1] Gini Index or Gini Impurity :-

$$= 1 - \sum p_i^2$$

$$GI = 1 - \left[P(Y)^2 + P(N)^2 \right]$$

$$= 1 - P(Y)^2 - P(N)^2$$

1] Sygon

$$GI = 1 - \left[(5/5)^2 + (0/5)^2\right]$$

$$= 1 - \left[(1)^2 + 0\right]$$

$$= 1 - 1$$

$$GI = 0$$

11] 5y & 5N

$$= 1 - \left[(5/10)^2 + (5/10)^2 \right]$$

$$= 1 - [0.25 + 0.25]$$

$$= 1 - 0.5$$

$$\text{ans} = 0.5$$

iii) 3Y & 2N

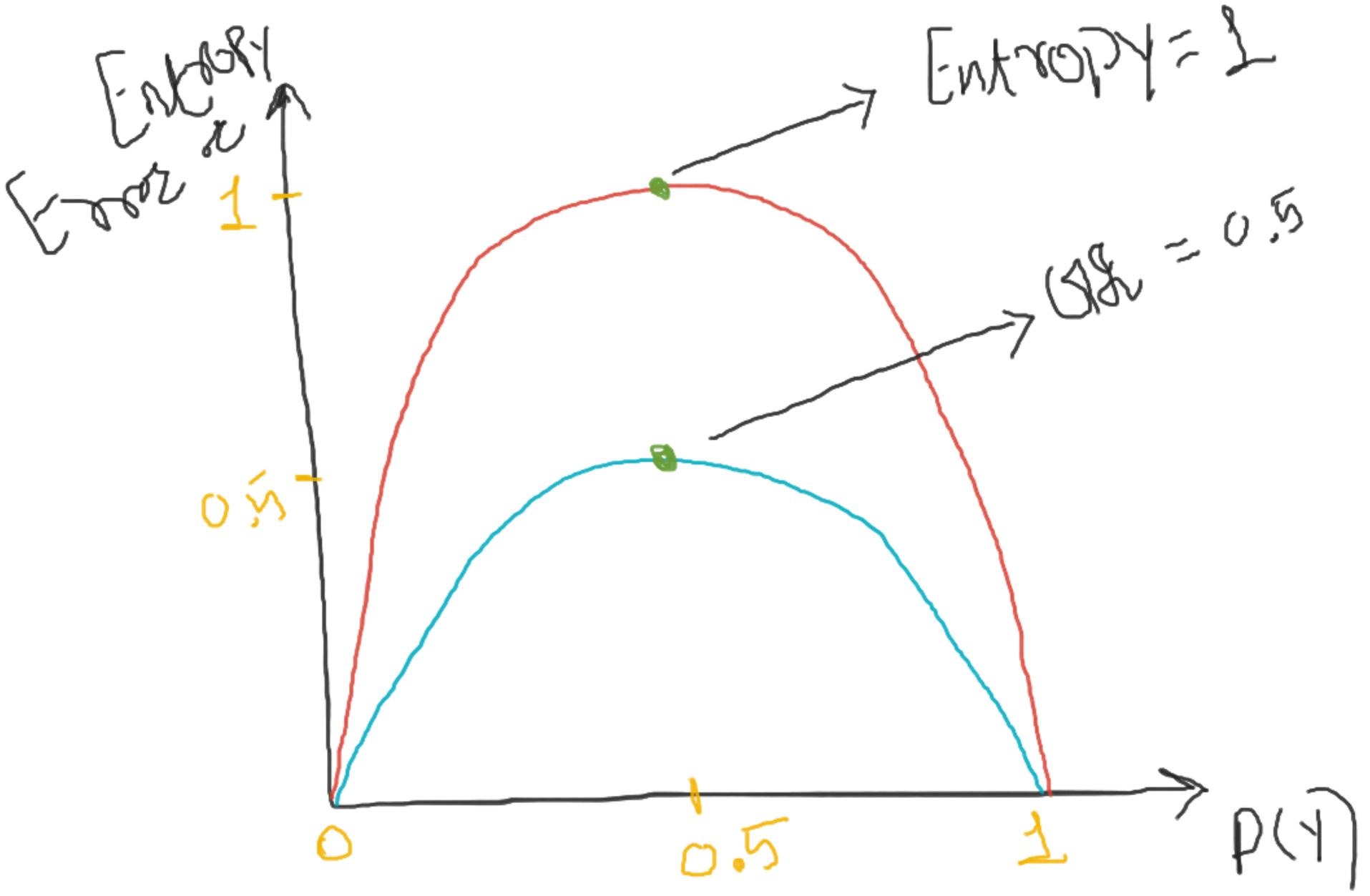
$$= 1 - \left[(3/5)^2 + (2/5)^2\right]$$

