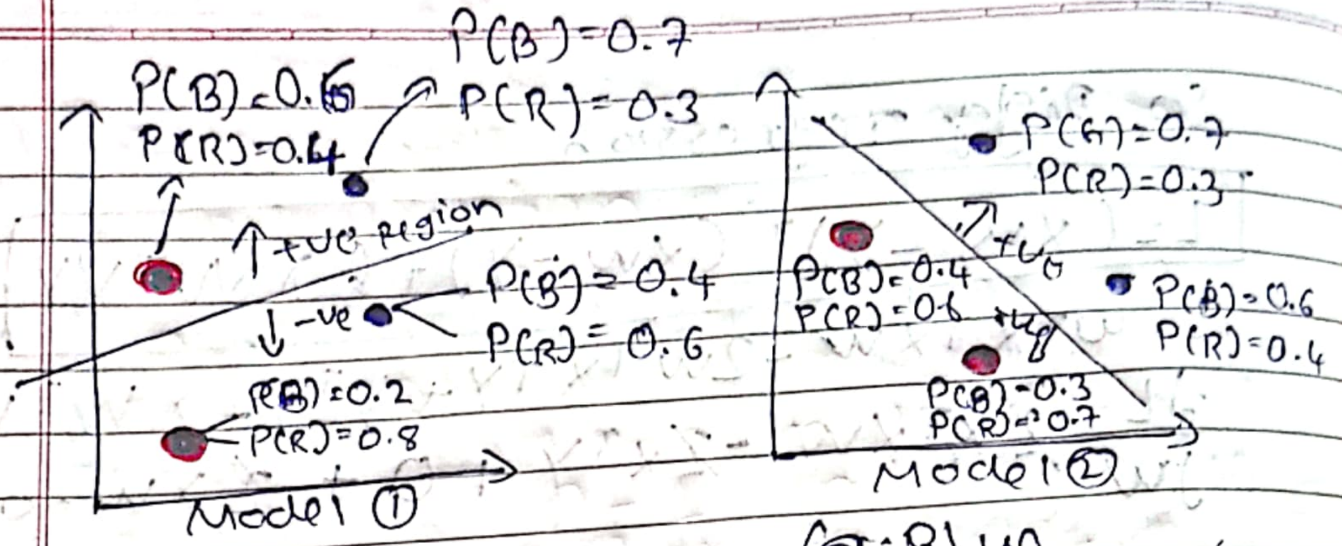
Maximum Likelihood



P(B) = Probability of Blue

P(R) = Probability of Red.

~~Maximum Likelihood says~~

* If you want to find which model is better than the others using Maximum Likelihood. (calculate product)

Maximum Likelihood says multiple all the prediction of the model and then compare that with the other model.

whichever model has higher value that model is the better model.

* Calculating Maximum Likelihood of model ①

> Note: In maximum likelihood we only consider that value of point which is Actual its colour/charactor) Blue meaning IF the point is Blue we only only consider P(B) not the P(R)

① Calculating ~~Ⓘ~~ Maximum L q Model ①

As we can see their are 2 Red and 2 Blue points.

$= 0.7 \times 0.4 \times 0.4 \times 0.8$

ML1 $= 0.089$

② Calculating ~~M~~ M L q Model ②

2 Blue , 2 Red

$= 0.7 \times 0.6 \times 0.6 \times 0.7$

ML2 $= 0.176$

ML 2 > ML 1

meaning Model 2 is Better than Model 1

But the problem with this method is if there is a large Dataset with large no. of Rows and Columns the product (Multiple) get too small. So, some how we have to convert this product (x) to sum (+).

So As we know

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

Now

$\log(\text{max}) = \log(0.7) + \log(0.4) + \log(0.4) + \log(0.8)$

Page No.
Date

But we still have problem with this equation

$\log(Max) = \log(0.7) + \log() + \log() + \log()$.

Because the Maximum likelihood Range is 0 - 1 and for every value from 0-1 there logs is ~~M~~ $\boxed{\text{NEGATIVE}}$

So we have to change positive to -ve this called as Cross Entropy.

The summation of ~~...~~ Negative logs of Maximum likelihood is called as Cross Entropy.

Now,
$\log(min) = \boxed{-\log(0.7) - \log(0.4) - \log(0.4) - \log(0.8}$

In Maximum likelihood we choose ~~...~~ Better Model by comparing ~~...~~ product whoever model have larger product than the other that model is Better.

But,
In ~~...~~ Cross Entropy we ~~...~~ minimize the value because
$\log 0.1$ value $> \log 0.9$ value...
$1 > 0.04$

In Maximum L we Maximize the Value and in Cross Entropy we Minimize the Value.

Now, we will choose that model whose cross Entropy is less than the other model's cross Entropy

Formula is

① $= -y_i \log(\hat{y}_i) - (1-y_i)\log(1-\hat{y}_i)$

Putting point in the formula.

$P(B) = 0.7$
$P(R) = 0.3$

For this

Blue point correctly classified as Blue (1) ($y_i = 1$)

Target value

Predicted value

$= -y_1 \log(\hat{y}_1) - (1-y_1)\log(1-\hat{y}_1)$

$= -y_1 \log(\hat{y}_1) - (1-1)\log(1-\hat{y})$

$= -y_1 \log(\hat{y}_1) - 0$

$= -1 \log(\hat{y}_1)$

$= \boxed{-\log(0.7)}$

we are seeing that what the probability of getting Blue that why in $\hat{y}$ we are putting Blue value

(2) Now for the 2nd point putting in the formula.

$P(B) = 0.6$ $\quad y_2 = 0$ ← Target Value

$P(R) = 0.4$

So,

$= -y_2 \log(\hat{y}_2) - (1-y_2)\log(1-\hat{y}_2)$

$= -\underbrace{0 \log(\hat{y}_2)}_{0} - (1-y_2)\log(1-\hat{y}_2)$

$= -(1-y_2)\log(1-\hat{y}_2)$

$= -(1-0)\log(1-\hat{y}_2)$

$= -1\log(1-\hat{y}_2)$

$= -\log(1-0.6)$

$= \boxed{-\log(0.4)}$

(3). 3rd point

$P(B) = 0.4$

$P(R) = 0.6$ $\quad y_3 = 1$

$= -y_3 \log(\hat{y}_3) - (1-y_3)\log(1-\hat{y}_3)$

$= -y_3 \log(\hat{y}_3) - 0$

$= -y_3 \log(\hat{y}_3)$

$= -1\log(\hat{y}_3) = \boxed{-\log(0.4)}$

$= -\log(0.4)$

(4) 4^{th} point

$P(B) = 0.2$

$P(R) = 0.8$

$y_4 = 0$

$= -y_4 \log(\hat{y}_4) - (1 - y_4) \log(1 - \hat{y}_4)$

$= -0 \log(\hat{y}_4) - (1 - 0) \log(1 - \hat{y}_4)$

$= 0 - (1 - 0) \log(1 - \hat{y}_4)$

$= -\cancel{(1 \log \hat{y}_4)}$

$= -1 \log(1 - \hat{y}_4)$

$= \log(1 - 0.2)$

$- \log(0.8)$

$= -\log(0.7) - \log(0.4) - \log(0.4) - \log(0.8)$

Our Loss function is

$$\boxed{L = -y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i)}$$

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

For Average Loss / Error

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

These loss func also known as $\boxed{\text{Log-Loss Error}}$ and $\boxed{\text{Binary Cross Entropy}}$