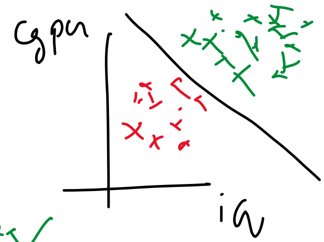
# Introduction

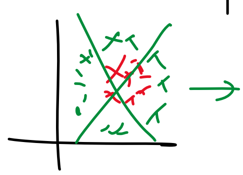
28 June 2023 16:35

Logistic Regression → Interviews
Logistic Regression → Deep Learning

- → classification → categorical → binary 1 0
  - categorical → multi YNM
- → linear model
  - Linear reg
  - logistic → lines/planes → linear | sort of linear
- → multiple interpretation

cgpa | iq | placem

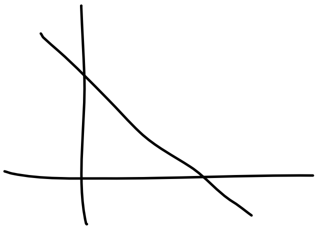


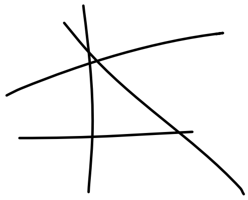
# Some Basic Geometry

28 June 2023 16:41

1. Every line has a positive side and a negative side.



$2x+3y+2=0$



$\boxed{y = mx+b}$

$\downarrow$

$\boxed{Ax+By+C=0} \rightarrow$ general eqn of line

2. How to find out if a given point lies on a given line?

$4x+3y+5=0$ $\quad (5,2)$

$4(5)+3(2)+5=0$

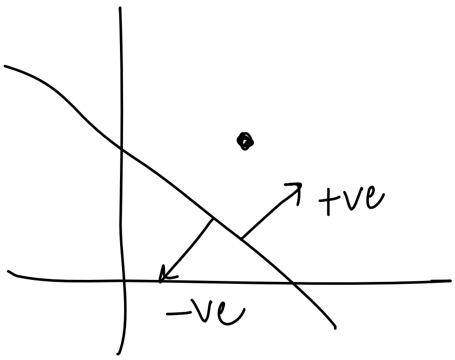
$(0)$

$Ax+By+C=0$

$(x_1,y_1) \rightarrow (x_1,y_1,1)$

$\rightarrow \boxed{Ax_1+By_1+C(1)}$

3. How to find out if a given point is on the positive side of the line or the negative side of the line.




$Ax_1+By_1+C>0 \rightarrow$ +ve region

$Ax_1+By_1+C<0 \rightarrow$ -ve region

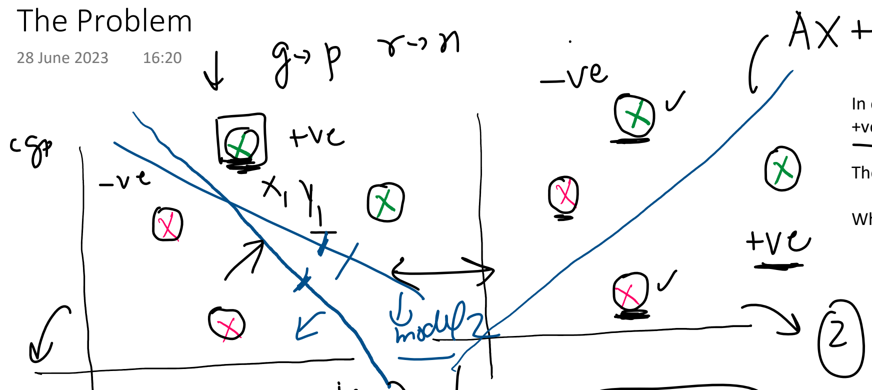
# The Problem

28 June 2023 16:20

In classification problem we have to make sure the +ve points land in the +ve region and -ve point lands in the -ve region

The Algorithm

What is the problem?




