Machine Learning

Supervised Learning: Linear Regression

Regression:

Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).

Regression

- Linear Regression
- Logistic Regression

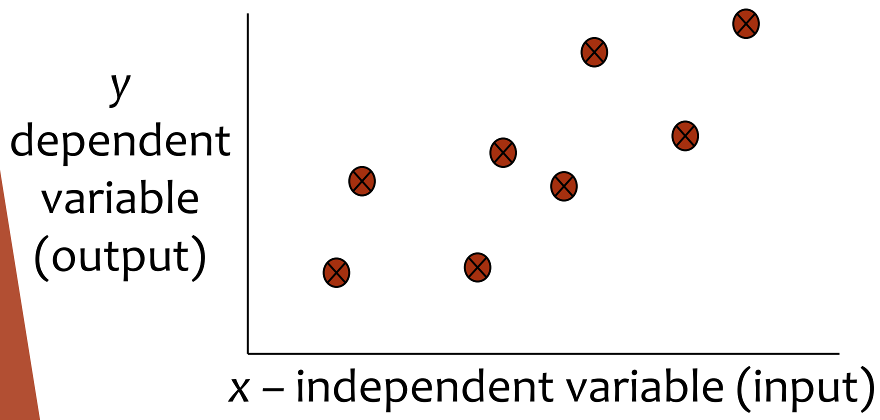
Simple Linear Regression

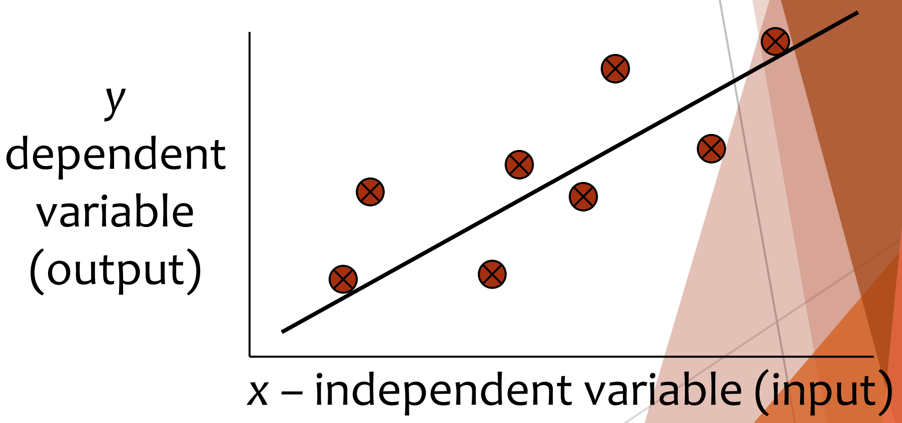
SLR is a statistical method that allows us to summarize and study the relationship between two continuous(quantitative) variables.

1. **The first variable denoted by x, is regarded as the predictor, explanatory, or independent variable.**
2. **The second variable , denoted by y, is regarded as the response, outcome, or dependent variable.**

Regression

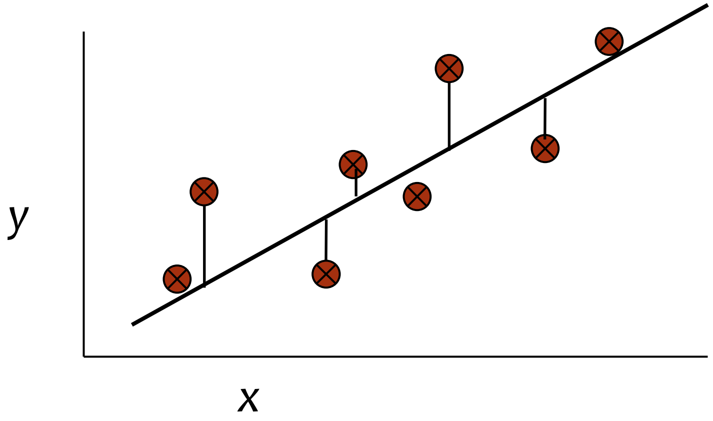
- For classification the output(s) is nominal
- In regression the output is continuous
  - Function Approximation
- Many models could be used – Simplest is linear regression
  - Fit data with the best hyper-plane which "goes through" the points





