Logistic Regression logistic classification

Arunkumar Nair

¿ Log

1

10^1

100 → 10^2

1000 → 10^3

10000

⋮

→ $\log_{10} 1 = 1$

$\log_{10} 100 = 2$

$\log_{10} 1000 = 3$

$= 4$

$8 = 2^3$

$2 \times 2 \times 2$

$\log_2 8 = 3$

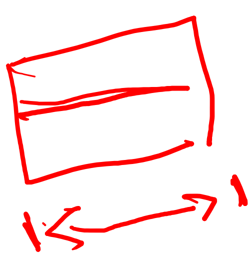
Expo

10^1

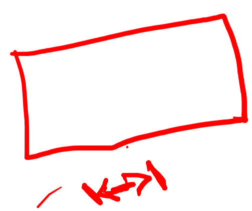
10^2

10^3

$2^3 = 8$



log



Log formula's

$$\log(ab) = \log(a) + \log(b)$$

$$\log\left(\frac{a}{b}\right) = \log(a) - \log(b)$$

$$\log(a^b) = b * \log(a)$$

$$\log\left(\frac{1}{x^a}\right) = -a$$

COMMON LOGARITHM TABLE

	0	1	2	3	4	5	6	7	8	9	Mean Difference								
											1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374	4	8	12	17	21	25	29	33	37
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755	4	8	11	15	19	23	26	30	34
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106	3	7	10	14	17	21	24	28	31
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430	3	6	10	13	16	19	23	26	29
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732	3	6	9	12	15	18	21	24	27
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014	3	6	8	11	14	17	20	22	25
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279	3	5	8	11	13	16	18	21	24
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529	2	5	7	10	12	15	17	20	22
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765	2	5	7	9	12	14	16	19	21
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989	2	4	7	9	11	13	16	18	20
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	12	14	15
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	9	11	13	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	11	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	12	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12

Probability

$P(x) = 0$ to 1

$P = \dfrac{\text{Number of events}}{\text{Total num of events}}$

$\text{odd} = \dfrac{\text{Number of events that will occur}}{\text{No. events that will NOT occur}}$

$= \dfrac{\text{Success}}{\text{Failure}}$

$= \dfrac{P}{1-P}$

$=$

Probability

$\underset{2}{\text{e}}$

H T

1 1

$\Rightarrow \dfrac{P(H)}{P(C)} = \dfrac{P(E)}{\text{Total}} = \dfrac{1}{2}$

odd

$\Rightarrow \dfrac{.5}{.5} = 1$

$\log(\text{odd}) = \ln\left(\dfrac{P}{1-P}\right)$

Logistic Regression

Logistic Regression, also known as **Logit Regression** or **Logit** Model, is a mathematical model used in statistics to estimate (guess) the probability of an event occurring having been given some previous data. **Logistic Regression** works with binary data, where either the event happens (1) or the event does not happen (0).

Logistic Regression

Putting z value to sigmoid function

Linear Regression Equation

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k$$

Sigmoid Function

$$p = \frac{1}{1 + e^{-z}}$$

$$p = \frac{e^z}{e^z + 1}$$

$$p = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k}}{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k} + 1}$$

Replace p in odd ratio and solve

$$\text{Odds Ratio S} = \frac{\textit{Probability of Success}}{\textit{Probability of Failure}} = \frac{p}{1-p}$$

Logistic Regression

Putting z value to sigmoid function

Linear Regression Equation

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k$$

Sigmoid Function

$$p = \frac{1}{1 + e^{-z}}$$

$$p = \frac{e^z}{e^z + 1}$$

$$p = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k}}{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k} + 1}$$

Replace p in odd ratio and solve

$$\text{Odds Ratio } S = \frac{\textit{Probability of Success}}{\textit{Probability of Failure}} = \frac{p}{1-p}$$

$$S = \frac{\dfrac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k}}{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k} + 1}}{1 - \dfrac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k}}{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k} + 1}}$$

$$S = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k}$$

Take log each side and solve

$$\mathrm{Ln}(S) = \boldsymbol{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k}$$

Transformed into Linear Regression known as log of Odds

Logistic Regression

Putting z value to sigmoid function

Linear Regression Equation

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k$$

Sigmoid Function

$$p = \frac{1}{1 + e^{-z}}$$

$$p = \frac{e^z}{e^z + 1}$$

$$p = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k}}{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k} + 1}$$

Replace p in odd ratio and solve

$$\text{Odds Ratio S} = \frac{\textit{Probability of Success}}{\textit{Probability of Failure}} = \frac{p}{1-p}$$

$$\frac{\frac{e^{\beta_0}}{e^{\beta_0}+1}}{\frac{1}{1} - \frac{e^{\beta_0}}{e^{\beta_0}+1}}$$

$$S = \frac{\dfrac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k}}{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k} + 1}}{1 - \dfrac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k}}{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k} + 1}}$$

$$S = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k}$$

·5 ·2 ·

Take log each side and solve

$$\frac{\frac{e^{\beta_0+1}}{e^{\beta_0}+1}}{1 + e^{\beta_0}+1}$$

$$\mathbf{Ln(S)} = \boldsymbol{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k}$$

Transformed into Linear Regression known as log of Odds

$$S = \frac{\dfrac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\dots+\beta_k x_k}}{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\dots+\beta_k x_k}+1}}{1-\dfrac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\dots+\beta_k x_k}}{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\dots+\beta_k x_k}+1}}$$

$$\Downarrow$$

$$\log(S) = \log\left(e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\dots+\beta_k x_k}\right)$$

$\Downarrow$ Take log each side and solve

$$\ln(S) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k$$

Transformed into Linear Regression known as log of Odds

$$S = \frac{\dfrac{e^{\beta_0}}{e^{\beta_0}+1}}{\dfrac{1}{1}-\dfrac{e^{\beta_0}}{e^{\beta_0}+1}} = \frac{e^{\beta_0}}{e^{\beta_0}+1} \div \left(\frac{1}{1} - \frac{e^{\beta_0}}{e^{\beta_0}+1}\right)$$

$$\frac{1 * (e^{\beta_0}+1) - e^{\beta_0} * 1}{1 * (e^{\beta_0}+1)}$$

Case-Study Data

We are provided a sample of 1000 customers.

We need to predict the probability whether a **customer of a Particular Age** will buy (y) a sim card or

not.

As we've a categorical outcome variable, we'll use logistic regression.