$$= 1 - \left[(0.36) + (0.16)\right]$$

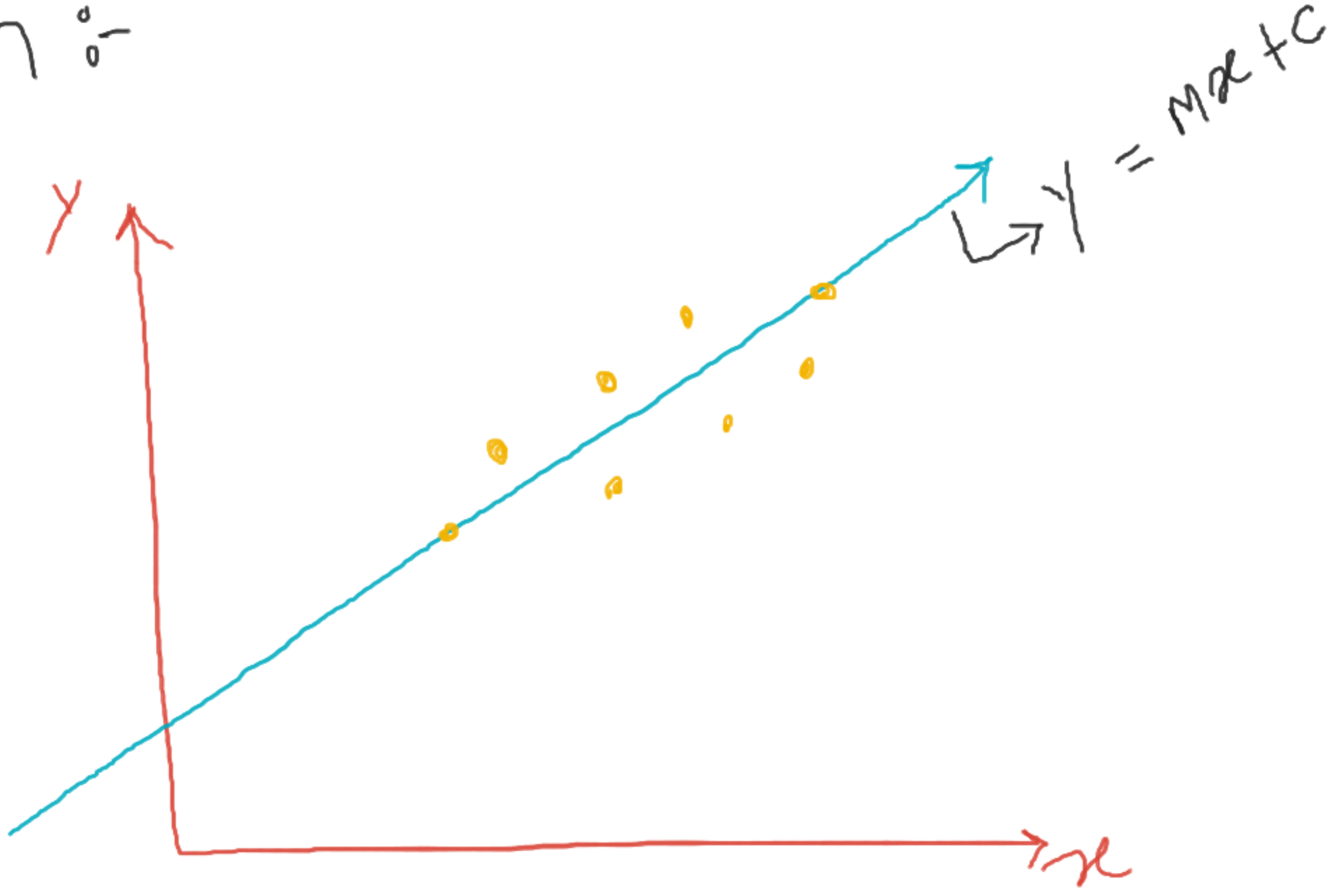
$$= 1 - 0.52$$

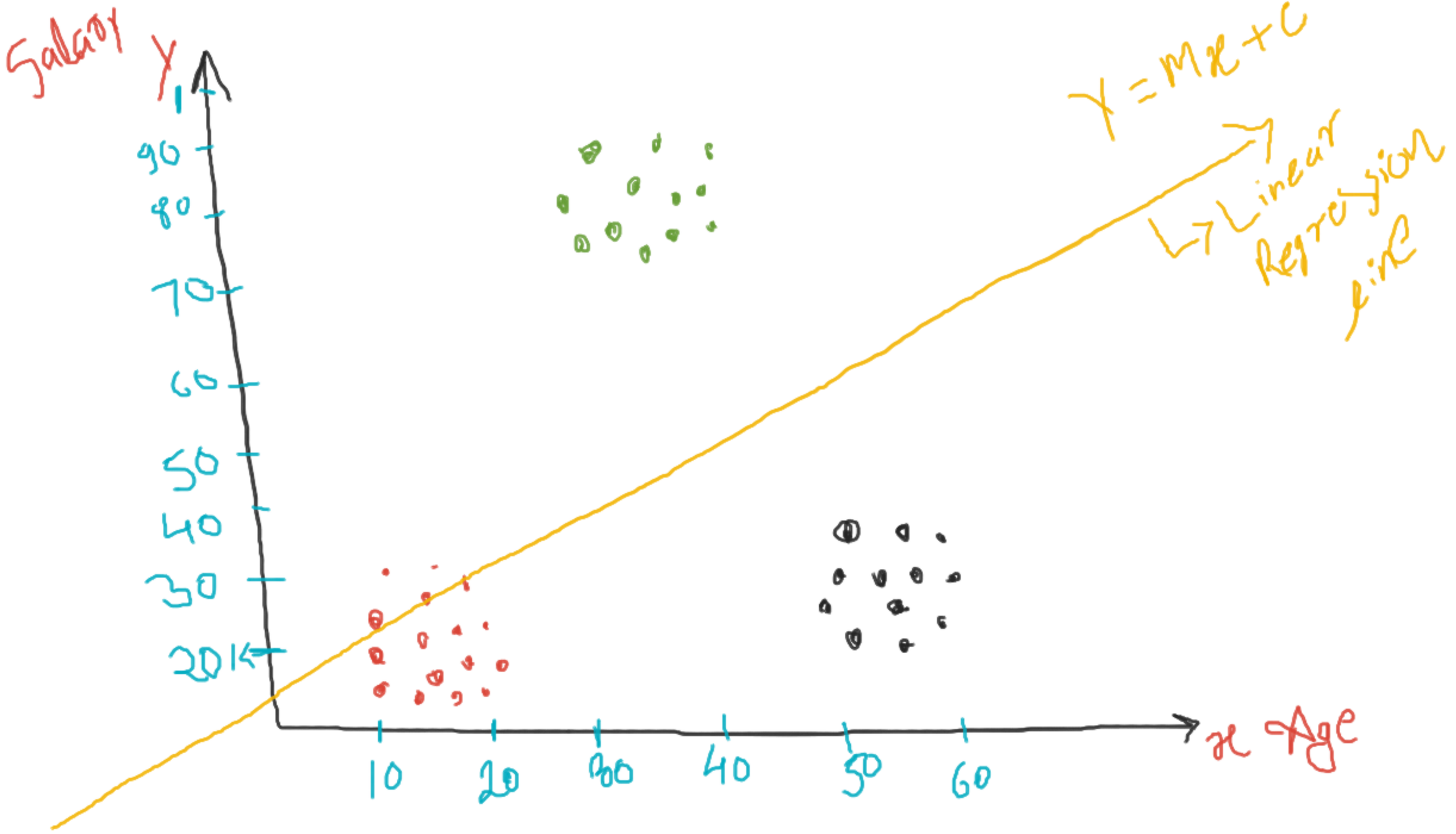