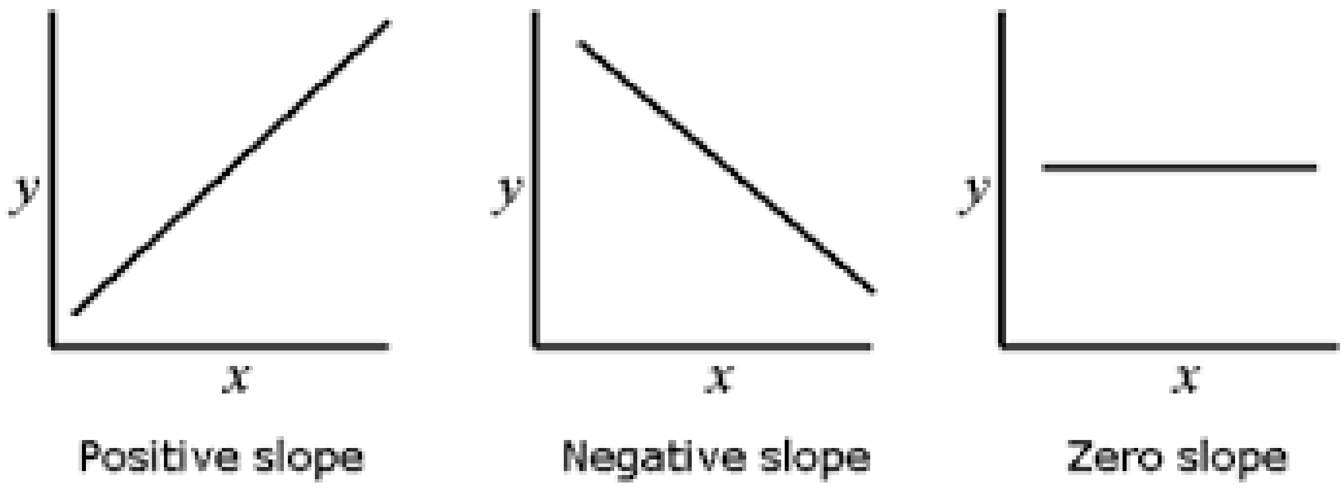
Regression

- For classification the output(s) is nominal
- In regression the output is continuous
  - Function Approximation
- Many models could be used – Simplest is linear regression
  - Fit data with the best hyper-plane which "goes through" the points
  - For each point the differences between the predicted point and the actual observation is the *residue*



Regression Line

Linear regression consists of finding the best-fitting straight line through the points. The best-fitting line is called a *regression line*.



Linear Regression Using Least Squares

Regression Line : y = c+ mx

y = b0 + b1 x

$$m = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$c = \bar{y} - m\bar{x}$$

Here, m and c can also be denoted as b1 and b0.

So, y = b0 + b1 x

After computing b0 and b1, we can find the new value for ypred for any given x.

Evaluation of Model Estimators

1. **Karl Pearson's Coefficient of Correlation**
2. **R-Square**
3. **Standard Error of the Estimate**

Evaluation of Model Estimators

1. **Karl Pearson's Coefficient of Correlation**

- The Karl Pearson's correlation coefficient (or simply, the Pearson's correlation coefficient) is a measure of the strength of a linear association between two variables and is denoted by *r* or rxy(x and y being the two variables involved).
- Here, x and y are variables and N is the no. of instances we have to compute the coefficient.

Correlation Coefficient Formula

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

Continue….

- **The value of *r* always lies between +1 and -1.** Depending on its exact value, we see the following degrees of association between the variables-

R value variation

Association	Negative	Positive
Weak	-0.1 to -0.3	0.1 to 0.3
Average	-0.3 to -0.5	0.3 to 0.5
Strong	-0.5 to -1.0	0.5 to 1.0

- A value greater than 0 indicates a positive association i.e. as the value of one variable increases, so does the value of the other variable. A value less than 0 indicates a negative association i.e. as the value of one variable increases, the value of the other variable decreases.

Evaluation of Model Estimators

2. R-Square

- **R-squared** is a statistical measure of how close the data are to the fitted **regression** line. It is also known as the coefficient of determination.
- R-squared measures the strength of the relationship between your model and the dependent variable on a convenient 0 – 100% scale.
- High value of r-Squared indicates a strong linear relationship.

$$\boldsymbol{R^2 = \frac{\sum(ypred - ymean)^2}{\sum(y - ymean)^2}}$$

Evaluation of Model Estimators

3. Standard Error of the Estimate

- The **standard error of the estimate** is a measure of the accuracy of predictions.
- It is used to check the accuracy of predictions made with the regression line.

$$\boldsymbol{Standard\ Error\ of\ the\ Estimate} = \sqrt{\frac{\sum(\boldsymbol{ypred} - \boldsymbol{y})^2}{\boldsymbol{n} - 2}}$$

Supervised Learning: Linear Regression Example

Example :

