Logistic Regression

Linear Regression :

Type of Target : Continuous

What if : Discrete

Classify something

- Email is spam or not
- Online transaction frad or not
- Credit card fraud or not
- Marketing Call buy or not
- Sales - buy or not
- Operation - Attrition or not
- Healthcare - malign or not

Logistic Regression

1. Classification Algorithm
2. Name is misleading
3. Probablistic way of predicting

Linear Regression

$$\text{Model} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots$$

Continuous $\rightarrow$ $-\infty$ — $+\infty$

$\Downarrow$

Probablity

Linear Regression Output range

$(-\infty \quad , \quad +\infty)$

$\Downarrow$

$(0 \quad , \quad 1)$

1. Apply exponential to convert the range between $(0, \infty)$

$e^{\hat{y}} \Rightarrow (0, \infty) = \exp(y)$

2. Any number divided by number + 1

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

$\boxed{\frac{\exp(y)}{\exp(y) + 1}} \Rightarrow (0, 1)$

Probablity

1. proportion $= 0, 1$

2. odds $= \boxed{\frac{\text{Success}}{\text{failure}}} = 0, \infty$

3. $\log[\text{odds}] = \log\left(\frac{p}{1-p}\right) = (-\infty, +\infty)$

$$y = \beta_0 + \beta_1 x \quad - \quad (-\infty, +\infty)$$

$$e^{(\beta_0 + \beta x)} \quad - \quad (0, \infty)$$

$$\boxed{p = \frac{e^{(\beta_0 + \beta_1 x)}}{e^{(\beta_0 + \beta_1 x)} + 1}} \quad - \quad (0, 1)$$

$$p = \frac{e^{y}}{e^{y} + 1} \qquad \Big| \qquad q = 1 - p$$

$$= 1 - \left(\frac{e^{y}}{e^{y} + 1} \right)$$

$$(e^{y} + 1)\, p = e^{y}$$

$$e^{y} p + p = e^{y}$$

$$e^{y} p - e^{y} = -p$$

$$e^{y} [p - 1] = -p$$

$$e^{y} = \frac{-p}{p - 1}$$

$$\boxed{e^{y} = \left[\frac{p}{1 - p} \right]}$$

success $= p$

Failure $= 1 - p$

$$e^y = \frac{P}{1-P}$$

apply log

$$\log e^y = \log\left(\frac{P}{1-P}\right)$$

$$y = \log\left(\frac{P}{1-P}\right)$$

$$\log\left(\frac{P}{1-P}\right) = \underbrace{\beta_0 + \beta_1 x}_{\text{linear Regression}}$$