$$= 0.48$$

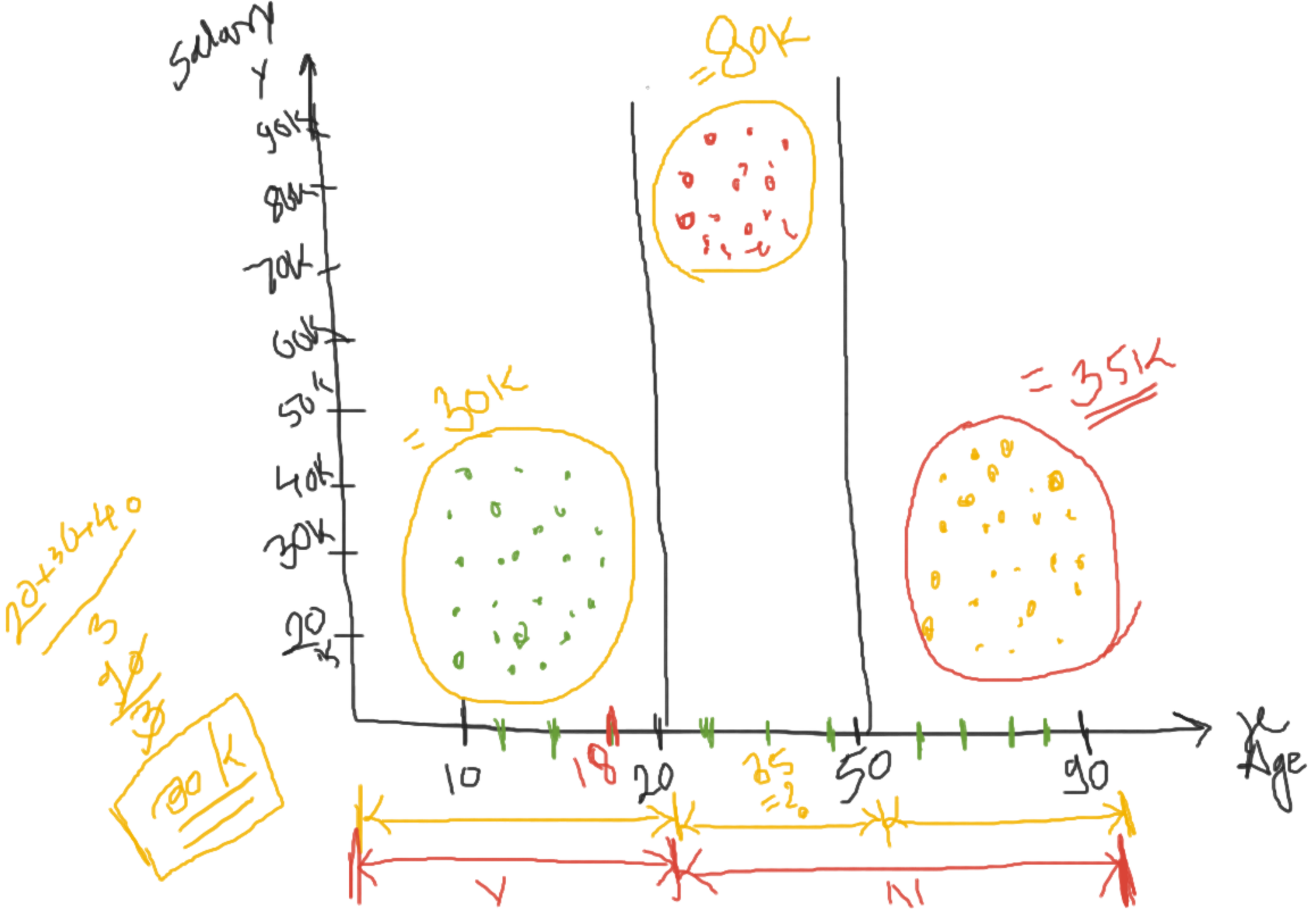
Ans =

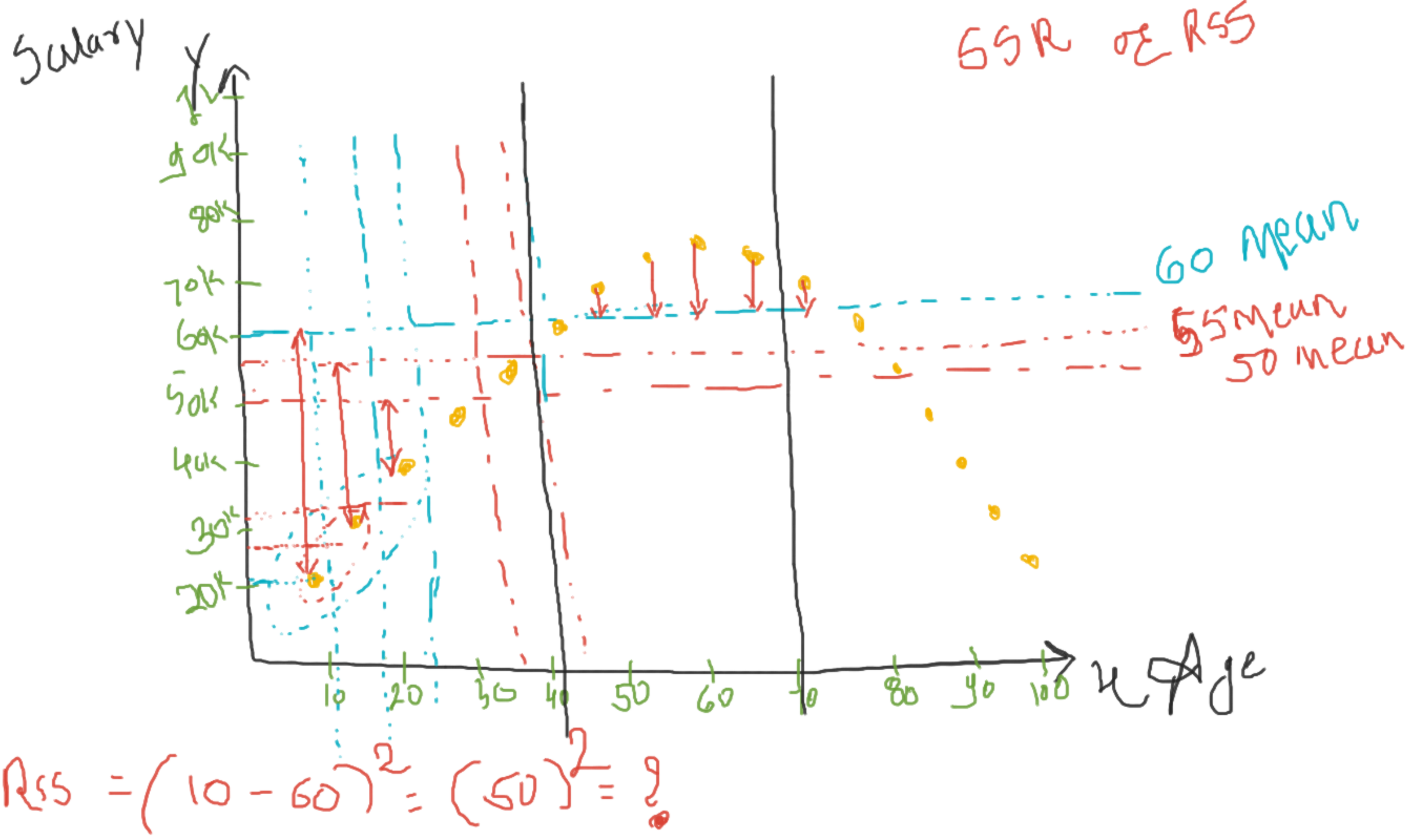

Entropy
or
Error
1
0.5
0
0.5
1
p(+)
Entropy = 1