Linear to Logistic – (a)

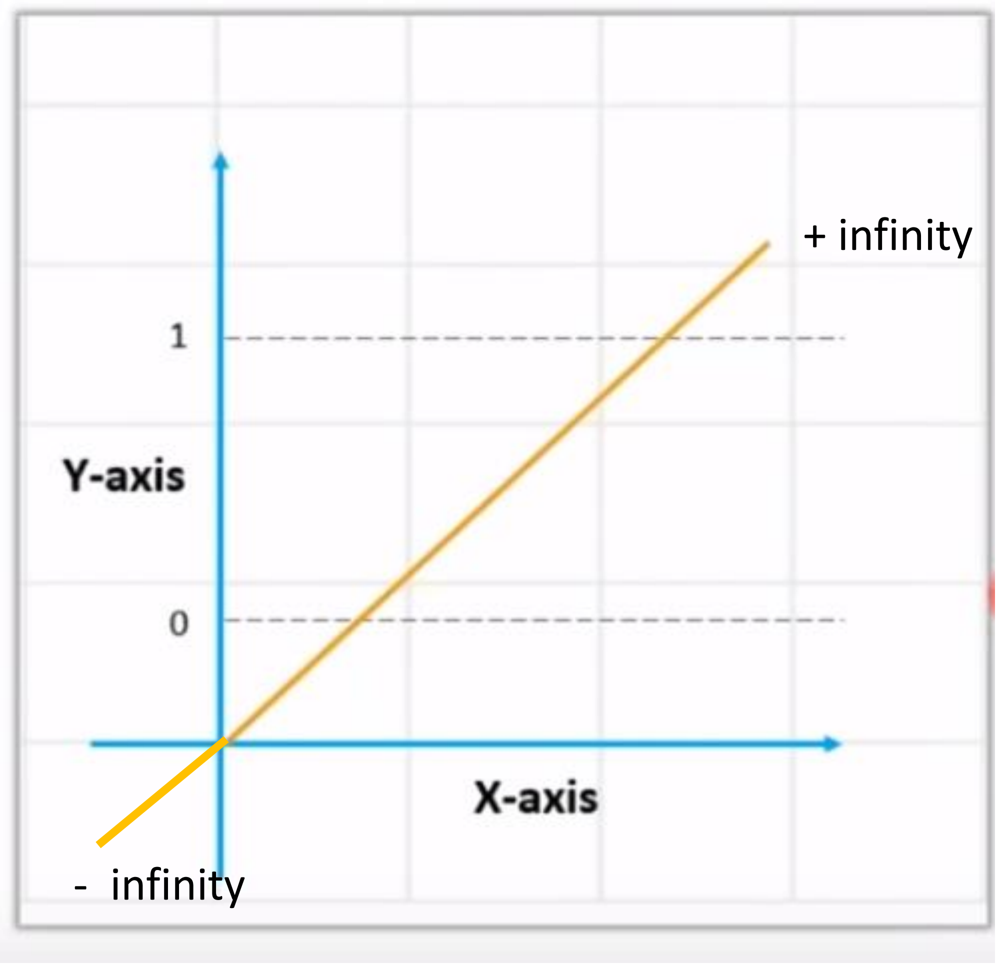
- To start with logistic regression, first write the simple linear regression equation with dependent variable enclosed in a link function:

$$Y = c + m X$$

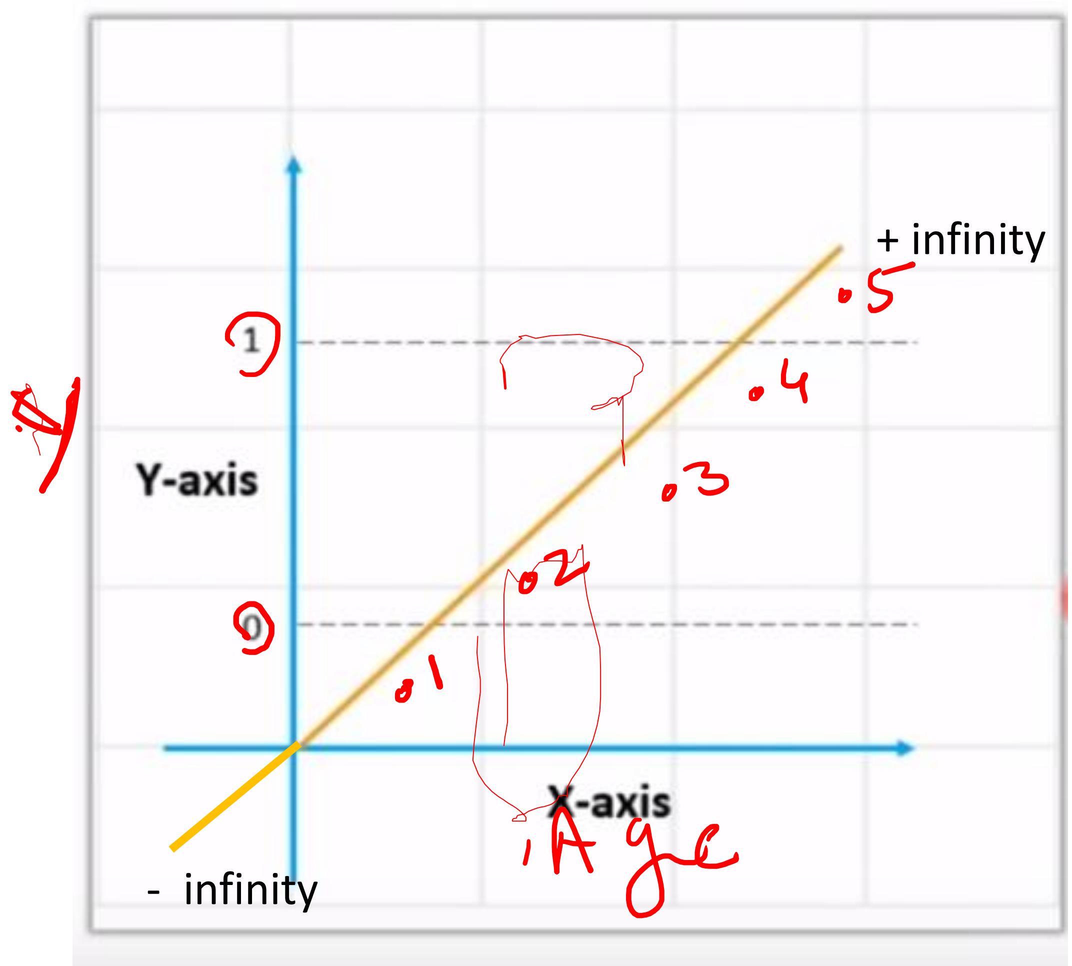
$$g(y) = \beta o + \beta(\text{Age}) \text{——} (a)$$

For understanding, consider '*Age*' as independent variable.

