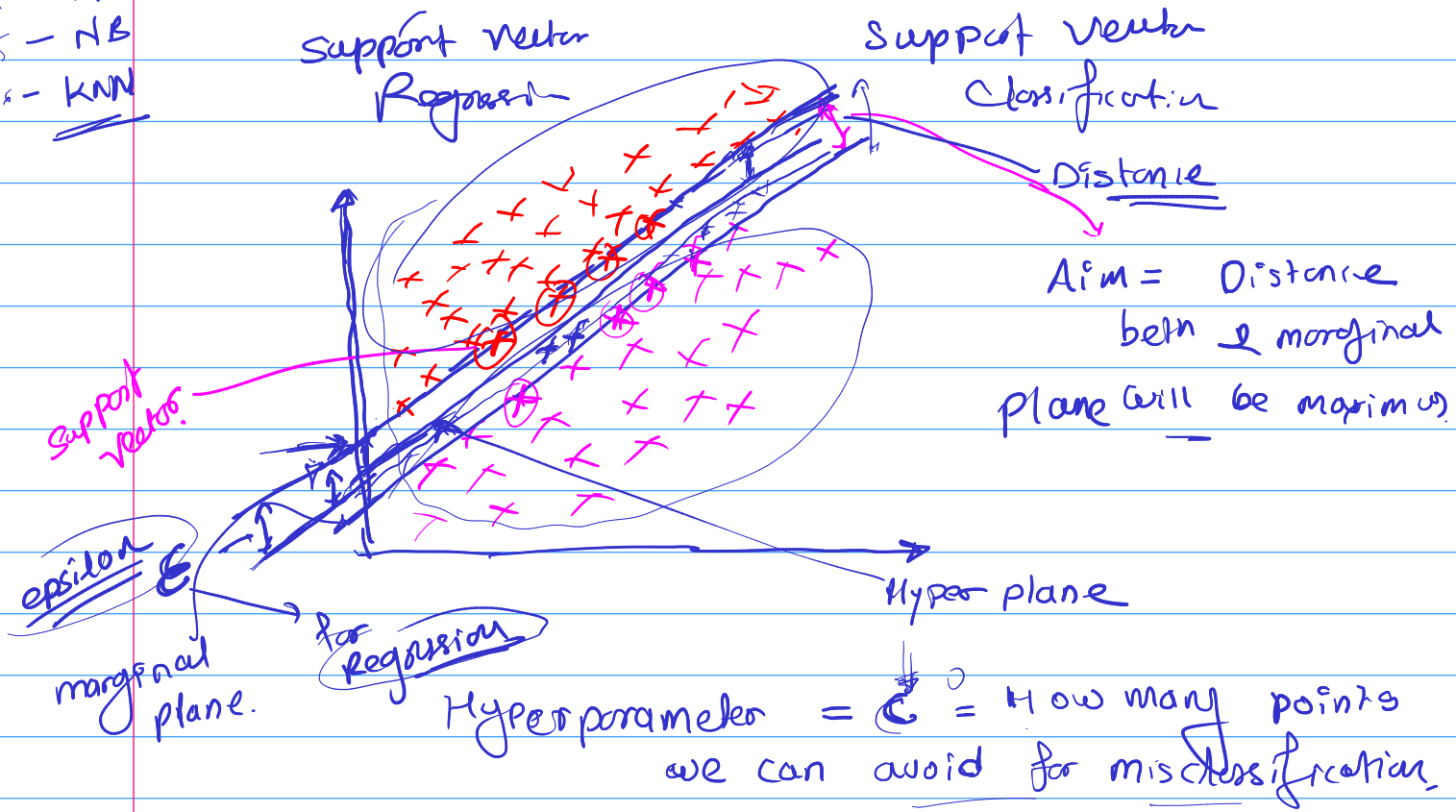


- 1 - LR
- 2 - LogR
- 3 - Poly
- 4 - L1 & L2
- 5 - NB
- 6 - KNN

⑦ SVM = Support Vector Machine.

