Independent variables

$\underbrace{\hspace{3em}}_{x}$

Target variable

$\underbrace{\hspace{3em}}_{y}$

↓

continuous

(1) Regression → continuous

Predict new type of target.

Predict continuous

(1) Regression /

Predict = Discrete

Classification

Eg) Classification

(1) Threat / Not a threat.

Real life M.L applic..

NLP / DL

(1) ~~Spam~~ / ham → Gmail

(2) Loan approval → Loan App. / Not approved

(3) Take up the sentiment analysis

C.V / D.L ← (4) T.B analysis — X-Ray

Classification → Classifying the data set
↓ ↓
[0, 1, 2] discrete target.
0.58
(Spam, ham) → Category.

Come up with words or questions that can ask what are the chances of being

① What is the ~~probability~~ chance ? that India wins the toss

Events
T | H

⇒ Possibility
⇒ Chances of occurrence
⇒ Outcome.
⇒ Expectation
⇒ Likelihood.
⇒ Odds. → what are the odds?

Odds
⇒ what are the odds of a mail being Spam?
↓
Spam/ham

⇒ Logistic Regression

Given :- The independent variables

Todo :- ~~Utalt~~ We will have to classify

⇒ Probability — $\frac{\text{chance of the event}}{\text{All possible event}}$ =

H T ⇒ $P(\text{head}) = \frac{1}{2} = 0.5$

Rolling a dice

$P(1 \text{ (or) } 2) = \frac{2}{6} = \frac{1}{3} = 0.33$

What is the range of probability?

$0 \rightarrow 1$

$x_1, x_2, x_3, x_4, 4$

0.8

$0.5 > 1$

$0.5 < 0$

$0 \rightarrow 1$

classify (0 (or) 1)

$0 - 1$

$$\text{Odds} \Rightarrow \frac{P(\text{occurance})}{P(\text{not-occurance})} = \frac{P}{1-P} = \frac{p}{q}$$

$P(\text{heads}) = 1/2 = 0.5$

$P(\text{tails}) = (1-0.5) = 0.5$

$$\Rightarrow \frac{0.5}{0.5} = 1$$

fair die

Odds (1 or 2)

$P(1 \text{ or } 2) = 2/6$

not (2/6)

1 or 2

$1 - 2/6$

$= 4/6$

$$\text{odds} \Rightarrow \frac{0.333}{0.666} = 1/2 = 0.5$$

coin

$P(\text{heads}) = 7/10 = 0.7$

$$\text{Odds (heads)} = \frac{0.7}{0.3} = 2.33$$

Odds range $\rightarrow$ ~~0 to 1~~

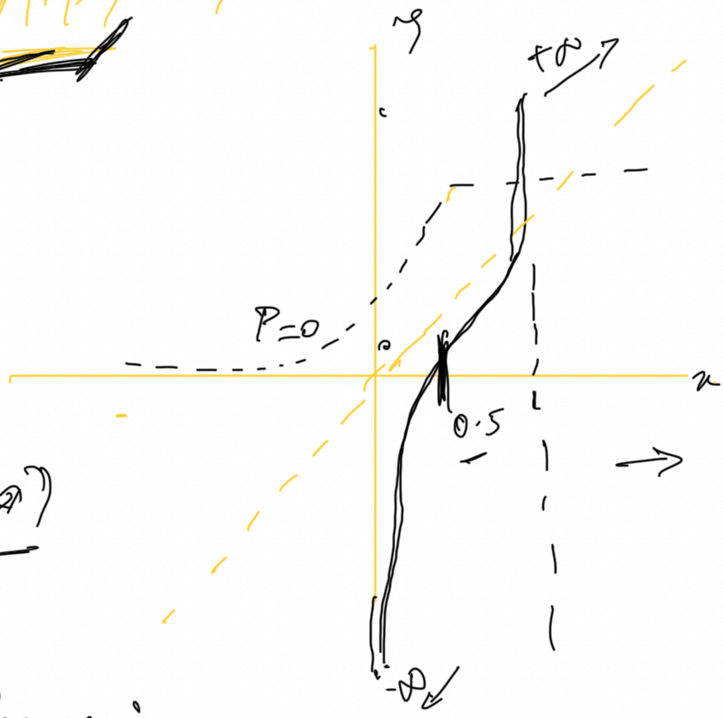
$\boxed{0 \rightarrow +\infty}$

$\log($

$(0 \rightarrow \infty)$

$\log_e(P/1-P) \Rightarrow$

$\log_e \rightarrow \log$



Range $\Rightarrow$ $(-\infty, \infty)$

$(-\infty, \infty)$

Linear Regression $(-\infty, \infty)$ - Continuous.