Linear Regression

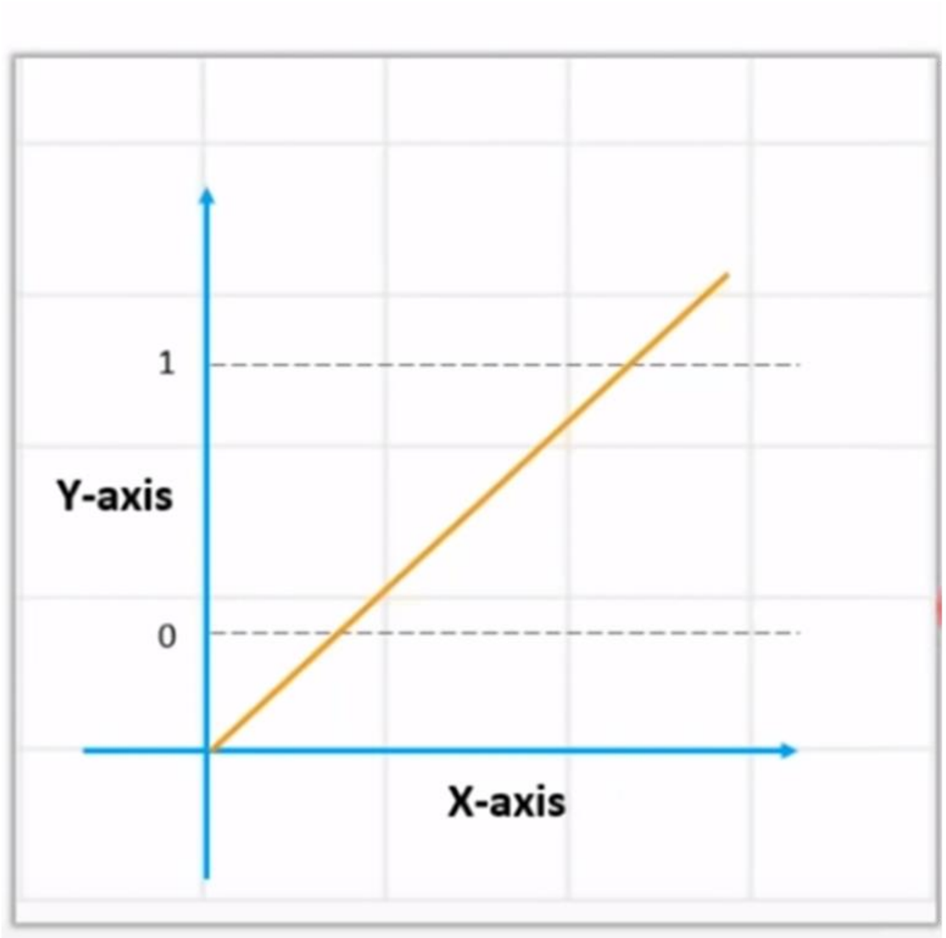


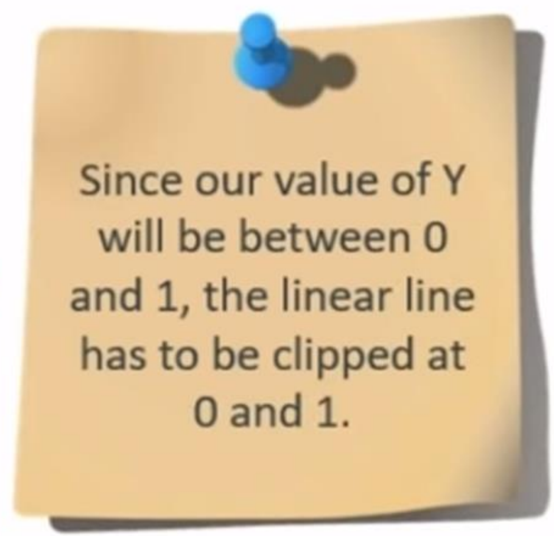
Linear Regression



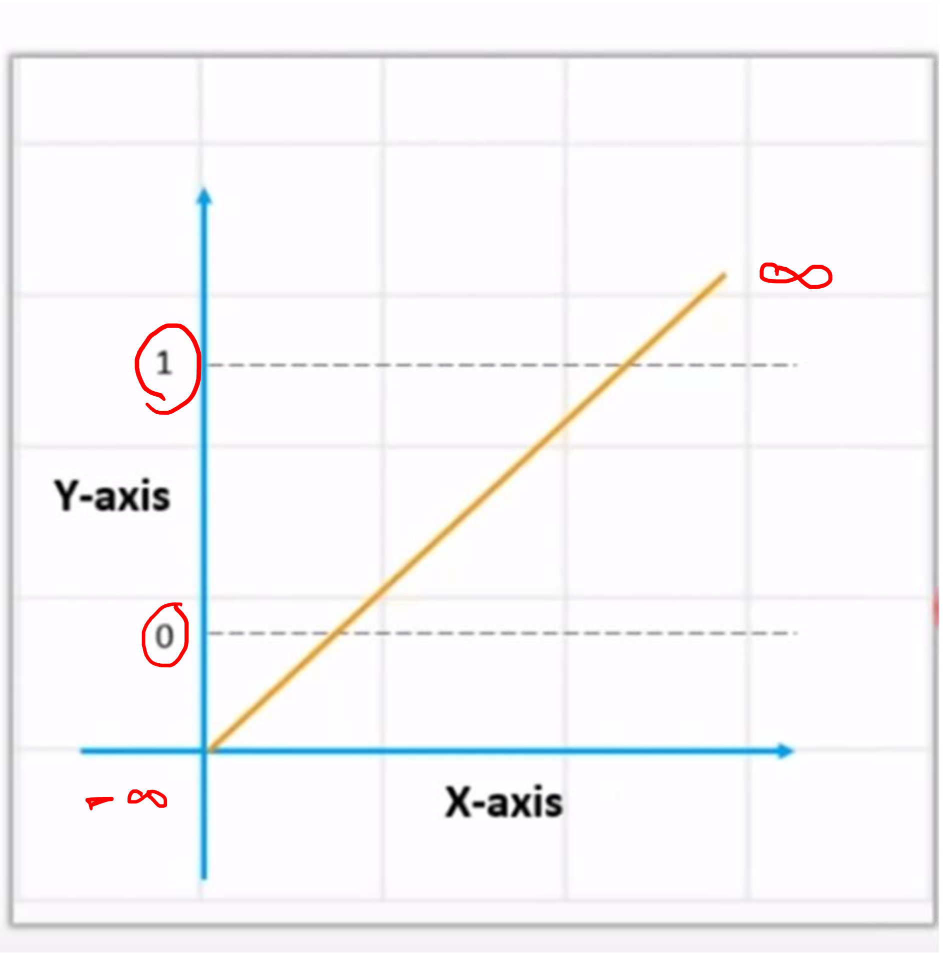
1 or 0

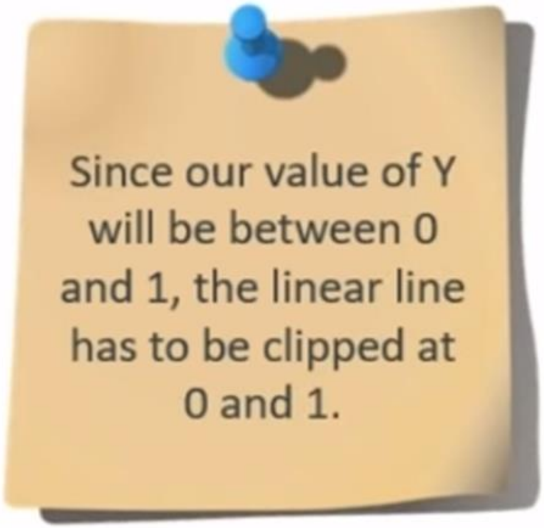