SVR — SVC



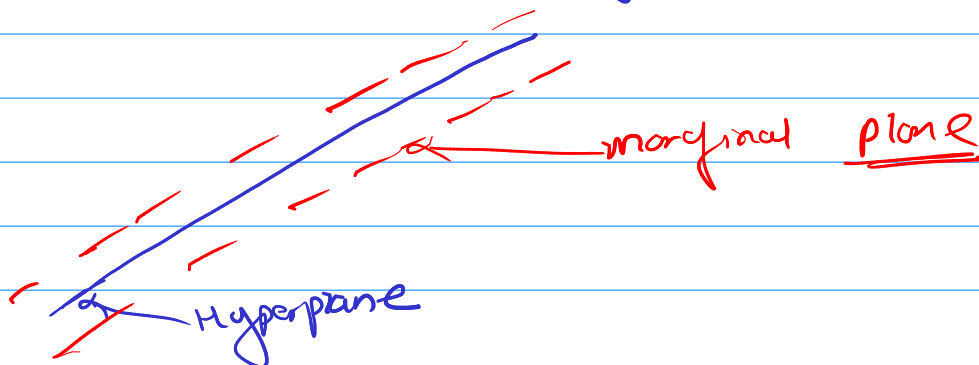
$$\frac{mx+c}{mx+c+\epsilon} = 0$$

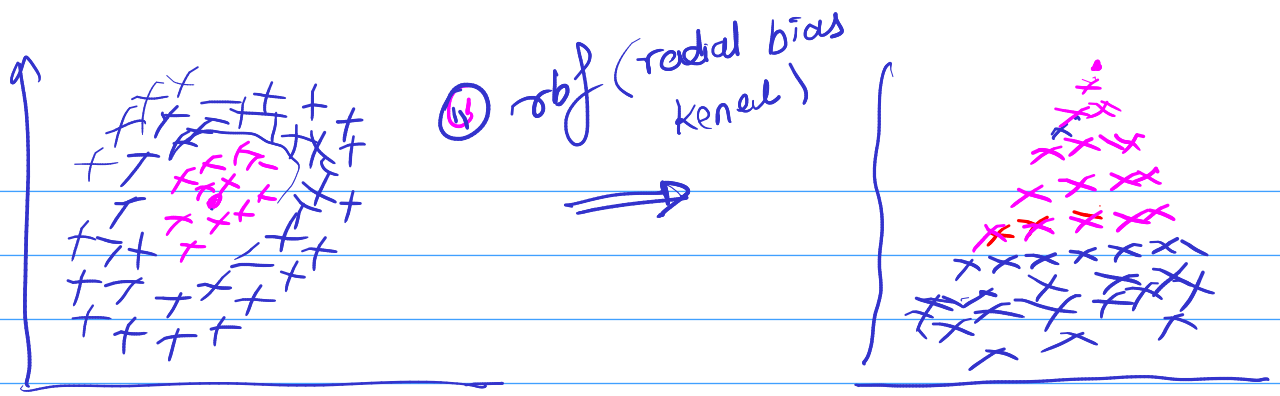
η (eta) = distance of misclassified from marginal plane

classification

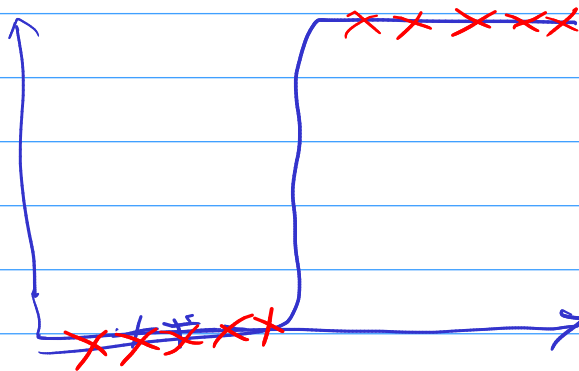
SVM - kernel \rightarrow (Transformation of data)