$\boxed{\text{model 1}}$

count = 0

correctly classifying all the points

$\boxed{\text{model 2}}$

misclassi

count = 2

loop → (n)

→ $\boxed{\text{green}}$ and $\boxed{Ax_1 + By_1 + c > 0}$ → +ve → +ve correct classi

→ red → -ve

$Ax_1 + By_1 + c > 0$ → +ve → misclass

# misclassifications

counter = 0

→ loop

if pt → red and $Ax_1 + By_1 + c > 0$

misclas → counter + 1 = 1

if pt - green and $Ax_1 + By_1 + C < 0$

mis cl counter

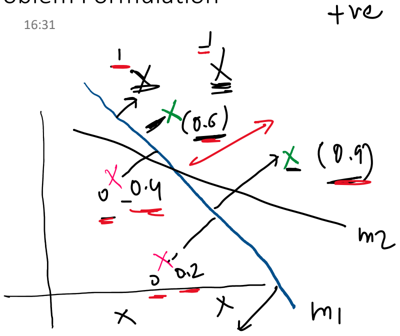
#misclassi

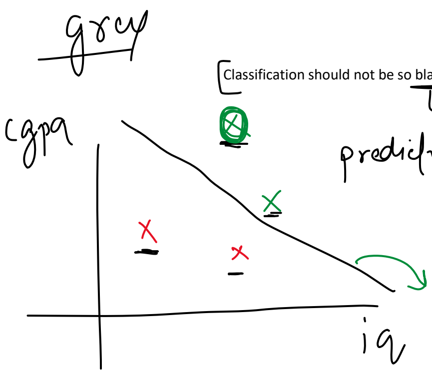
# New Problem Formulation

28 June 2023 16:31

+ve

grey




Classification should not be so black and white

predicting → 1, 0

green red

| iq | cgpa | placed |
|---|---|---|
| 6 | 60 | 0 |
| 4 | 40 | 0 |
| 8 | 80 | 1 |
| 9 | 90 | 1 |

$Ax + By + C = 0 \rightarrow \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$3x + 5y + 6 = 0$

$\{9, 90\}$ → $x_1$, $x_2$

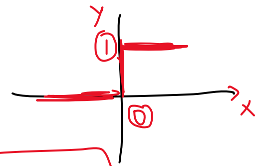
$3 \times 9 + 5 \times 90 + 6$

$150 > 0 \rightarrow 1$

$< 0 \rightarrow 0$

$\boxed{\beta_0 + \beta_1 x_1 + \beta_2 x_2} = z$

Step $\boxed{\begin{array}{ll} z > 0 & 1 \\ z < 0 & 0 \end{array}}$




$z \rightarrow \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$\boxed{\text{Step}(z)} \rightarrow 0, 1$

$\boxed{1, 0}$

$z \uparrow$ 0.9

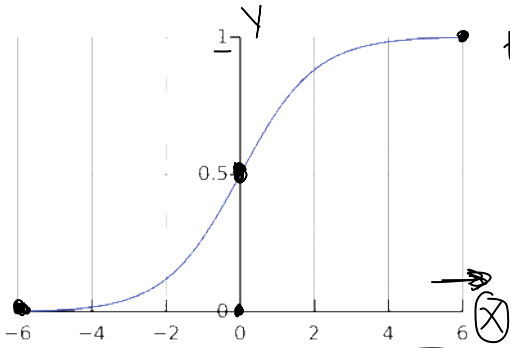
$z \uparrow$ 0.6

$z \downarrow$ 0.1

# Sigmoid Function

28 June 2023 16:32

step → sigmoid (σ)




tries to touch 1

X → ∞

Y → 1

X → 0

Y → 0.5

$$\boxed{Y = \frac{1}{1+e^{-x}}}$$

sigmoid

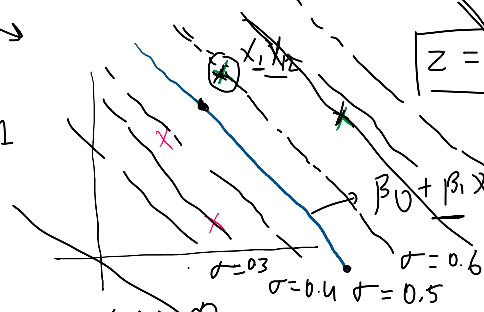
Z >> 0

Z > 0 = σ(z) = 0.5 →

0.