Linear regression equation: $y = \beta 0 + \beta 1X1 + \beta 2X2 \ldots. + \beta nXn$



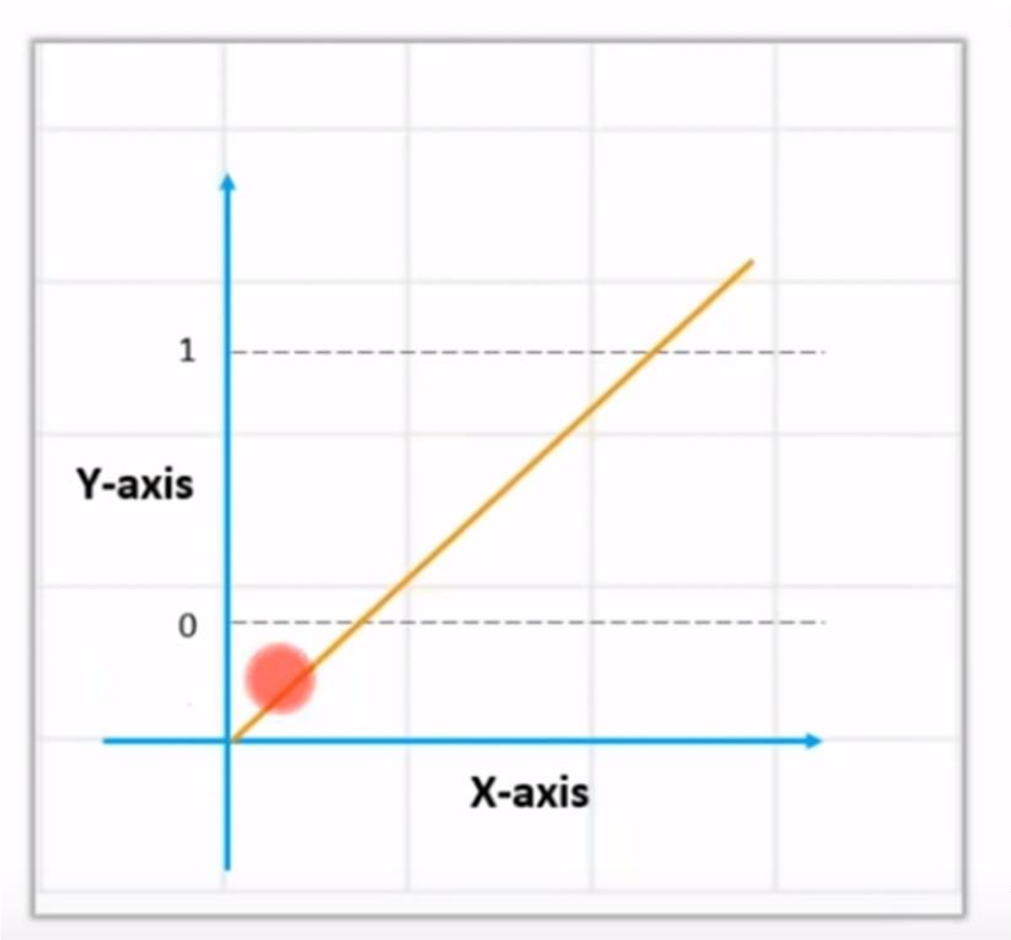


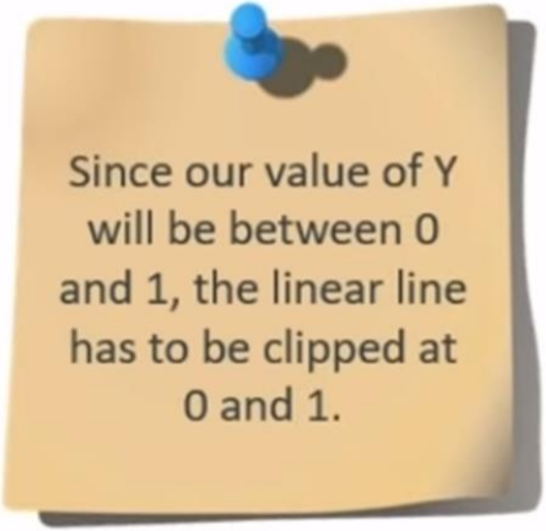
Linear regression equation: **$y = \beta 0 + \beta 1X1 + \beta 2X2 \ldots. + \beta nXn$**