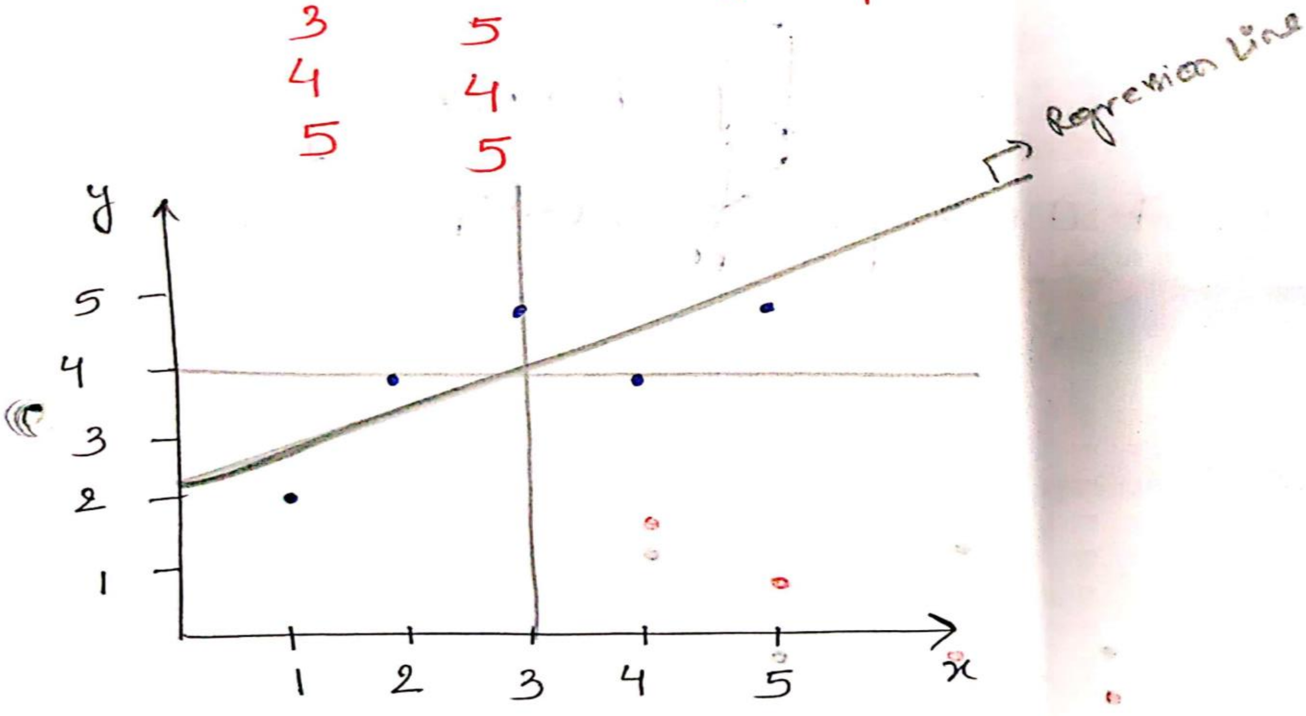
Example — Least Square Method

Q:-

x	y
1	2
2	4
3	5
4	4
5	5

x independent Variable

y dependent Variable



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

Independent Variable x	Dependent Variable y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
1	2	-2	-2	4	4	4
2	4	-1	0	1	0	0
3	5	0	1	0	1	0
4	4	1	0	1	0	0
5	5	2	1	4	1	2
$\bar{x} = 3$	$\bar{y} = 4$			10		6

$$b_1 = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{6}{10} = 0.6$$

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

$$4 = b_0 + 0.6(3)$$

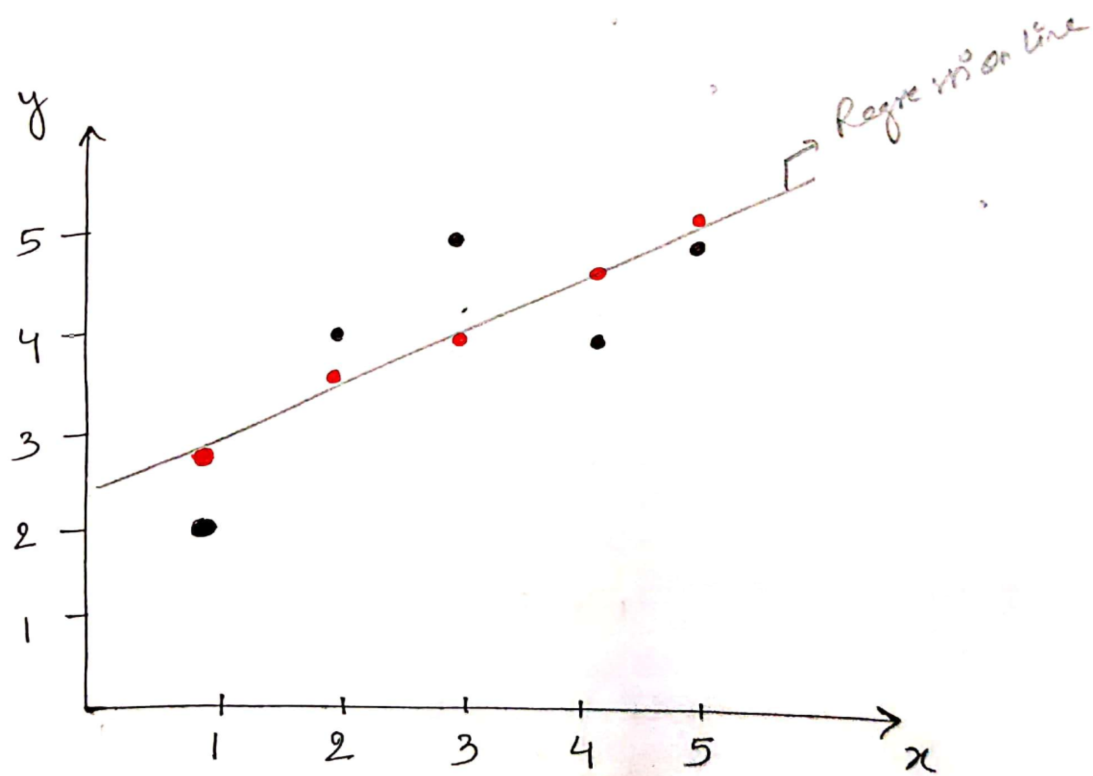
$$b_0 = 4 - 1.8$$

$$= 2.2$$

$$\boxed{\hat{y} = 2.2 + 0.6x}$$

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

Predicted y.