$$\boxed{Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2}$$

Z = 0




σ(z)

↓

0.5

0 → 1

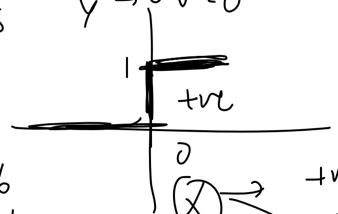
X → −∞

Y → 0 σ = 0

X +ve Y → 1

X -ve Y → 0

$\boxed{\text{step}}$



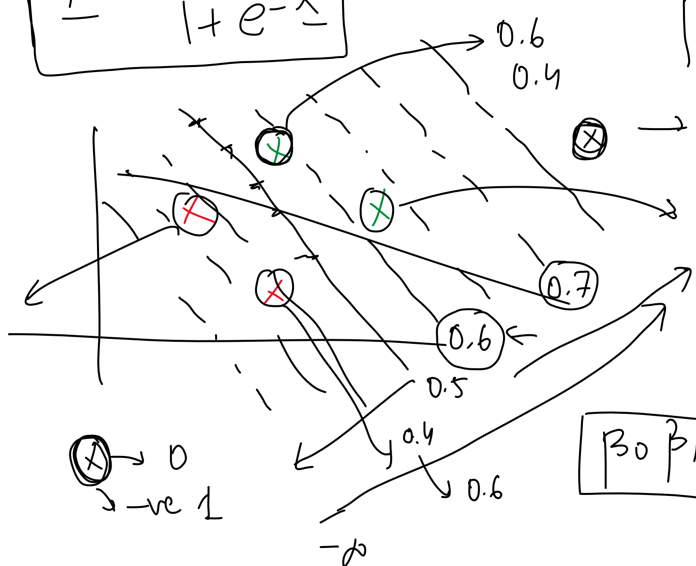

X → +ve (1)

→ -ve 0

X → +ve hinga

probabiblity

inter




0.3

↓

0.7

X → 0

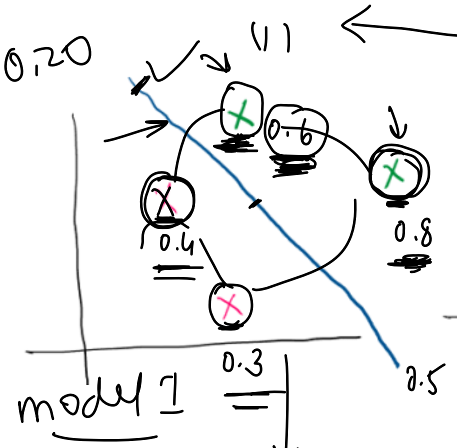
→ -ve 1

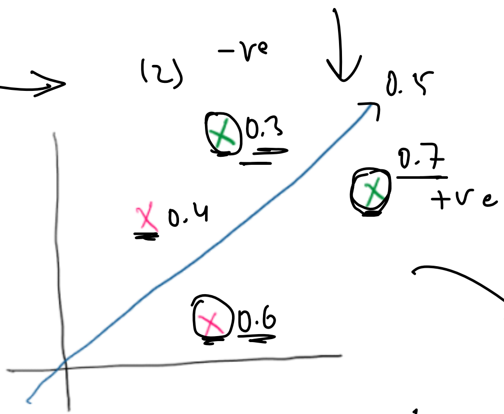
σ = 0

$\boxed{\beta_0 \; \beta_1 \; \beta_2}$

# Maximum Likelihood

28 June 2023 16:32




The likelihood function is the product of the predicted probabilities for the actual class of each observation

$0.7 \times 0.4 \times 0.3 \times 0.6$

0.05

likelihood

$\boxed{0.8 \times 0.6 \times 0.7 \times 0.6}$ → maximum

→ best line

→ logistic ln

$\boxed{0.20}$ → maximum

→ better model

$$Y = \frac{1}{1+e^{-x}}$$

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

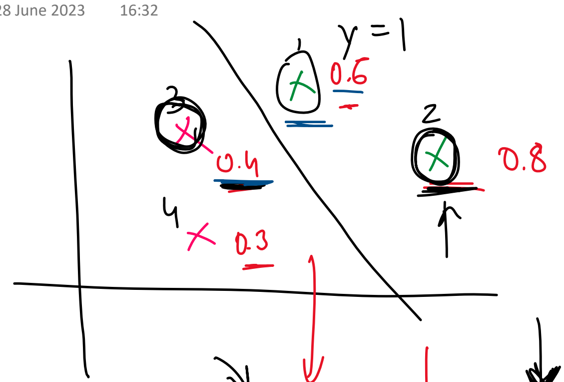
$$\boxed{Y} = \boxed{\frac{1}{1+e^{-\beta_0+\beta_1 x_1+\beta_2 x_2}}}$$