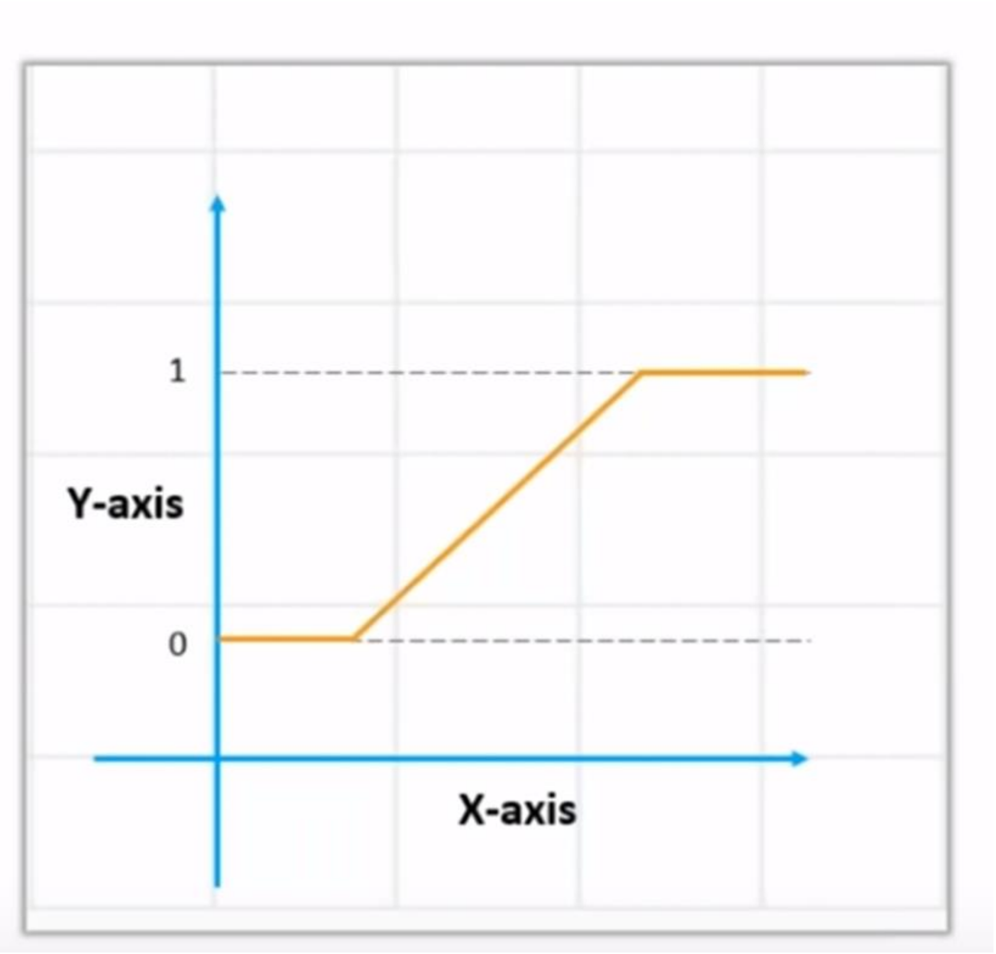


Value of Y – between 0 and 1

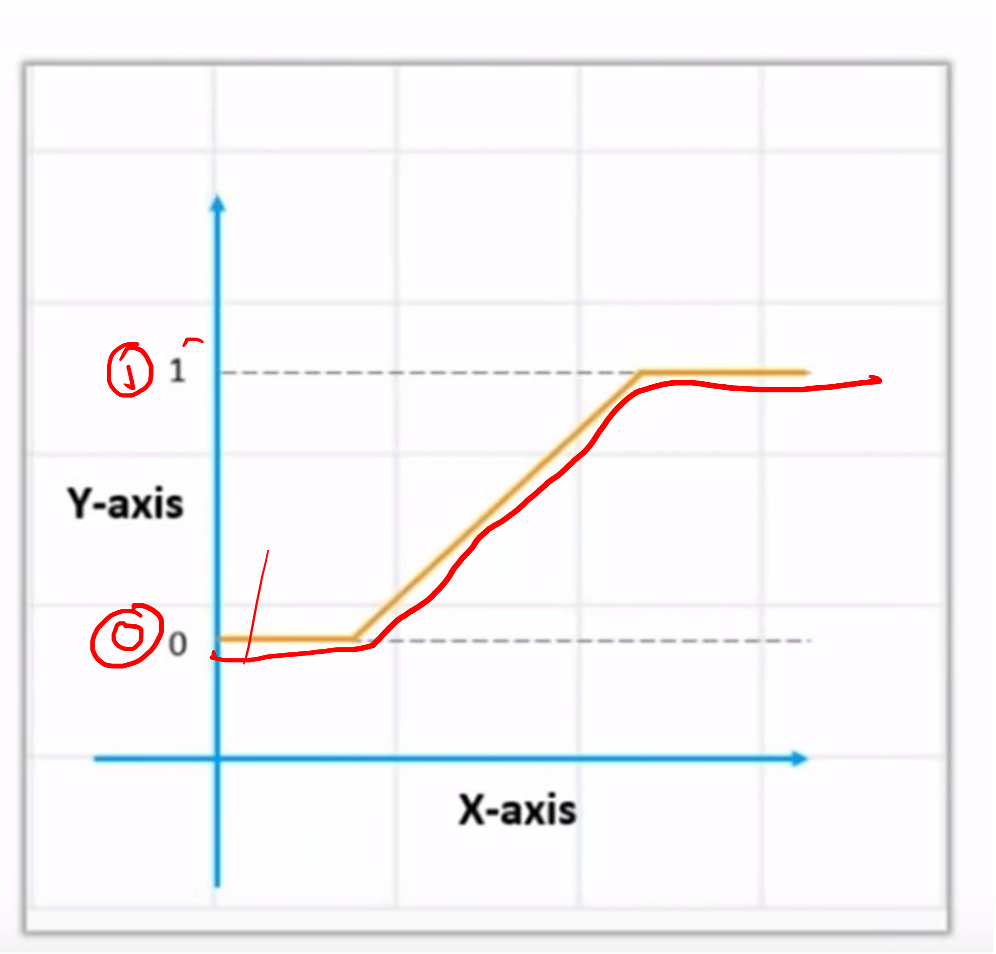




How to get the value of 0 and 1



How to get the value of 0 and 1



How to get the value of 0 and 1?
Use Sigmoid

- We Apply sigmoid function on the linear regression equation to get the S-curve so that it lies between 0 and 1

Sigmoid function: $\boldsymbol{p = 1 / 1 + e^{-y}}$

- A sigmoid function is a mathematical function/equation having a characteristic "S"-shaped curve or sigmoid curve.

How to get the value of 0 and 1?
Use Sigmoid

- We Apply sigmoid function on the linear regression equation to get the S-curve so that it lies between 0 and 1

Sigmoid function: $\boldsymbol{p = 1 / 1 + e^{-y}}$

- A sigmoid function is a mathematical function/equation having a characteristic "S"-shaped curve or sigmoid curve.

How to get the value of 0 and 1?
Use Sigmoid

- We Apply sigmoid function on the linear regression equation to get the S-curve so that it lies between 0 and 1

Sigmoid function: $\boldsymbol{p = 1 / 1 + e^{-y}}$

- A sigmoid function is a mathematical function/equation having a characteristic "S"-shaped curve or sigmoid curve.