$$\underbrace{\log\left[\text{Odds}\right]}$$

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta x$$

$$e^{\left[\log\left(\frac{p}{1-p}\right)\right]} = e^{\beta_0 + \beta x}$$

$$\frac{p}{1-p} = e^{\beta_0 + \beta x}$$

$$p = (1-p)\left(e^{\beta_0 + \beta x}\right)$$

$$p = e^{\beta_0 + \beta x} - p \cdot e^{\beta_0 + \beta x}$$

$$p + p\, e^{\beta_0 + \beta x} = e^{\beta_0 + \beta x}$$

$$p\left(1 + e^{\beta_0 + \beta x}\right) = e^{\beta_0 + \beta x}$$

$$p = \frac{e^{\beta_0 + \beta x}}{1 + e^{\beta_0 + \beta x}}$$

$\div \; e^{\beta_0 + \beta x}$ on Num and Deno

$$p = \frac{1}{\dfrac{1 + e^{\beta_0 + \beta x}}{e^{\beta_0 + \beta x}}}$$

$$= \frac{1}{e^{-(\beta_0 + \beta x)} + 1}$$

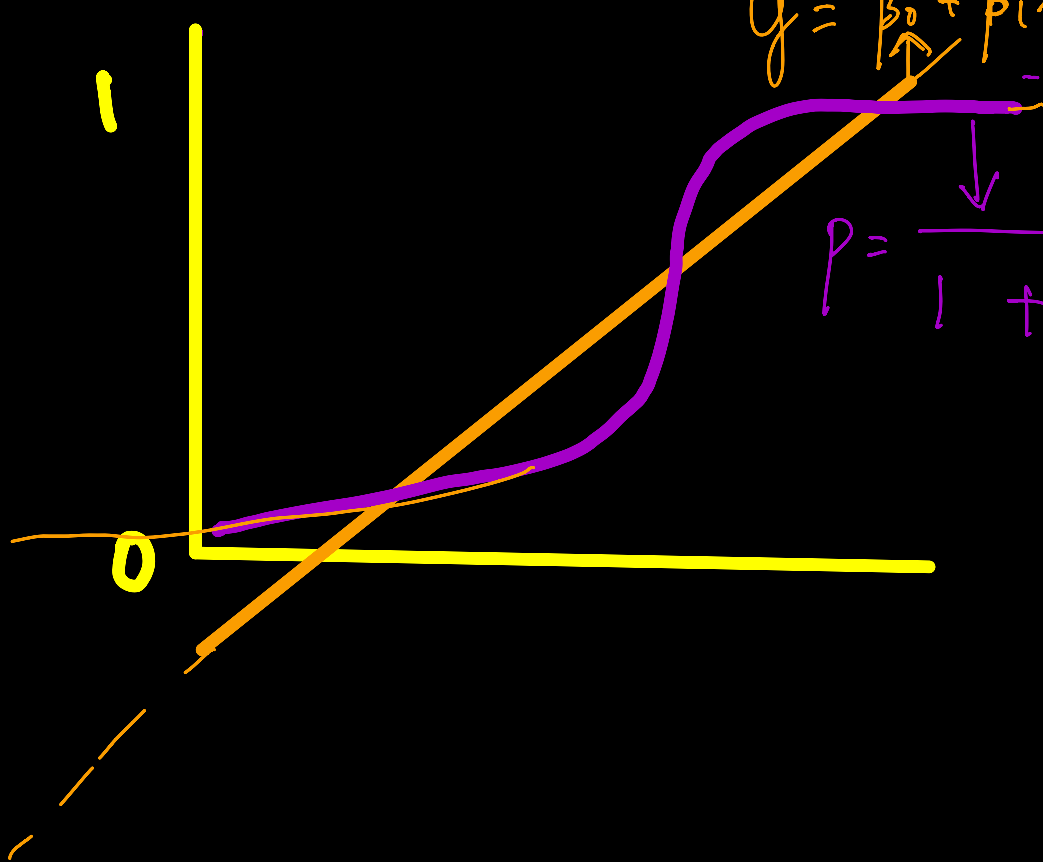
$$\boxed{p = \frac{1}{1 + e^{-(\underbrace{\beta_0 + \beta x}_{\text{linear Regression}})}}}$$

Sigmoid Function

Logit Function

(0, 1)

$$y = \beta_0 + \beta_1 x$$



$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

Age	Status
23	Died
40	Survived
42	Survived
28	Died

Intercept = 1.8185

Slope (Age) = −0.0665

Age = 0 = ?

Died = 0

Survied = 1

⇒ 0.5

⇒ 0.86

⇒ 0.48

$bx + b$

$$\log\left(\frac{p}{1-p}\right) = 1.8188 - 0.0668\,(\text{Age})$$

$$= 1.8188 - 0.0668\,(0)$$

$$\log\left(\frac{p}{1-p}\right) = 1.8185$$

$$\frac{p}{1-p} = e^{1.8185}$$

$$\frac{p}{1-p} = 6.16$$

$$p = 6.16 - 6.16p$$

$$p(1+6.16) = 6.16$$

$$p = \frac{6.16}{1+6.16}$$

$$p = \frac{6.16}{7.16} = 0.86$$

the 0.5

Survived

$$p = \frac{1}{1 + e^{-(1.8165 - 0.0668(Age))}}$$

$$\frac{1}{1 + e^{-(1.8185 - 0.0668(0))}}$$

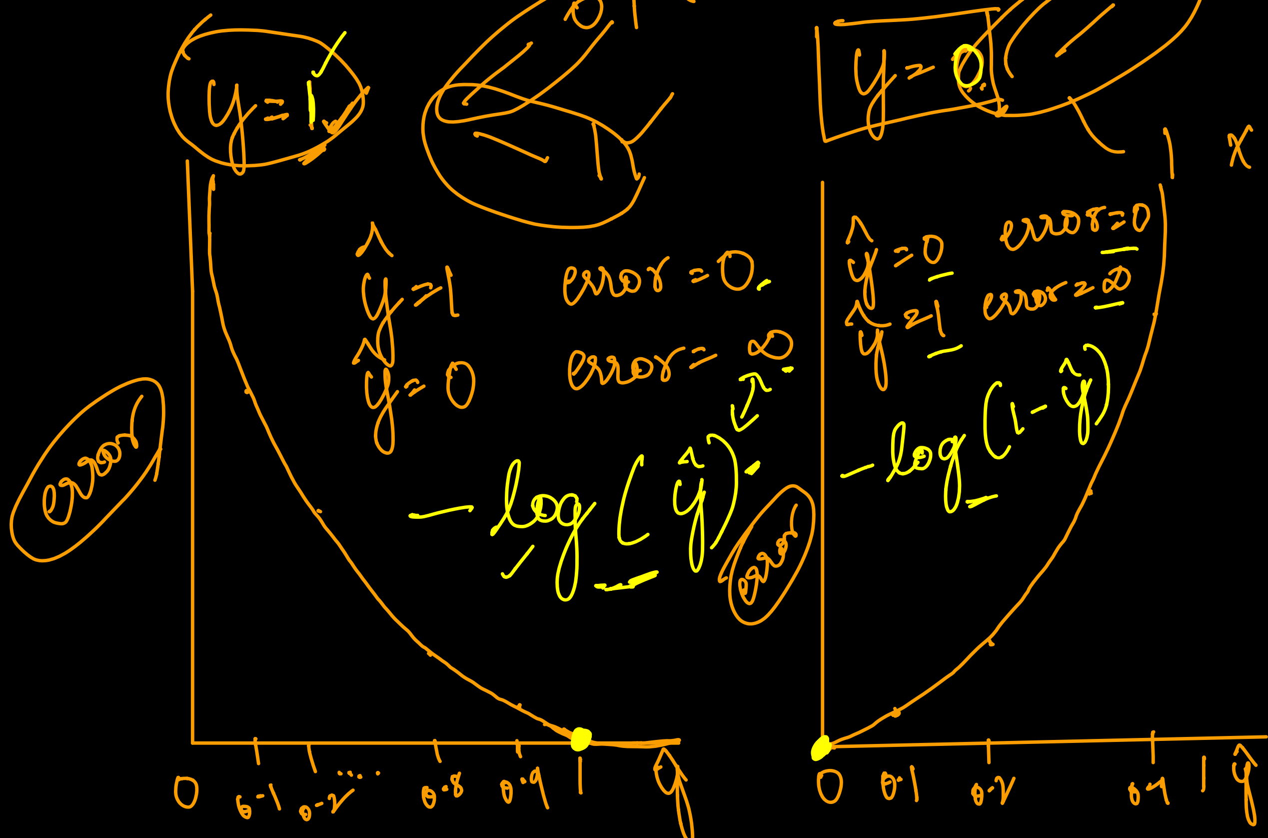
$$\frac{1}{1 + \underline{e^{-(1.8185)}}}$$