Aim \rightarrow Hyper plane & marginal plane

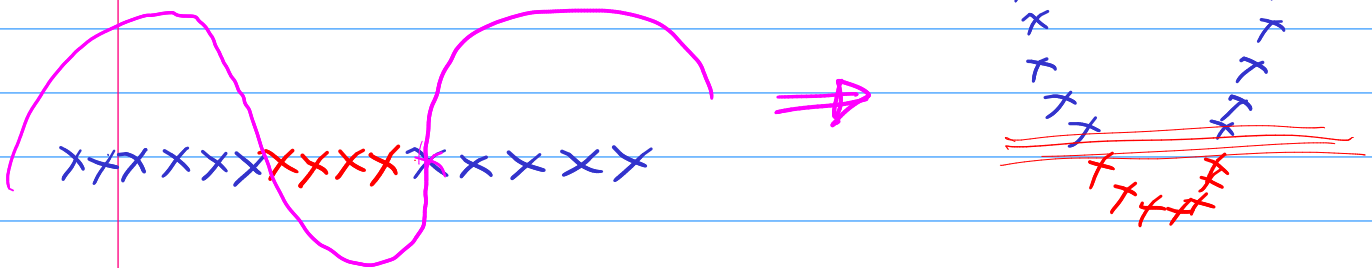




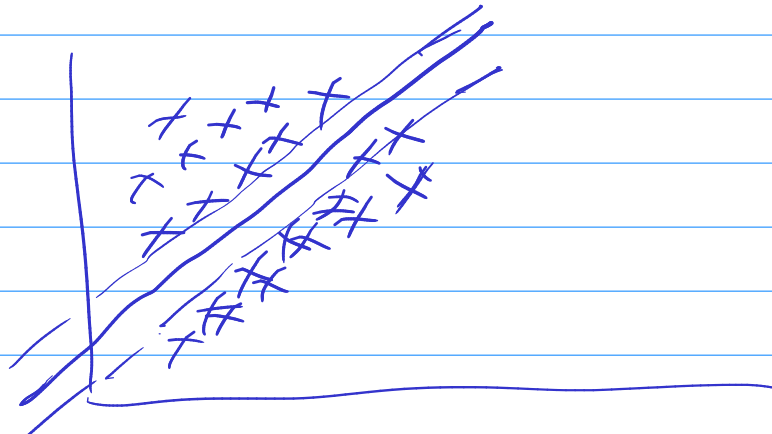
② Sigmoid



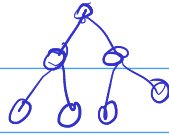
③ poly kernel



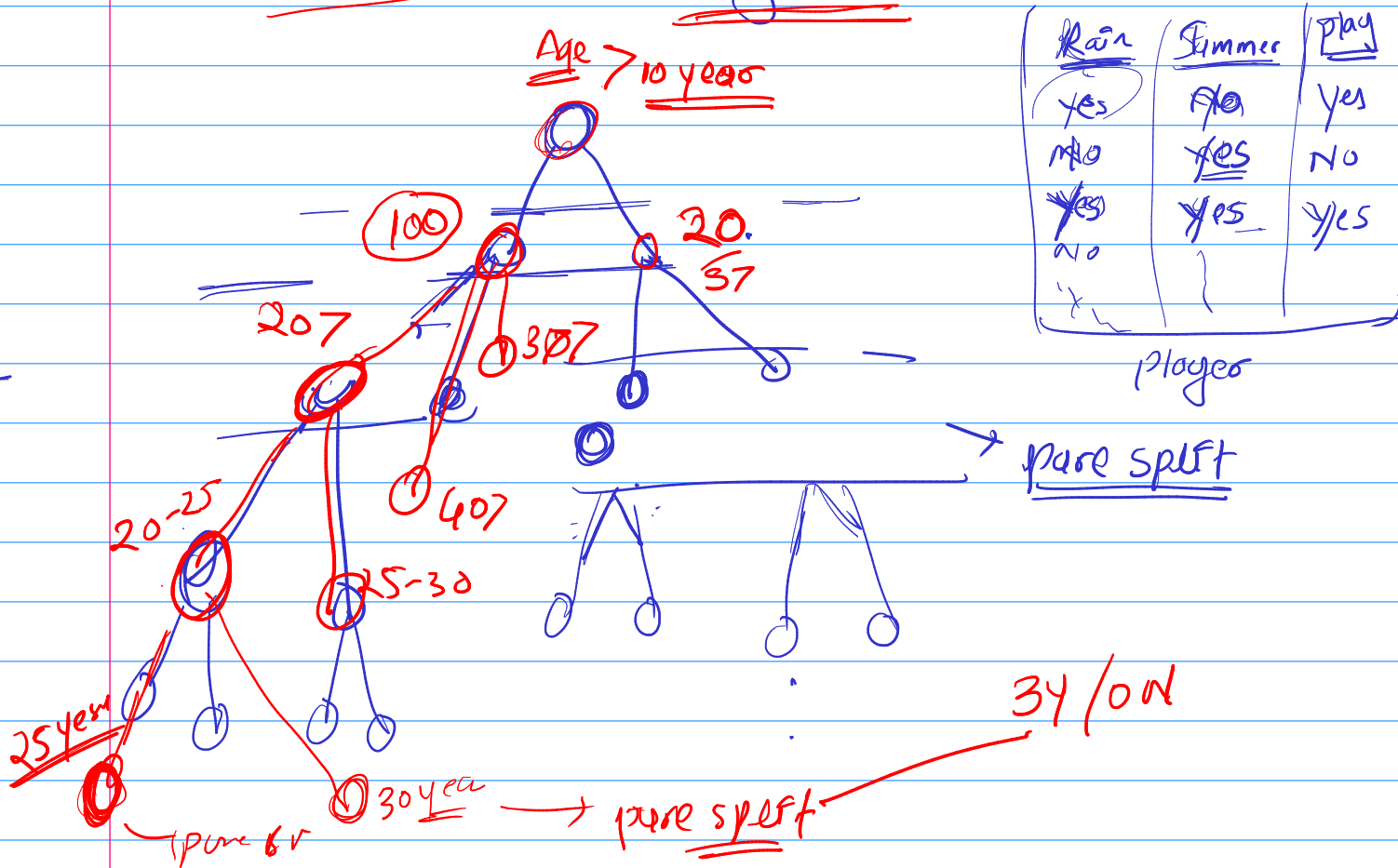
④ Linear kernel



⑧ Decision Tree (Reg - Class) If-else



- ① Entropy & Gini Impurity \rightarrow purity split
- ② Information gain \rightarrow feature Decision tree split
- ③ Pure split ? \rightarrow (Leaf Node)



Entropy

$$H(s) = -p_{+ve} \log p_{+ve} - p_{-ve} \log p_{-ve}$$

p_{+} → probability of yes

p_{-} → probability of No

If 3 yes / 0 No = 3

$$H(s) = \frac{-3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3} = 0 \rightarrow \text{Pure split}$$

0.7
0.2
0.3