$\uparrow \rightarrow$ Classification

$\log_e(p/1-p) = y(mx+c)$

$$\cancel{\exp}\left(\cancel{\log_e}(p/(1-p))\right) = e^y$$

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

$$p = e^y(1-p)$$

$$p = e^y - e^y p$$

$$p + e^y p = e^y$$

$$p(1+e^y) = e^y$$

$$p = e^y / (1+e^y)$$

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

$> 0 \to 1$

$< 0 \to 0$

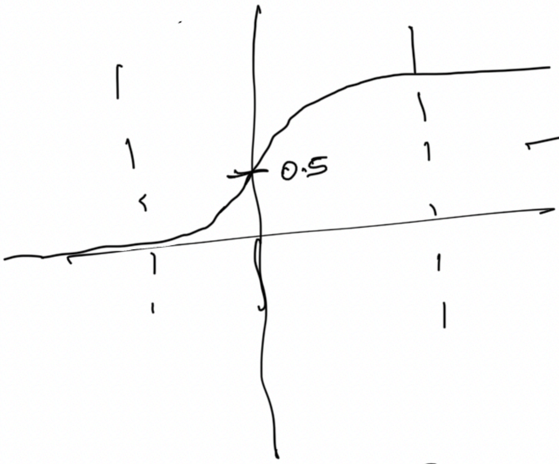
$$\boxed{p = \frac{1}{1+e^{-\underbrace{(mx+c)}_{y}}}}$$

Sigmoid

(x)

Logistic Regression

↳ Derived from linear reg.



→ Sigmoid Graph

$p = 1/1+e^{-y}$

→ Is Logistic Regression linear model?

→ when looking at in ~~terms~~ of log-odds it is linear

0.5 → 0, 1 → Classification

a linear model

loss function

~~cannot use MSE?~~

⇒ Write a blog on why we cannot use MSE in logistic Regres?

classification 0.0001, 0.0002

act $y = 0$ → $\hat{y}$: 0, 1; 0.8, 0.7, 0.6

loss fun, $y=0$: 0, 1

$fn(\hat{y})$ → Error

P → 0.3, 0.05, 0.02

$\hat{y}$	Error
0	0
1	∞

$y = 0$

(1) Real value / predic. ⇒ $\frac{0}{0}$ = Error / undefined

$\log(0) = -\infty$

$\log(1) = 0$

actual $y = 0$

y	$\hat{y}$	Error
0	0	0

$\log(\ )$
$\log(1) = 0$

actual $y = 0$

y	$\hat{y}$	Error
0	0	0
0	1	∞

$\log(1-\hat{y})$

$\log(\hat{y}) \Rightarrow \log(1) \Rightarrow 0$

$\log(0) = -\infty$

$y = 0$

$\Rightarrow$

$\log(1) = 0$

$y = 0$

$\log(1-1) = \log(0) = -\infty$

Case-1 :- $-(1-y)\log(1-\hat{y})$

$\hat{y} = 0$
$y = 0$

$-(1)\log(1-0)$

$-\log(1) = -0 = 0$

Case-2 :- $-(1-0)\log(1-1)$

$\hat{y} = 1$
$y = 0$

$\Rightarrow -1\log(0)$

$= \infty$

When actual = 1

Case 1 :- $\hat{y} = 1$ $(1-y)\log(1-\hat{y}) \Rightarrow (1-1) \longrightarrow 0$

$\hat{y} = 0$ $\quad 0 \rightarrow$

$-y(\log(\hat{y})) \Rightarrow$ Case 1 $\hat{y} = 1$ $-1(\log(1))$ $\;0$

Case 2 $\hat{y} = 0$ $-1(\log(0))$ $\;+\infty$

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

Log-loss

$$\text{loss} = -y\log(\hat{y}) + (1-y)\log(1-\hat{y})$$

$y = 1$

$\Rightarrow$ D.L !!!

$\hat{y}$ P $\longrightarrow$

$$-\left(\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \log(p(y_i)) + (1-y_i)\cdot\log(1-p(y_i))\right]\right)$$

0 or 1

Weather man

	31st of Dec	1st Jan	2nd Jan
Pred.	0	1	1
Act.	1	1	0

Log loss → 1566

Optimize ⇑ Cost

Weather → $\frac{36}{70}$ → Positive Correctly

Internal pur. ← Error

① Optimization

Metrics → Humans

① Persp. → Quantified measure humans use to understand.

Metrics :

Loss / Log

Mean / Loss.

Metrics :-

Loss/Cost
Mean/ Cost

correct

Confusion matrix

Actual values

		Pos	Neg
Predicted	Pos	TP	FP
	Neg	FN	TN

Wrong

correct

Wrong

Very specific to L.R?

All Classification

$$\text{Accuracy} = \frac{TP+TN}{\text{Total values}}$$

1 Churn Prediction.

1 Consumer

2 Telecom

1, 2

Probability

< 25 will not churn

26-50 not lik

50-75 likely

will churn

Actual

Will not | not lik | likel | will

Predict