low	SE	No

Pure Subset

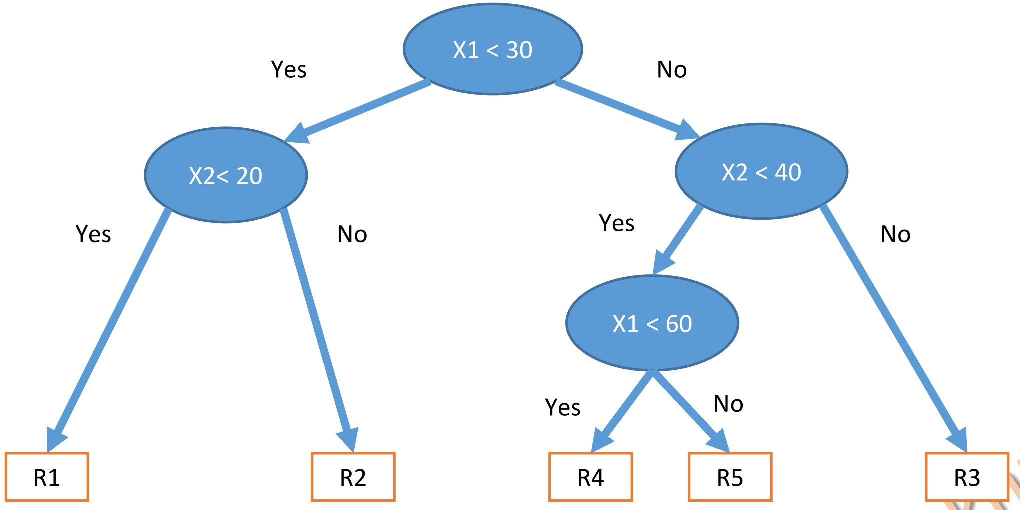
LID	IL	CS	ET	Status
L4	Medium	Low	Salaried	Yes

