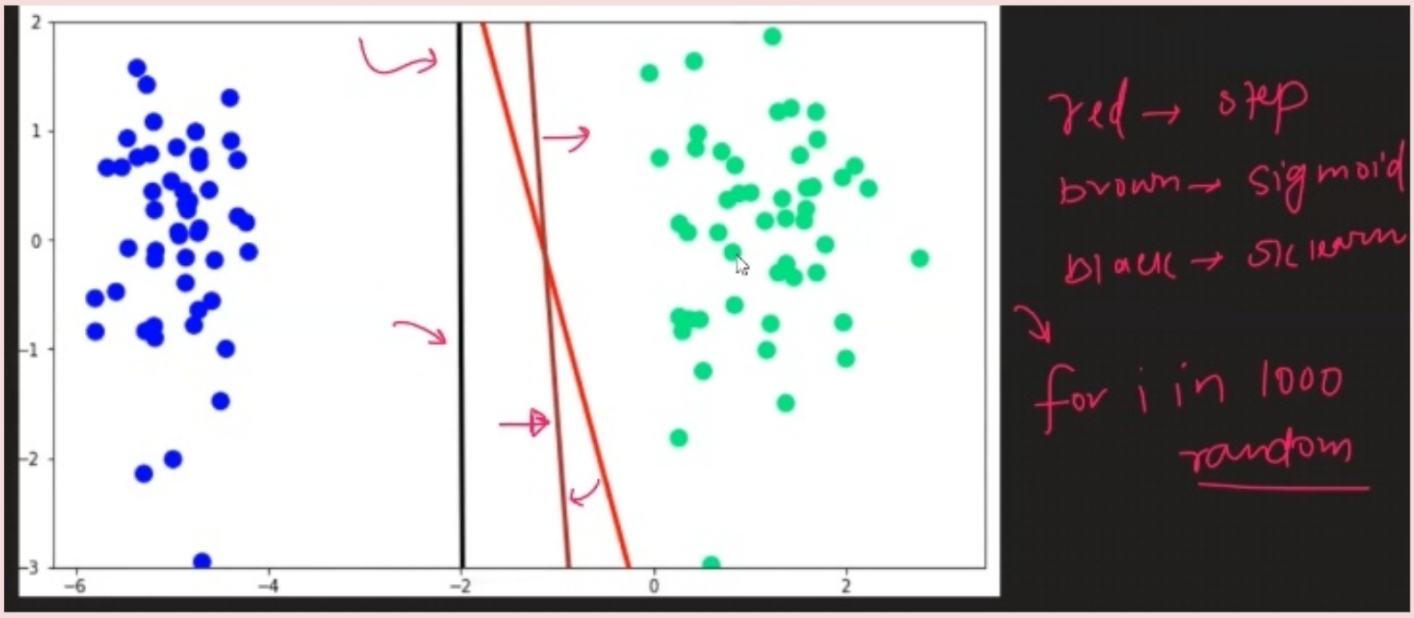



→ Here is our Previous approaches result.