ypred = 2.2 + 0.6 x

R SQUARED

x	y	$y-\bar{y}$	$(y-\bar{y})^2$	Predicted y $\hat{y}$	$\hat{y}-\bar{y}$	$(\hat{y}-\bar{y})^2$
1	2	-2	4	2.8	-1.2	1.44
2	4	0	0	3.4	~~3.4-4~~ -0.6	0.36
3	5	1	1	4	0	0
4	4	0	0	4.6	0.6	0.36
5	5	1	1	5.2	1.2	1.44
			6		0	3.60

$$R^2 = \frac{\Sigma(\hat{y}-\bar{y})^2}{\Sigma(y-\bar{y})^2} = \frac{3.6}{6} = 0.6$$

R^2 Coefficient for multiple determination.

Coefficient Correlation

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

x	y	x^2	y^2	xy
1	2	1	4	2
2	4	4	16	8
3	5	9	25	15
4	4	16	16	16
5	5	25	25	25
$\Sigma x = 15$	$\Sigma y = 20$	$\Sigma x^2 = 55$	$\Sigma y^2 = 86$	$\Sigma xy = 66$

$$r = \frac{(5 \times 66) - (15 \times 20)}{\sqrt{(5 \times 55 - 225)(5 \times 86 - 400)}}$$

$$= \frac{330 - 300}{\sqrt{(275 - 225)(430 - 400)}}$$

$$= \frac{30}{\sqrt{50 \times 30}} = \frac{30}{38.729} = 0.9762$$

Standard Error of the Estimate

Estimated values are compared to the actual value. Distance between estimated & actual is error. We have to minimize the error.

$$\text{Standard Error of the Estimate} = \sqrt{\frac{\Sigma(\hat{y}-y)^2}{n-2}}$$

x	y	$\hat{y}$	$\hat{y}-y$	$(\hat{y}-y)^2$
1	2	2.8	0.8	0.64
2	4	3.4	-0.6	0.36
3	5	4	-1	1
4	4	4.6	0.6	0.36
5	5	5.2	0.2	0.04
				2.4

Standard Error of the estimate $= \sqrt{\dfrac{\Sigma(\hat{y}-y)^2}{n-2}}$

$$= \sqrt{\frac{2.4}{5-2}} = \sqrt{\frac{2.4}{3}}$$

$$= \sqrt{0.8} = 0.89$$

Example 2: Create the relationship model for the given dataset to find the relation between height and weight parameters. Predict Y for X=154,161,178

Sr No.	Height(X)	Weight(y)
1	151	63
2	174	81
3	138	56
4	186	91
5	128	47
6	136	57
7	179	76
8	163	72
9	152	62
10	131	48

Coefficient Computation

Sr No.	Height(X)	Weight(Y)	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$
1	151	63	-2.8	-2.3	6.44	7.84
2	174	81	20.2	15.7	317.14	408.04
3	138	56	-15.8	-9.3	146.94	249.64
4	186	91	32.2	25.7	827.54	1036.84
5	128	47	-25.8	-18.3	472.14	665.64
6	136	57	-17.8	-8.3	147.74	316.84
7	179	76	25.2	10.7	269.64	635.04
8	163	72	9.2	6.7	61.64	84.64
9	152	62	-1.8	-3.3	5.94	3.24
10	131	48	-22.8	-17.3	394.44	519.84
	$\bar{X}$ **153.8**	$\bar{Y}$ **65.3**			**2649.6**	**3927.6**

b1=	**0.67461**
b0=	**-38.4535**

Regression Line: **y = -38.45348 + 0.674 x**

Karl Pearson coefficient

Sr No.	Height(X)	Weight(Y)	X^2	Y^2	XY
1	151	63	22801	3969	9513
2	174	81	30276	6561	14094
3	138	56	19044	3136	7728
4	186	91	34596	8281	16926
5	128	47	16384	2209	6016
6	136	57	18496	3249	7752
7	179	76	32041	5776	13604
8	163	72	26569	5184	11736
9	152	62	23104	3844	9424
10	131	48	17161	2304	6288
	1538	**653**	**240472**	**44513**	**103081**

r=	**0.97713**

Standard Error of Estimate

Sr No.	Height(X)	Weight(Y)	$\hat{Y}$	$\hat{Y} - Y$	$(\hat{Y} - Y)^2$
1	151	63	63.3205	0.3205	0.10272025
2	174	81	78.8225	-2.1775	4.74150625
3	138	56	54.5585	-1.4415	2.07792225
4	186	91	86.9105	-4.0895	16.72401025
5	128	47	47.8185	0.8185	0.66994225
6	136	57	53.2105	-3.7895	14.36031025
7	179	76	82.1925	6.1925	38.34705625
8	163	72	71.4085	-0.5915	0.34987225
9	152	62	63.9945	1.9945	3.97803025
10	131	48	49.8405	1.8405	3.38744025
				-0.923	**84.73881**