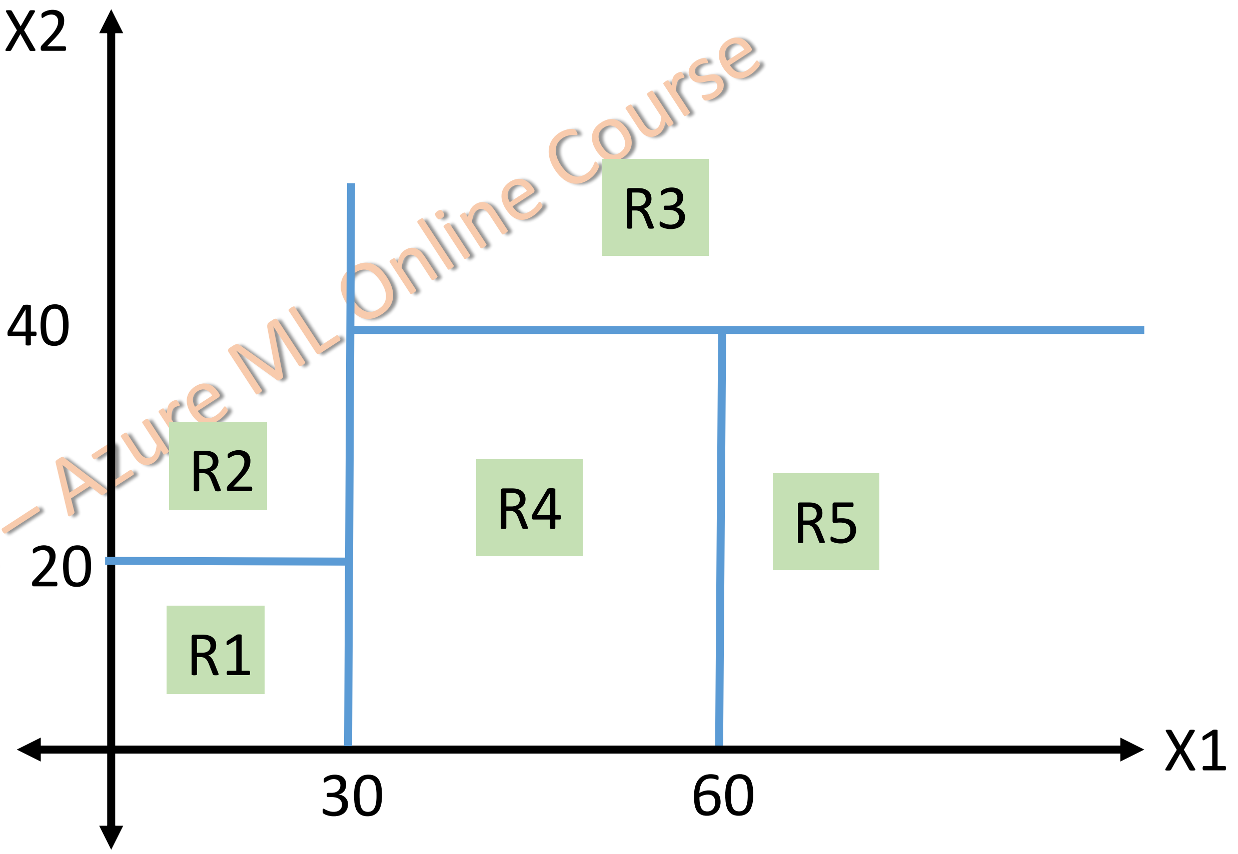
Pure Subset

Decision Tree Regression

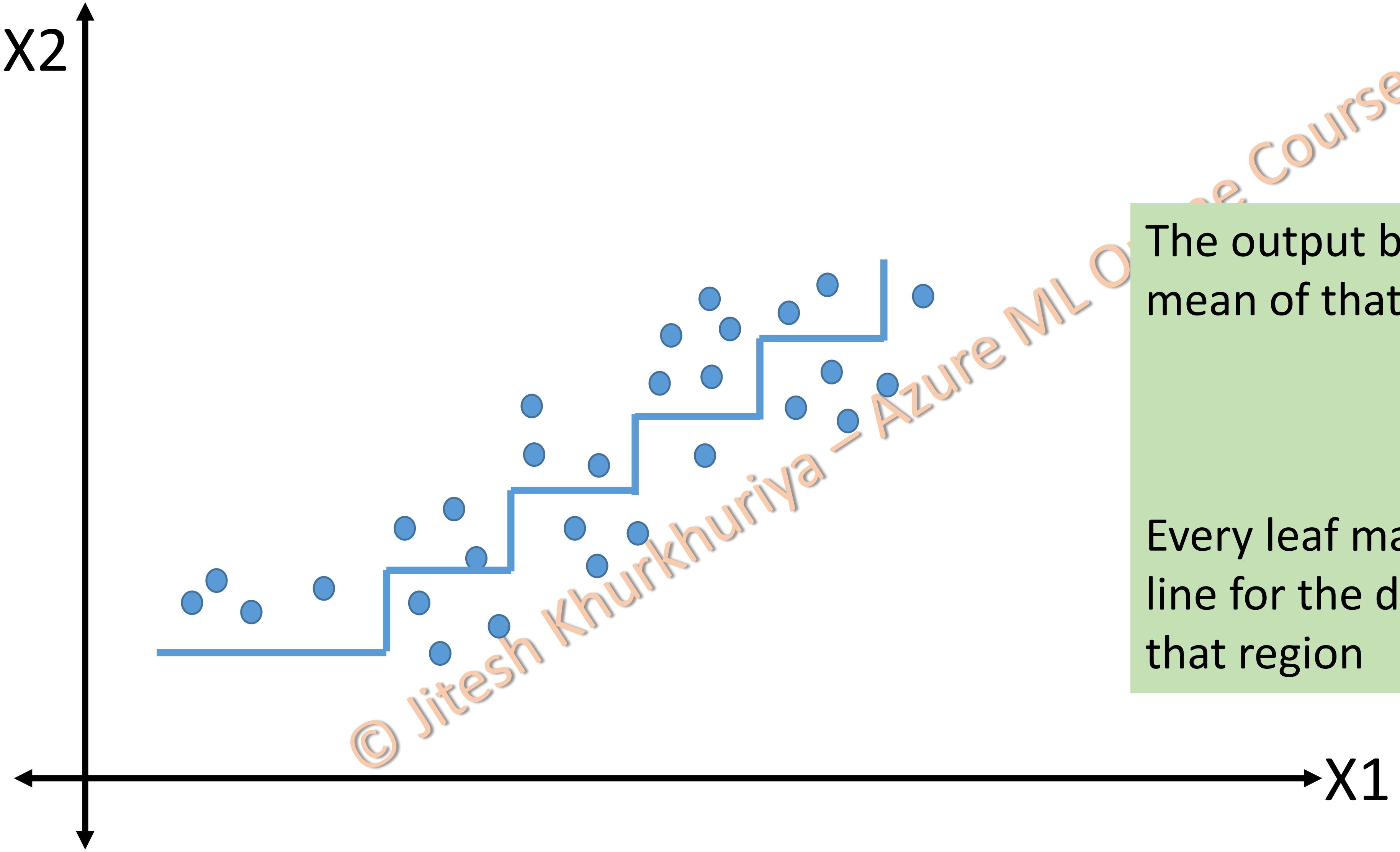


Decision Tree Regression





Decision Tree Regression



The output by every region is the mean of that region

Or

Every leaf may have a regression line for the data points within that region

Boosted Decision Tree Regression

- MART gradient boosting algorithm.
- Builds each regression tree in a step-wise fashion
- Predefined loss function to measure the error in each step

Thank You and Have a Great Time…!