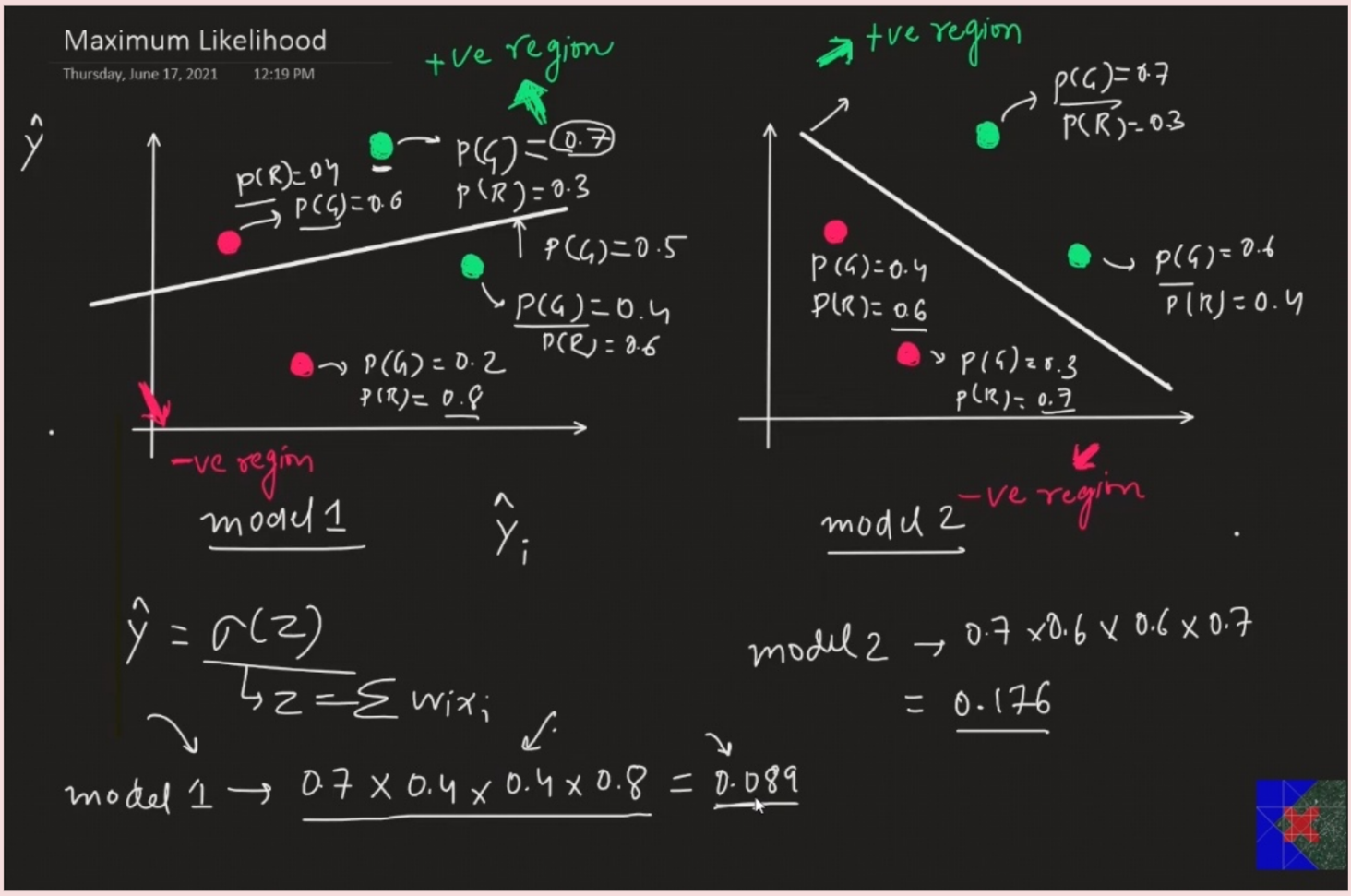
→ We can see we are not at the optimal solution yet.

→ Those aproach can't be the optimal solution.

→ Cause we pick point random and none aproach says that the approach search for the optimal solution.

→ That's why we will use lost function to find the optimal solution.

→ now we will use maximum likelihood to find the best solution.




→ First, we build a model and calculate the likelihood that is how likely it performing on our data.

→ We will find the likelihood value by multiplying the sigmoid function value for each point.

→ we will pick prediction of being 1 for 1 and prediction of being 0 for 0.

→ Then we will get the likelihood value and we can pick the maximum one and that will be our solution.

→ There is a problem in multiplying so many small values together.

→ The value becomes too small and it is difficult to find the maximum number from it.

→ Mean, it won't be efficient.

→ so, we will introduce log here to perform summation.

→ we know, $\log(ab) = \log a + \log b$

model 1 → $0.7 \times 0.4 \times 0.4 \times 0.8 = 0.089$

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

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

$0-1 = -ve$

cross entropy ~~minimize~~ max*

NO

$\log(0.1) > \log(0.01)$

$\Downarrow \qquad \Downarrow$

$-1 \qquad -0.04$

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

→ Another problem is, log of 0-1 value is always negative.

→ so instead of summation we will add -ve to each term.

→ And we can't directly add $\hat{y}$ to our formula, cause we are not taking $\hat{y}$ everytime, sometimes we are

taking $(1-\hat{y})$ where the point is 0 on Red.

→ So, we need to find a solution for that.

→ So, we will use:

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

→ How it works?

→ For 0 value the left part of the equation becomes 0 and we have, $\log(1-\hat{y}_i)$, $(1-\hat{y}_i)$ means probability of being 0.

→ For 1, the right part becomes 0 and we have $\log(\hat{y}_i)$.

→ Finally for the avg. error we

have the loss function.

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

→ It is also called,

* log loss error function
* Binary cross entropy function