Convert Linear to Logistics

- Linear regression equation: **y = β0 + β1X1 + β2X2 …. + βnXn**
- Sigmoid function: $\boldsymbol{p = 1 / 1 + e^{-y}}$

  e^{-y} y is replaced

- Logistic Regression equation: $\boldsymbol{p = 1 / 1 + e^{-(\beta 0 + \beta 1X1 + \beta 2X2 \dots + \beta nXn)}}$

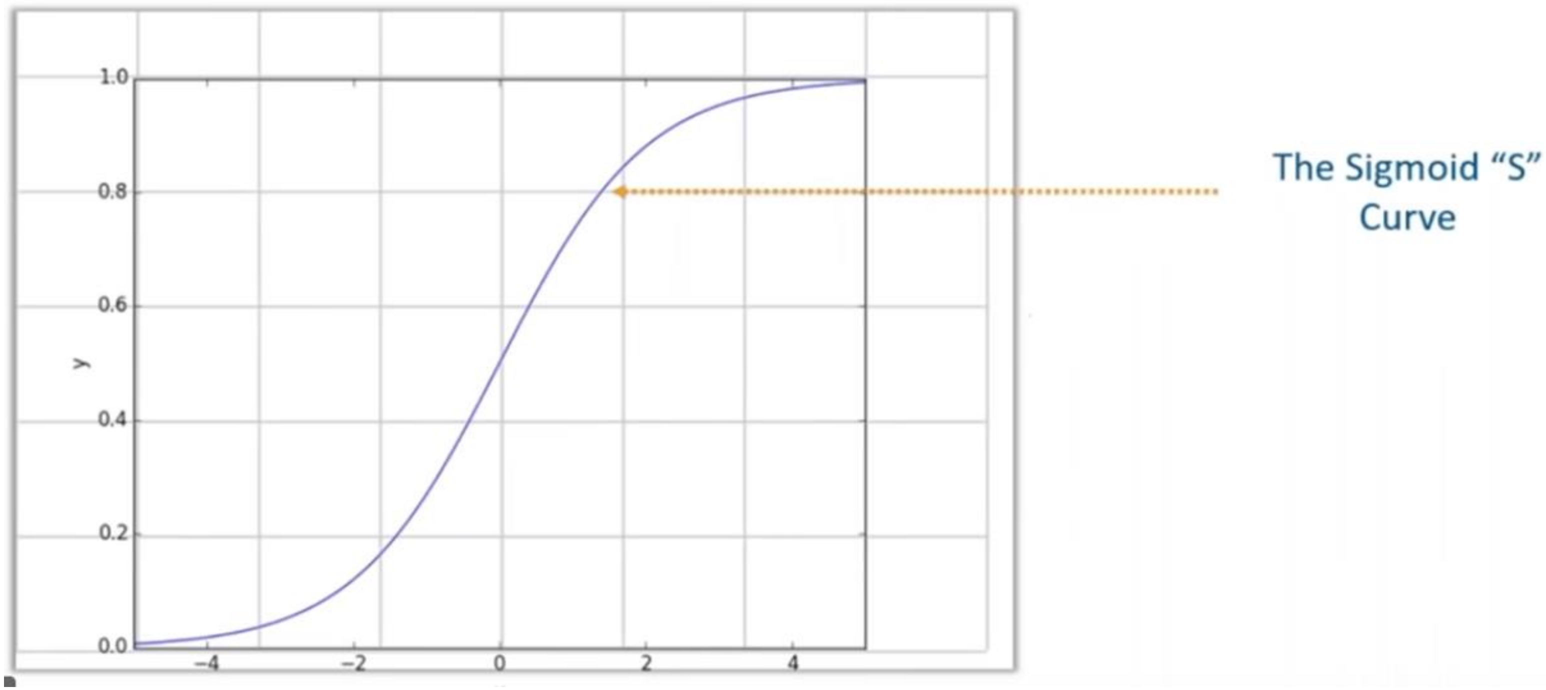
Convert Linear to Logistics

- Linear regression equation: **$y = \beta 0 + \beta 1X1 + \beta 2X2 \ldots. + \beta nXn$**
- Sigmoid function: **$p = 1 / 1 + e^{-y}$**

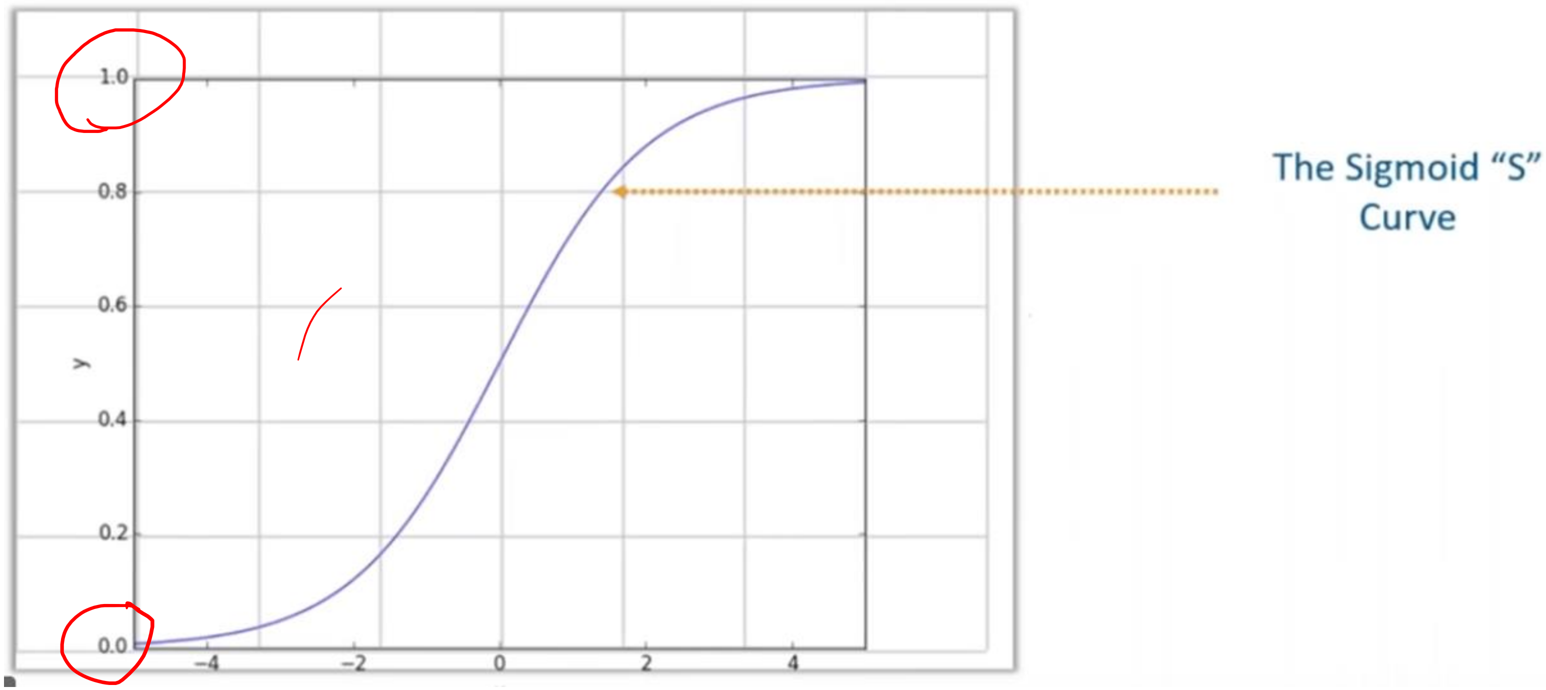
  e^{-y} y is replaced

- Logistic Regression equation: **$p = 1 / 1 + e^{-(\beta 0 + \beta 1X1 + \beta 2X2 \ldots. + \beta nXn)}$**