↓

$\boxed{0-1}$

# Log Loss

28 June 2023 16:32




1000's

y

→ 4 points

output → 1

$y_3 = 0$

$\boxed{0.20}$

$$ML = \boxed{0.8 \times 0.6 \times 0.6 \times 0.7} \rightarrow \text{problem}$$

small ≃ 0

↳ underflow

maximum log

↳ maximum

$$\log(ML) = \log(0.8 \times 0.6 \times 0.6 \times 0.7)$$

$$= \log 0.8 + \log 0.6 + \log 0.6 + \log 0.7$$

$$= \boxed{-\log 0.8 - \log 0.6 - \log 0.6 - \log 0.7}$$

positive minimum

minimum

minimize

$\log(0-1)$

-ve number

$0.1 \rightarrow 1$

$0.5 \rightarrow 0.3$

$$\boxed{\boxed{-\log(\hat{y}_1)} - \log(\hat{y}_2) - \log(\hat{y}_3) - \log(\hat{y}_4)}$$

NO X

$\hat{y}_1 = \sigma(z) \rightarrow (P) \rightarrow$ getting green

↳ $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$\left.\begin{array}{l}\hat{y}_1 \rightarrow P_1 \\ \hat{y}_2 \rightarrow P_2\end{array}\right\}$ getting green

$$\boxed{-y_3 \log \hat{y}_1 - (1-y_2)\log(1-\hat{y}_1)}$$

0

$(1-0)\log(1-\hat{y}_3)$

$$\boxed{-\log \hat{y}_1} - \log \hat{y}_2 - \log(1-\hat{y}_3) - \log(1-\hat{y}_4)$$

$$\boxed{-\log \hat{y}_1} - \log y_2 - \log(1-\hat{y}_3) - \log(1-\hat{y}_4)$$

$1-\hat{y}_i = p(\text{red})$

$$\frac{1}{n}\sum_{i=1}^{n} -y_i \log \hat{y}_i - (1-y_i)\log(1-\hat{y}_i)$$

$\cancel{\hat{y}_i = p(\text{green})}$

$\hat{y}_i = p(\text{green})$

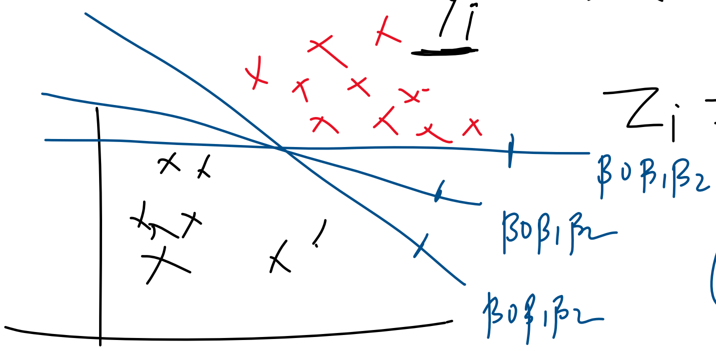
$$\boxed{L = -\frac{1}{n}\sum_{i=1}^{n} y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)}$$

→ log loss error

→ binary cross entropy

$\beta_0 \quad \beta_1 \quad \beta_2 \rightarrow L$ $\boxed{\text{minimise}}$

$$\hat{y}_i = \sigma(z_i)$$




$$z_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

min → logistic regssion

# Gradient Descent

28 June 2023 16:32

min

$$\boxed{L = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)} \rightarrow \text{Closed form solution}$$

mse →

$\beta_0\ \beta_1\ \beta_2$

| $x_1$ | $x_2$ | $y$ | $\hat{y}_i$ | |
|---|---|---|---|---|
| 28 | 32 | 0 | → 0.63 | → 0.37 |
| 15 | 12 | 1 | → 0.37 | → 0.63 |
| | | 0 | → 0.11 | → 0.89 |