Standard Error of Estimate = 3.25459

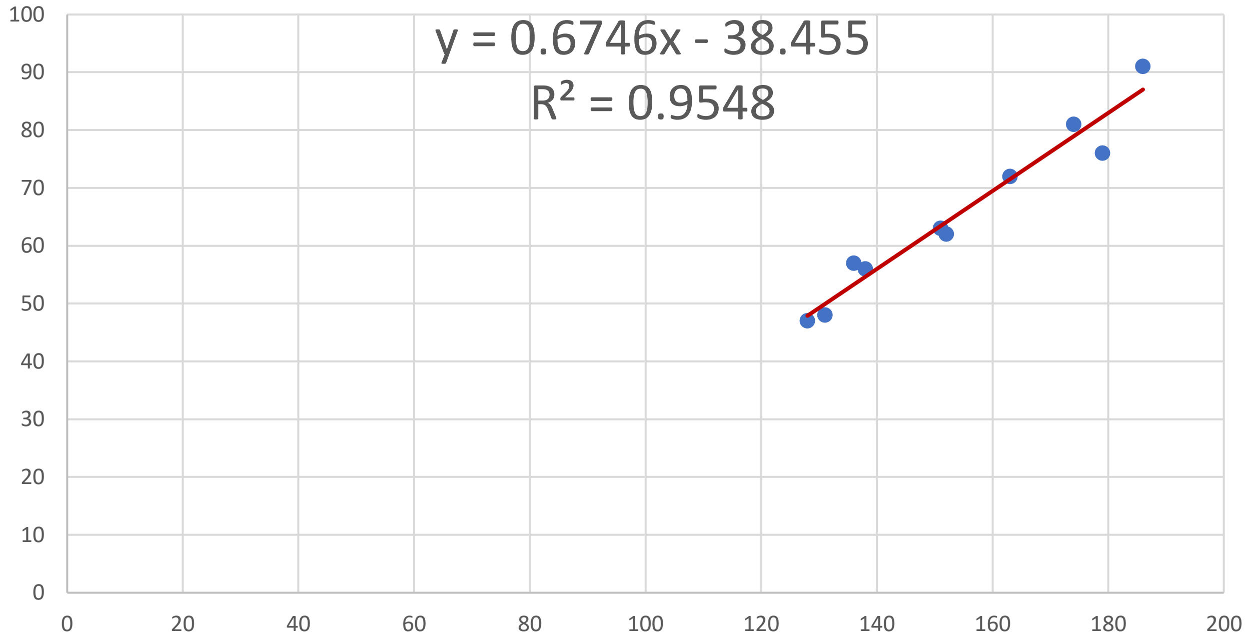
R Square

R-SQUARE

Sr No.	Height(X)	Weight(Y)	$\hat{Y}$	$\hat{Y}-Y$	$(\hat{Y}-\bar{Y})^2$	$(Y-\bar{Y})$	$(Y-\bar{Y})(Y-\bar{Y})$
1	151	63	63.3205	-1.9795	3.91842025	-2.3	5.29
2	174	81	78.8225	13.5225	182.8580063	15.7	246.49
3	138	56	54.5585	-10.7415	115.3798223	-9.3	86.49
4	186	91	86.9105	21.6105	467.0137103	25.7	660.49
5	128	47	47.8185	-17.4815	305.6028423	-18.3	334.89
6	136	57	53.2105	-12.0895	146.1560103	-8.3	68.89
7	179	76	82.1925	16.8925	285.3565563	10.7	114.49
8	163	72	71.4085	6.1085	37.31377225	6.7	44.89
9	152	62	63.9945	-1.3055	1.70433025	-3.3	10.89
10	131	48	49.8405	-15.4595	238.9961403	-17.3	299.29
				-0.923	**1784.2996**		**1872.1**

R-SQUARE= **0.9531006**

Plotting of Independent and Dependent Variable



X	Y
154	65.34252
161	70.06052
178	81.51852

Example 3: Create the relationship model for the given dataset to find the relation between x and y parameters. Predict the value of Y for X = 24,13,32.

Sr No.	X	Y
1	17	94
2	13	73
3	12	59
4	15	80
5	16	93
6	14	85
7	16	66
8	16	79
9	18	77
10	19	91

Example 4: Create the relationship model for the given dataset to find the relation between x and y parameters. Predict the value of Y for X = 68,75,89.

Sr No.	X	Y
1	65	105
2	65	125
3	62	110
4	67	120
5	69	140
6	65	135
7	61	95
8	67	130

LOGISTIC REGRESSION

Logistic Regression

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary).

Like all regression analyses, the logistic regression is a predictive analysis.

Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

Use of Logistic Regression

There are many important topics for which the dependent variable is "limited."

For example:

a) whether or not a mail is spam,

b) tumor is malignant or benign

c) student takes ML as a course or not.

d) How does the probability of getting lung cancer (yes vs. no) change for every additional pound a person is overweight and for every pack of cigarettes smoked per day?

e)Do body weight, calorie intake, fat intake, and age have an influence on the probability of having a heart attack (yes vs. no)?