$$\frac{1}{1 + 0.162} = \frac{1}{1.162}$$

$$= \underline{0.86}$$

Logistic Regression on Gradient descent

Cost function



$$\text{Cost} = \begin{cases} -\log(\hat{y}) & \text{if } y=1 \\ -\log(1-\hat{y}) & \text{if } y=0 \end{cases}$$

$$Cost = \sum y\left[-\log(\hat{y})\right] + (1-y)\left[-\log(1-\hat{y})\right]$$

log loss $\Rightarrow$

$$= -\left[\left[\sum y_i \log(\hat{y}_i)\right] + (1-y_i)\left[\sum \log(1-\hat{y}_i)\right]\right]$$

Gradient descent

Error Metrics

$y = 0$ → 0 ✓, 1 ✗

$y = 1$ → 0 ✗, 1 ✓

Confusion Matrix		Predicted class	
		Class = Yes (Female)	Class = No (Male)
Actual class	Class = yes (Female) 6	2 ✓ TP	4 ✗ FN
	Class = No (Male) 4	2 ✗ FP	2 ✓ TN
		4	6

10 Sample

height	Gender y	$\hat{y}$
.	F	F
.	F	F
.	M	F
.	F	M
.	M	F
	F	M
	F	M
	M	M
	F	M
	M	M

F = Positive
M = Negative

TP = 2
TN = 2
FP = 2
FN = 4

Confusion Matrix	Predicted	Class		Total
		Class = Yes (Female)	Class = No (Male)	
Actual Class	Class = yes (Female)	2 ✓ TP	4 X FN	6 ✓
	Class = No (Male)	2 X FP	2 ✓ TN	4 ✓
Total		4	6	10

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$= \frac{2 + 2}{2 + 2 + 2 + 4}$$

✓

$$= \frac{4}{10} = 0.40$$

$$= 40\%$$

Positive Female

$$\text{Recall} = \frac{TP}{\text{Actual Positive}} = \frac{TP}{TP + FN}$$

$$= \frac{2}{6} = \frac{1}{3}$$

$$= 33.3\%$$

$$\text{Recall}_{\text{Negative Male}} = \frac{TN}{\text{Actual Negative}} = \frac{TN}{FP+TN}$$

$$= \frac{2}{4} = \frac{1}{2}$$

$$= 50\%$$

$$\text{Precision}_{\text{Positive Female}} = \frac{TP}{\text{Predicted Positive}}$$

$$= \frac{TP}{TP+FP} = \frac{2}{2+2} = 0.5$$

$$50\%$$

$$\text{Precision}_{\text{Negative male}} = \frac{TN}{TN+FN} = \frac{2}{2+4}$$

$$= \frac{2}{6}$$

$$= 33.3\%$$

$$\text{F1-Score}_{\text{Positive Female}} = \frac{2PR}{P+R} = \frac{2 * 50 * 33.3}{50 + 33.3}$$

Harmonic mean

$$= \frac{3333}{83.3}$$

$$= 40\%$$

100 Patient 2 Cancer

100 = No Cancer

100 Patient = No Cancer

% Accuracy = 98%

Recall = 0% = $\frac{0}{}$

0% = $\frac{0}{0}$

Precision =

F1-Score = $\frac{2 * 2 \times 2}{2+2}$

100% Cancer

Accuracy = $\frac{2\%}{100}$

Recall = 2%

Prec $\frac{2}{100}$

Model => Covid Positive →

Positive

Negative

Recall 100% $\frac{TP}{AP}$ =

Gender prediction Balanced 1000

Accuracy 0 - 500

1 - 500

Spam Prediction

Spam ⇒ Primary

Primary ⇒ spam

Precision $\dfrac{TP}{PP}$

ROC Curve

Receiving Operating Characteristics