100 = n

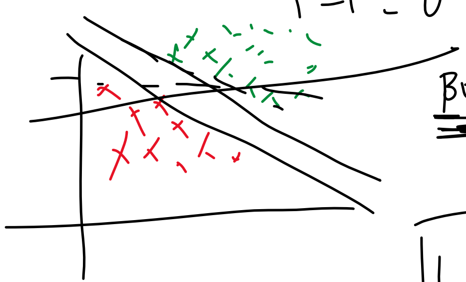
$\hat{y}_i \rightarrow P(\text{green})$

$(1-\hat{y}_i) \rightarrow P(\text{red})$

$1 - 0 = 1$

$1 - 1 = 0$

$\beta_0\ \beta_1\ \beta_2$



## Gradient Descent

$\beta_0\ \beta_1\ \beta_2 \rightarrow L$ min

$$\beta_0 = \beta_0 - \eta \boxed{\frac{\partial L}{\partial \beta_0}} \rightarrow \text{gradient}$$

learning rate

$$\boxed{L \rightarrow \frac{\partial L}{\partial \beta_0}}$$

$$\boxed{\begin{aligned}\beta_1 &= \beta_1 - \eta\frac{\partial L}{\partial \beta_1}\\ \beta_2 &= \beta_2 - \eta\frac{\partial L}{\partial \beta_2}\end{aligned}}$$

$$\boxed{\frac{\partial L}{\partial \beta_0} \quad \frac{\partial L}{\partial \beta_1} \quad \frac{\partial L}{\partial \beta_2}}$$

$\hat{y}_i = P(\text{green})$

$$\frac{\partial L}{\partial \beta_1} = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)$$

$n \rightarrow$ points

1 point

$\frac{\partial}{\partial x}\sigma(x) \rightarrow \sigma(x)[1-\sigma(x)]$

$\log x \rightarrow \frac{1}{x}$

$$\frac{\partial L}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left[-\boxed{y\log\hat{y}} - (1-y)\log(1-\hat{y})\right]$$

$\hat{y} = P(\text{green})$

$\sigma(z)$

$\beta_1$

$$-y\ \sigma(z)[1-\sigma(z)]\,x_1$$

$$\frac{-y}{\hat{y}} \sigma(z)[1-\sigma(z)] x_1$$

$\sigma(z)$ $\beta_1$

$\beta_0 + \boxed{\beta_1 x_1} + \beta_2 x_2$

$$\frac{\partial \hat{y}}{\partial \beta_1} \quad \frac{\partial \sigma(z)}{\partial \beta_1}$$

sig $-$ sig $(1-$sig$)$

$$\boxed{\frac{-y}{\cancel{\hat{y}}} \cancel{\hat{y}} (1-\hat{y}) x_1} \quad \boxed{-y(1-\hat{y}) x_1}$$

$z \to \beta_1 x_1$

$$\frac{\partial}{\partial \beta_1} - (1-y) \log(1-\hat{y}) = \frac{(1-y)}{\cancel{(1-\hat{y})}} \hat{y}\cancel{(1-\hat{y})} x_1$$

$$\boxed{(1-y)\hat{y} x_1}$$

$$-y(1-\hat{y}) x_1 + (1-y)\hat{y} x_1$$

$$[-y + \cancel{y\hat{y}} + \hat{y} - \cancel{y\hat{y}}] x_1$$

$$\boxed{\frac{\partial L}{\partial \beta_1} \quad (\hat{y}_i - y_i) x_{1i}} \to \frac{\partial L}{\partial \beta_2} = (\hat{y} - y) x_2$$

$$\frac{\partial L}{\partial \beta_0} = (\hat{y} - y) \longrightarrow$$

# Summary

28 June 2023 16:35

→ sigmoid

→ maximum likelihood

→ log → binary crossentropy → log loss

→ gradient desc → $\beta_0 \; \beta_1 \; \beta_2$ → L min