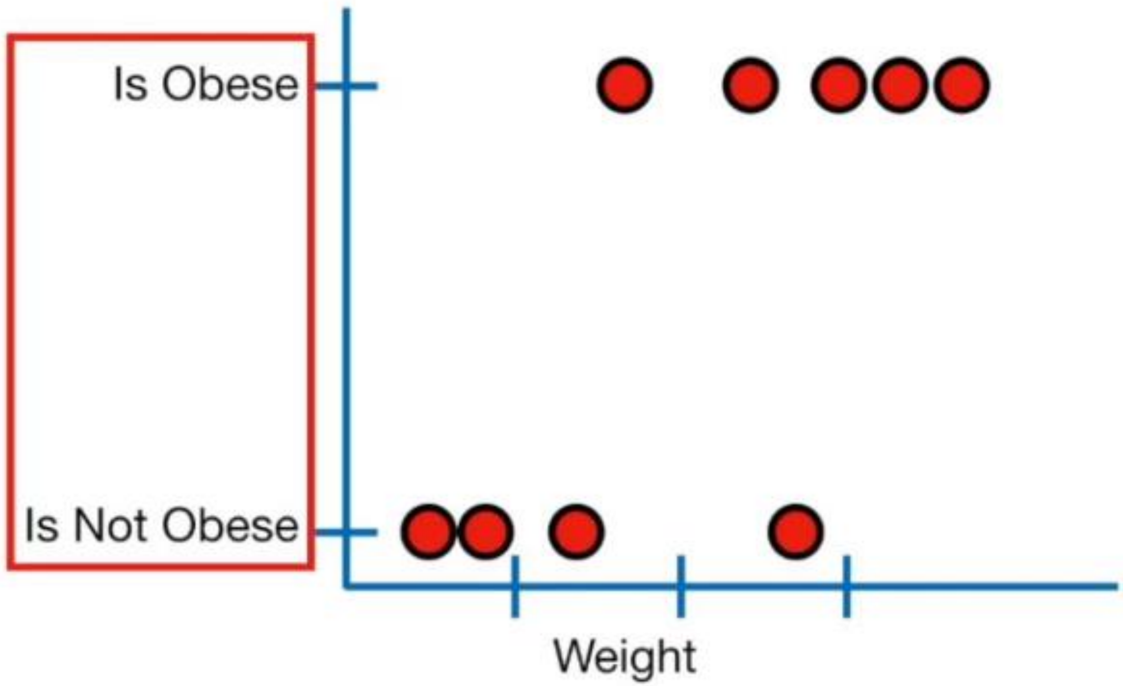
For these the outcome is not continuous or distributed normally.

Binary Logistic Regression major assumptions

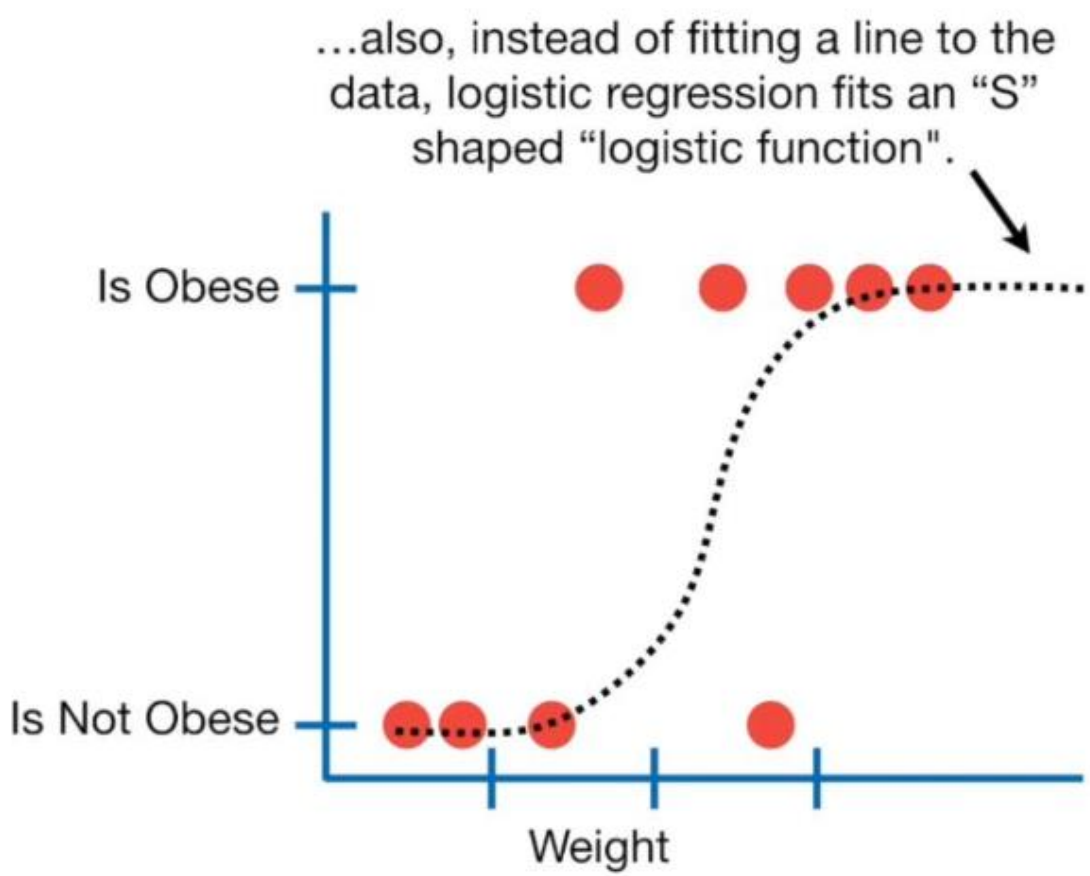
- The dependent variable should be dichotomous in nature (e.g., presence vs. absent).
- There should be no outliers in the data.
- There should be no high correlations (multicollinearity) among the predictors.