Gini = 0.5

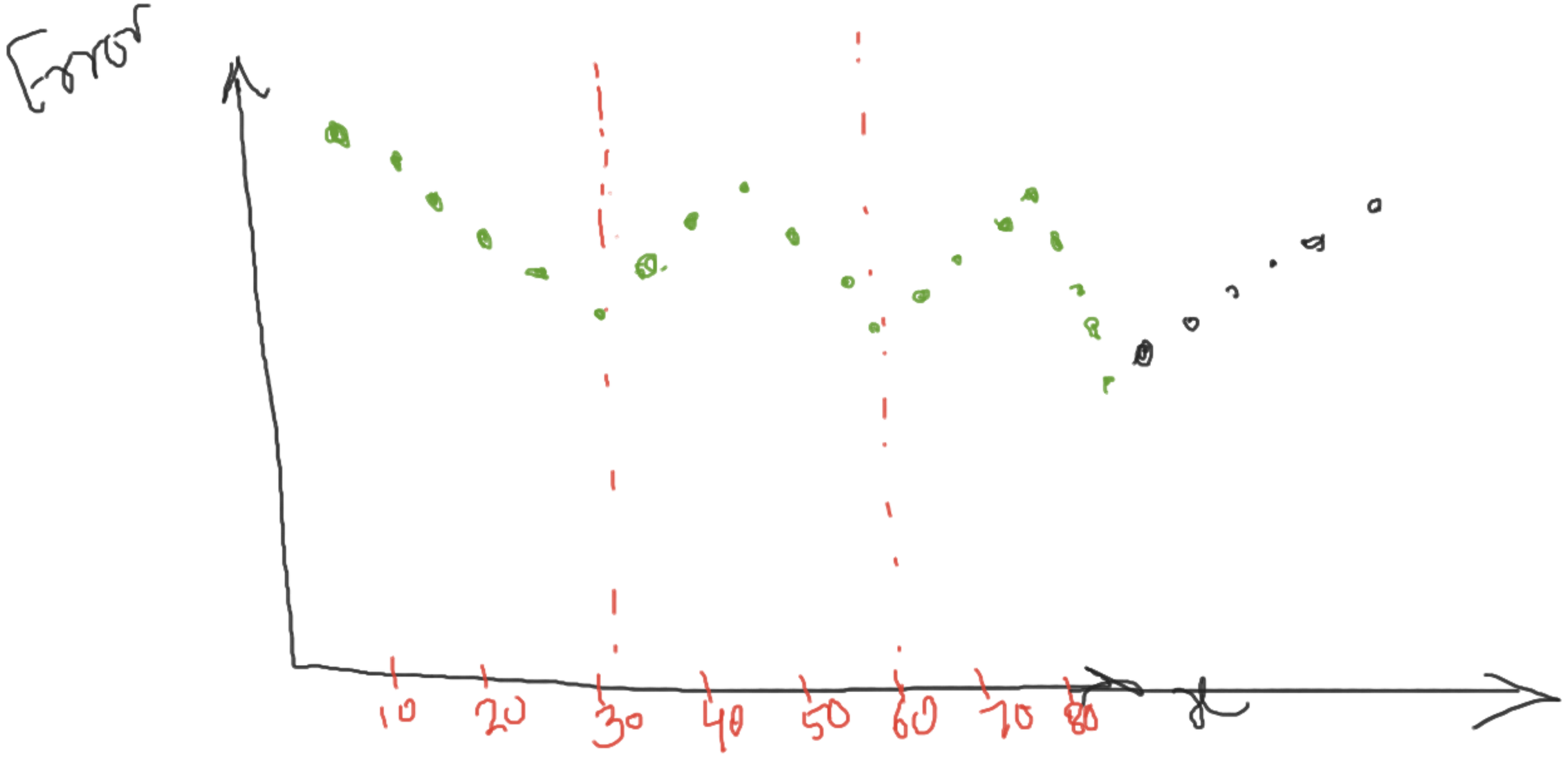
1] Regression :-

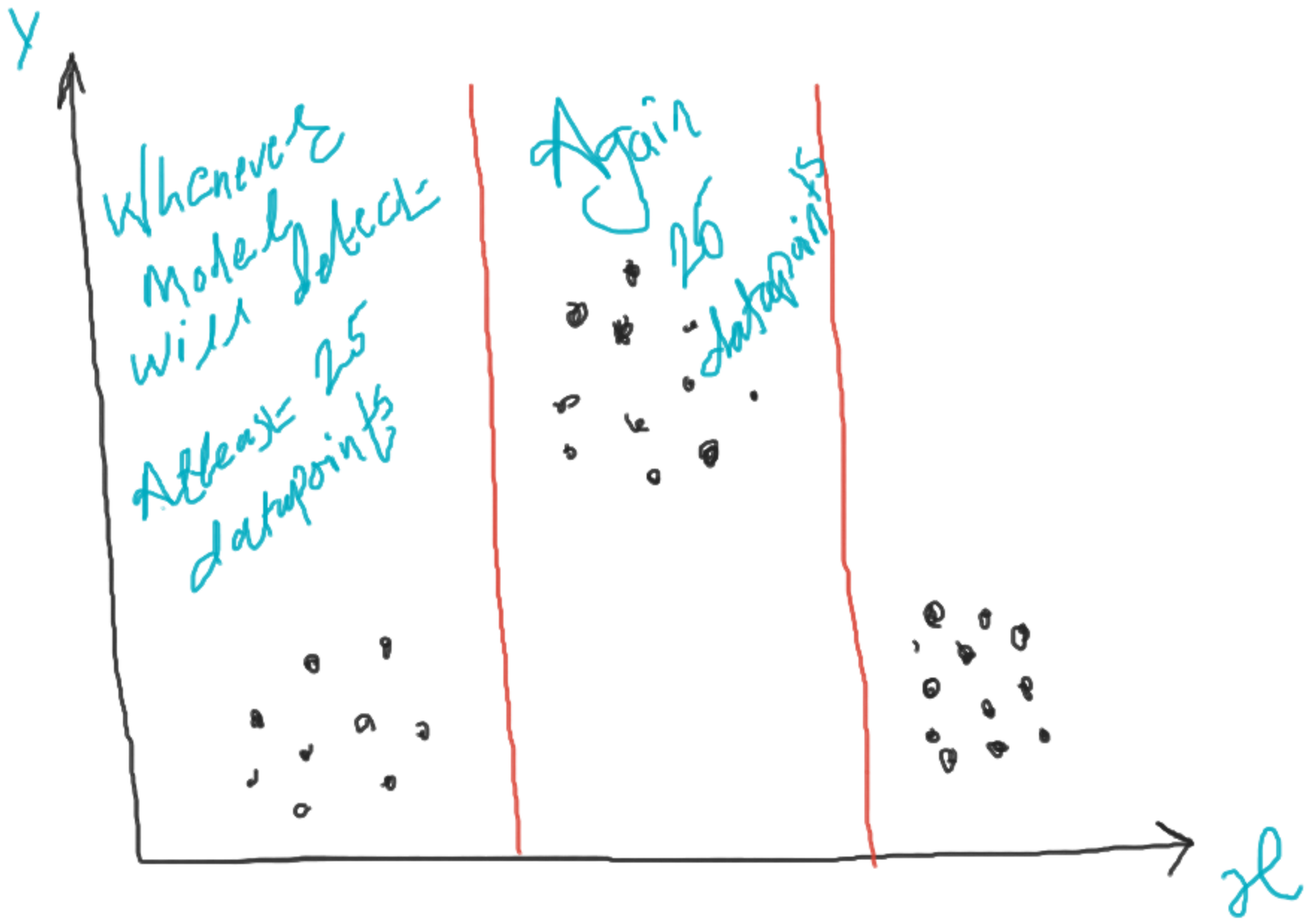



Salary
y
90
80
70
60
50
40
30
20
10
20
30
40
50
60
x Age
Y = mx + c
Linear Regression line


Salary
90K
80K
70K
60K
50K
40K
30K
20K
80K
30K
35K
10
18
20
25
50
90
Age
Y
N
20+36+40
3
96/3
80K


Salary
Y
100k
90k
80k
70k
60k
50k
40k
30k
20k
10
20
30
40
50
60
70
80
90
100
X Age
60 mean
55 mean
50 mean
55 R of RSS
RSS = (10 - 60)^2 = (50)^2 = ?


Error
10
20
30
40
50
60
70
80


y
Whenever Model will detect
Atleast 25 datapoints
Again 25 datapoints
x