Sigmoid – S-curve



Sigmoid – S-curve



Sigmoid

- Sigmoid curve converts any value from

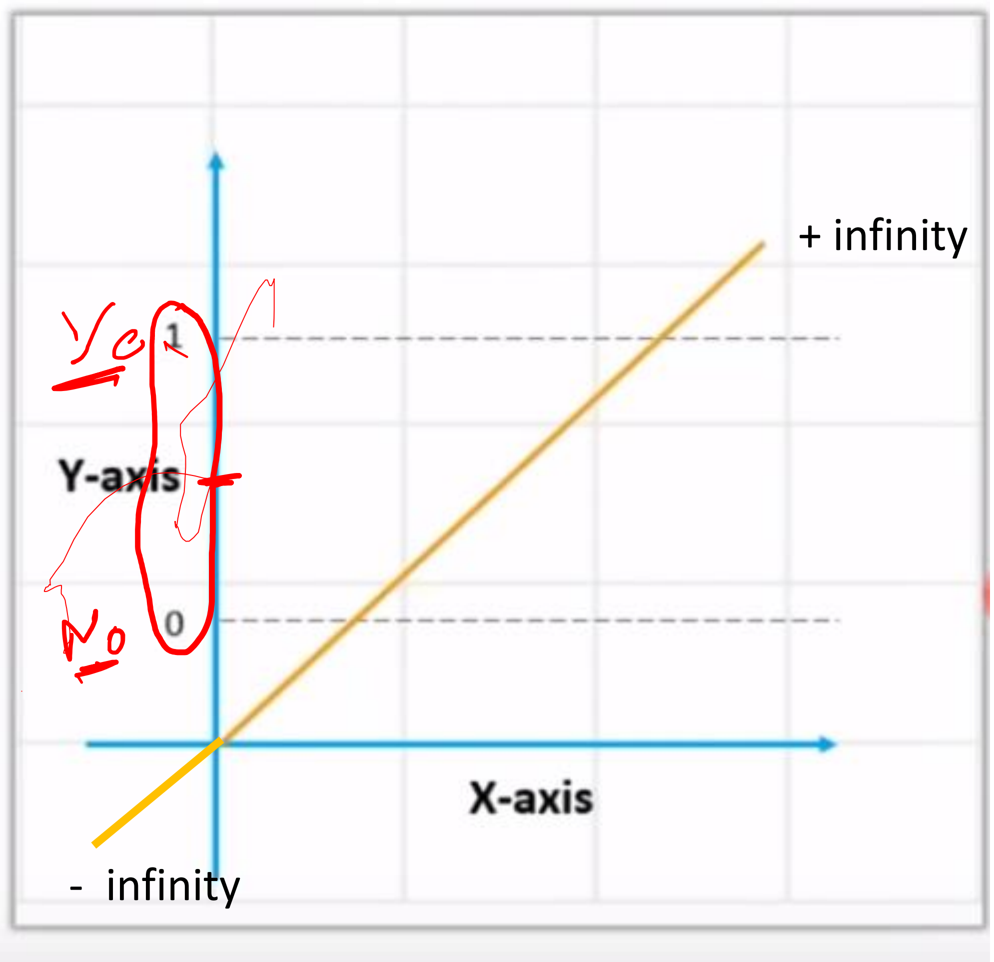
-infinity to +infinity to (0 to 1)

- Sigmoid will output:

'0' as x approaches $-\infty$

'1' as x approaches $+\infty$

Linear Regression

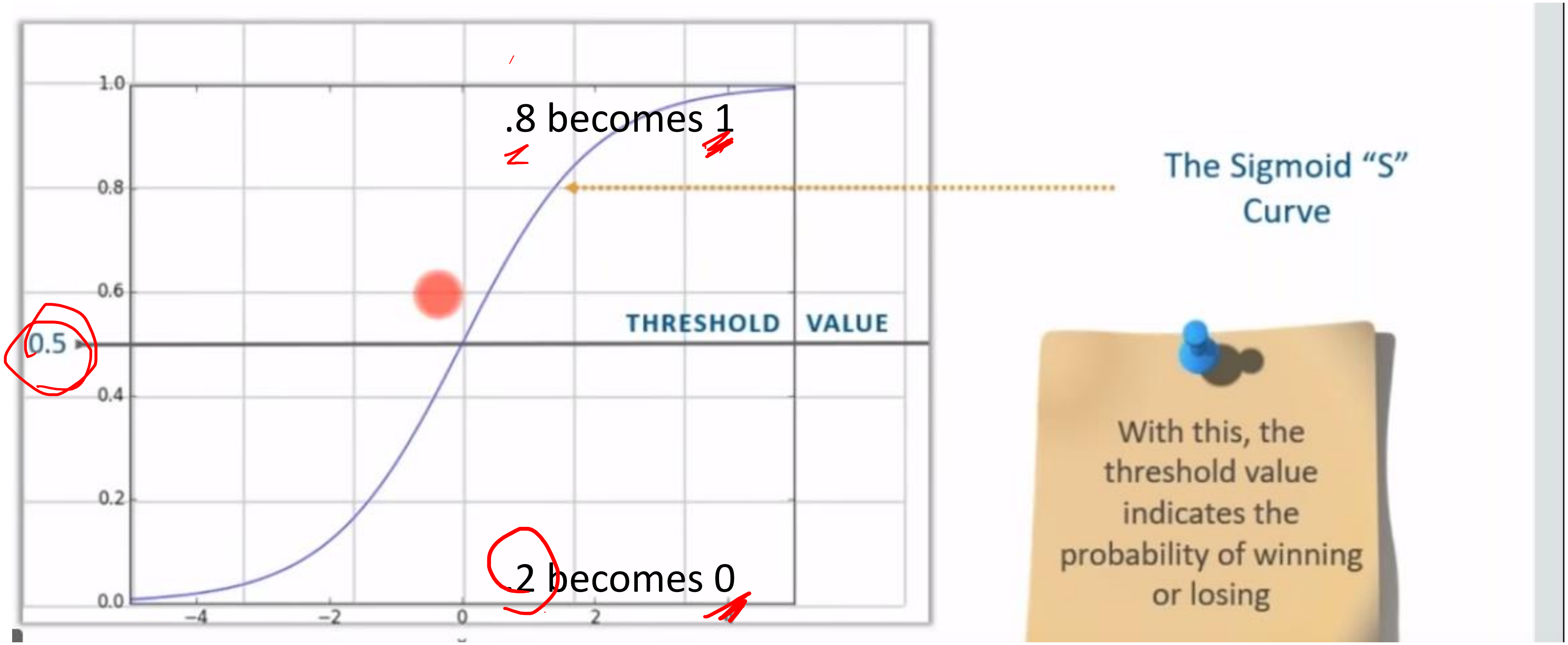


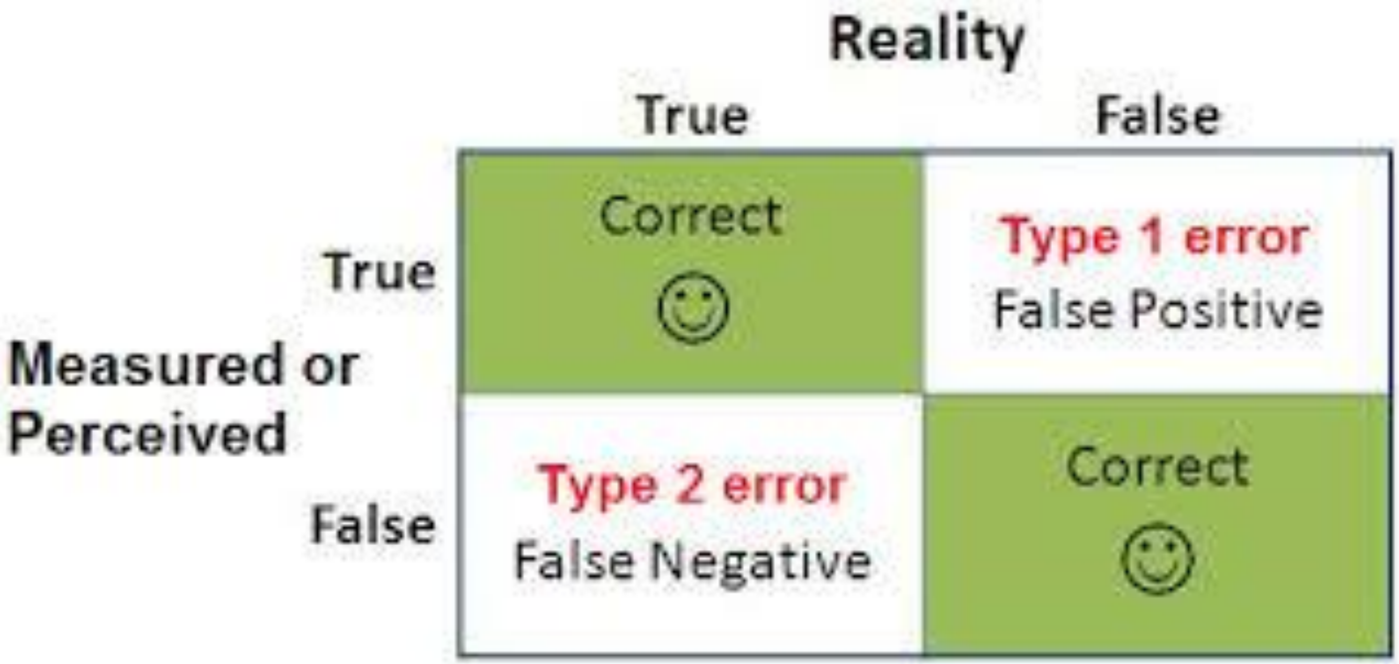
3 1 2 Yes

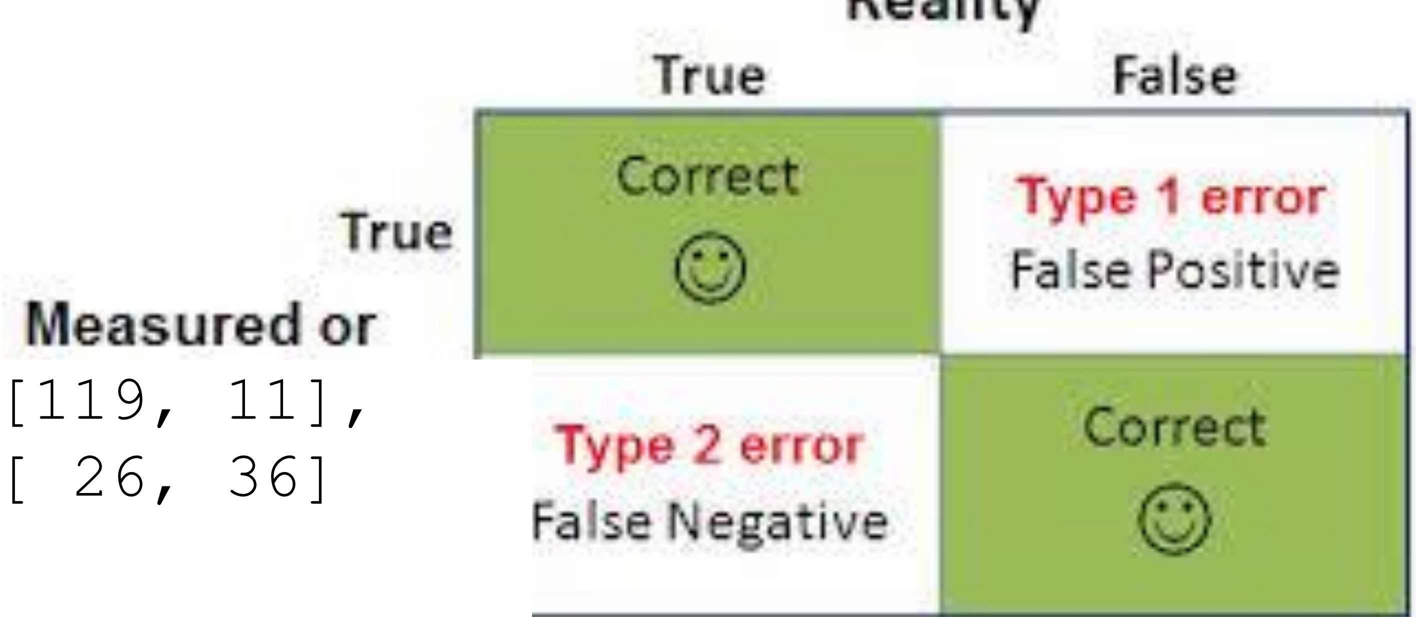
2

1 0 1 No

Probability values for the answers




Reality
True
False
Measured or Perceived
True
Correct
Type 1 error
False Positive
False
Type 2 error
False Negative
Correct




```
[[119, 11],
 [ 26, 36]
```

Accuracy=
(TP + TN)/
(TP+TN+FP+FN)

Thanks

End