If 3 yes / 3 No

$$H(s) = \frac{-3}{6} \log_2 \frac{3}{6} - \frac{3}{6} \log_2 \frac{3}{6} = 1 \rightarrow \text{Impure split}$$

Entropy range always 0 to 1

② Gini Impurity

$$GI = 1 - \sum_{i=1}^n (p_i)^2 = 1 - ((p_{+ve})^2 + (p_{-ve})^2)$$

for 3 yes / 3 No

$$= 1 - \left(\left(\frac{3}{6} \right)^2 + \left(\frac{3}{6} \right)^2 \right) = \underline{\underline{0.5}}$$

DC → If-else - Classification & Regression
to make decision

↓ technique → which feature select

① Gini Impurity (0-0.5)

③ Information Gain

② Entropy (0-1)

↓ log → max time
so, avoid

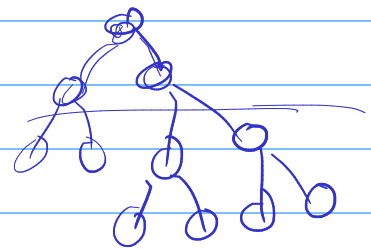
↓ max Gain
those feature selected
for splitting

Pure split (5y/0n)

Impure (5y/6n)

Post Pruning & Pre Pruning (MIMP Interview)