Logistic Regression

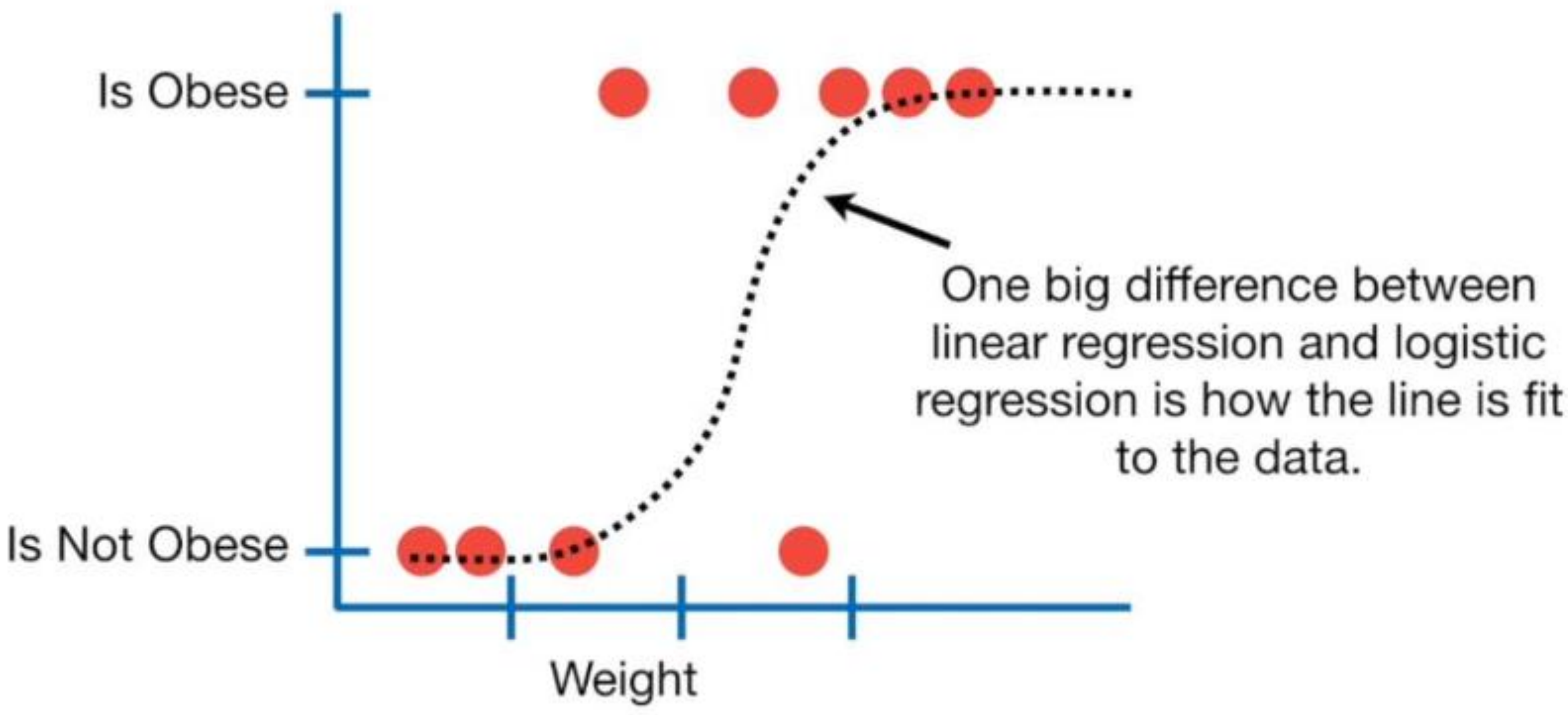
Logistic regression predicts whether something is ***True*** or ***False***, instead of predicting something continuous like ***size***.



Logistic Regression



Logistic Regression



LOGISTIC REGRESSION

Steps of Logistic Regression

Assume initial coefficients value as b0=b1=b2=0

Step 1: Calculate Prediction.

$$h(x) = \frac{1}{1 + e^{-x}}$$

Step 2: Calculate new coefficients.

Step 3: Repeat the process.

Step 4: Make Predictions

Steps of Logistic Regression

1. Assume initial coefficients value as b0=b1=b2=0

2. $$prediction = \frac{1}{1 + e^{-(B0 + B1 \times X1 + B2 \times X2)}}$$

3. b(new) = b(old) + α* (y-pred) * pred * (1-pred) *x

i.e b0(new) = b0(old) + α* (y-pred) * pred * (1-pred)

b1(new) = b1(old) + α* (y-pred) * pred * (1-pred) *x1

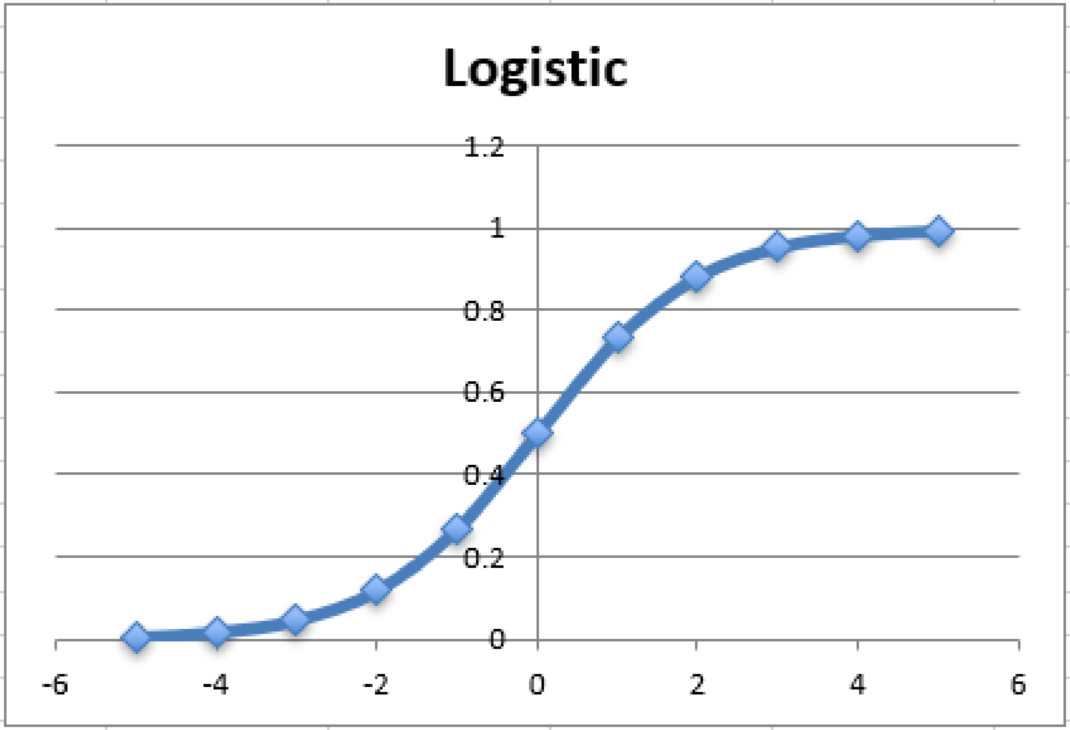
b2(new) = b2(old) + α* (y-pred) * pred * (1-pred) *x2

Where α is learning rate.

Example of Logistic Function

Logistic Function

Input	Logistic
-5	0.00669285
-4	0.01798621
-3	0.04742587
-2	0.11920292
-1	0.26894142
0	0.5
1	0.73105858
2	0.88079708
3	0.95257413
4	0.98201379
5	0.99330715



Example of Logistic Regression

X1	X2	Y
2.7	2.5	0
1.4	2.3	0
3.3	4.4	0
3.06	3.05	0
5.3	2.75	1

Step 1: b0=b1=b2=0

Step 2: $$prediction = \frac{1}{1 + e^{-(B0+B1\times X1+B2\times X2)}}$$

updated

21	Iteration	Bias	X1	X2	Y	B0	B1	B2	Prediction	B0(t+1)	B1(t+1)	B2(t+1)	Sharp Prediction	Squared Error	Error
22	1	1	2.7	2.5	0	0	0	0	0.5	-0.0375	-0.10125	-0.09375	1	0.25	1
23	1.1	1	1.4	2.3	0	-0.0375	-0.10125	-0.09375	0.402544	-0.06654	-0.14191	-0.16055	0	0.162042	0
24	1.2	1	3.3	4.4	0	-0.06654	-0.14191	-0.16055	0.224214	-0.07824	-0.18052	-0.21203	0	0.050272	0
25	1.3	1	3.06	3.05	0	-0.07824	-0.18052	-0.21203	0.218004	-0.08939	-0.21464	-0.24604	0	0.047526	0
26	1.4	1	5.3	2.75	1	-0.08939	-0.21464	-0.24604	0.129703	-0.05992	-0.05844	-0.16499	0	0.757416	1

Sharp Prediction = 1 if prediction >= 0.5
Sharp Prediction = 0 if prediction < 0.5

Squared Error = (prediction – Y)*(prediction – Y)

Error = 1 if Y not equal to sharp prediction
Error = 0 if Y equal to sharp prediction

Example of Logistic Regression

Dataset		
X1	**X2**	**Y**
2.7810836	2.550537	0
1.4654894	2.3621251	0
3.3965617	4.4002935	0
1.3880702	1.8502203	0
3.0640723	3.005306	0
7.6275312	2.7592622	1
5.3324412	2.0886268	1
6.9225967	1.7710637	1
8.6754187	-0.242069	1
7.6737565	3.508563	1
Learning Rate		
0.3		