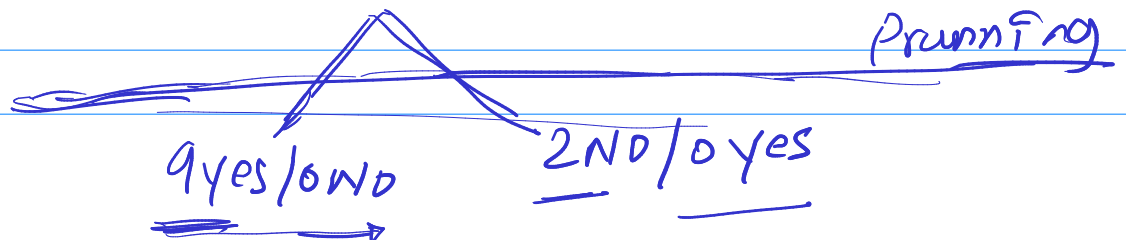
Pruning → Cutting the Branches



end splitting it overfitting

ex

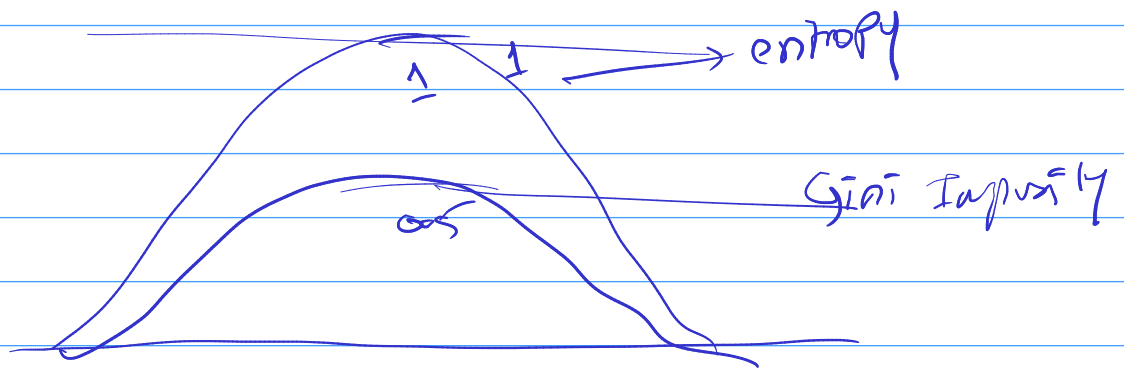
9yes/2no →



3 yes / 0 no

$$= 1 - \left(\frac{3}{3}\right)^2 + \left(\frac{0}{3}\right)^2 = 0$$

Gini Impurity = 0 to 0.5



which one use? → Gini Impurity we always use
in entropy we calculate \log_2 & it takes time

○ which feature

$f_1, f_2, f_3 \dots$ feature.

Rank Sum with
=

Information Gain

$$\text{Gain} = H(S) - \sum \frac{|S_v|}{|S|} H(S_v)$$

$H(S)$ → entropy of root node

$|S|$ → total sample $(94 + 56) = 14$

$|S_v|$ = sample of root node $(64 + 22) = 8$

gain = max

$\text{Gain}(f_1) = 0.049$
 $(f_2) = 0.051$ ✓ next feature

Hyperparameter

- max-depth

pre

(earlier)

post (after)

small data

for big data

Cross Validation

Train data

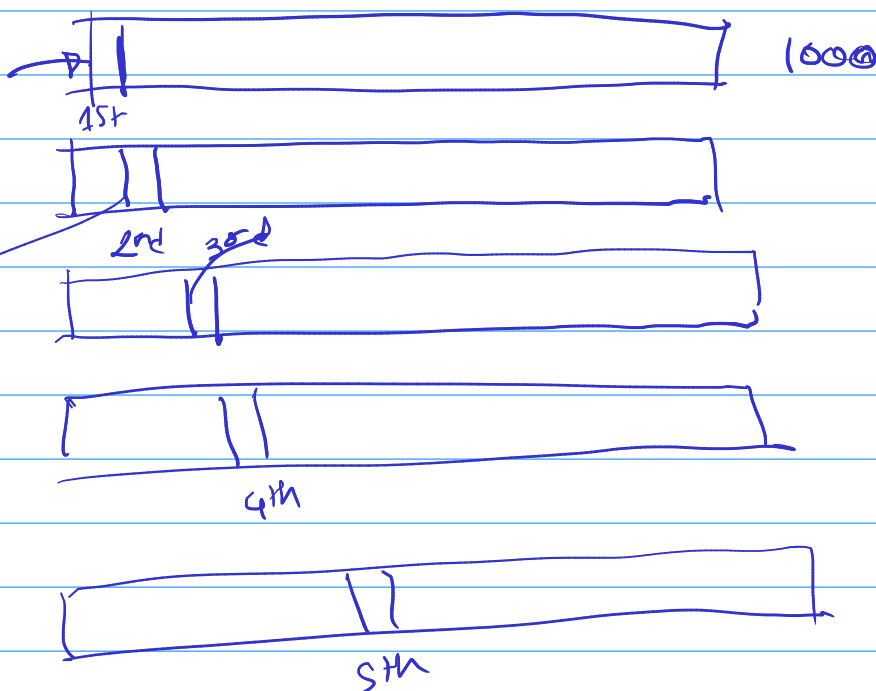
Test data

validation

① LOOCV → leave one out cross validation

CV=5

only 1 data point



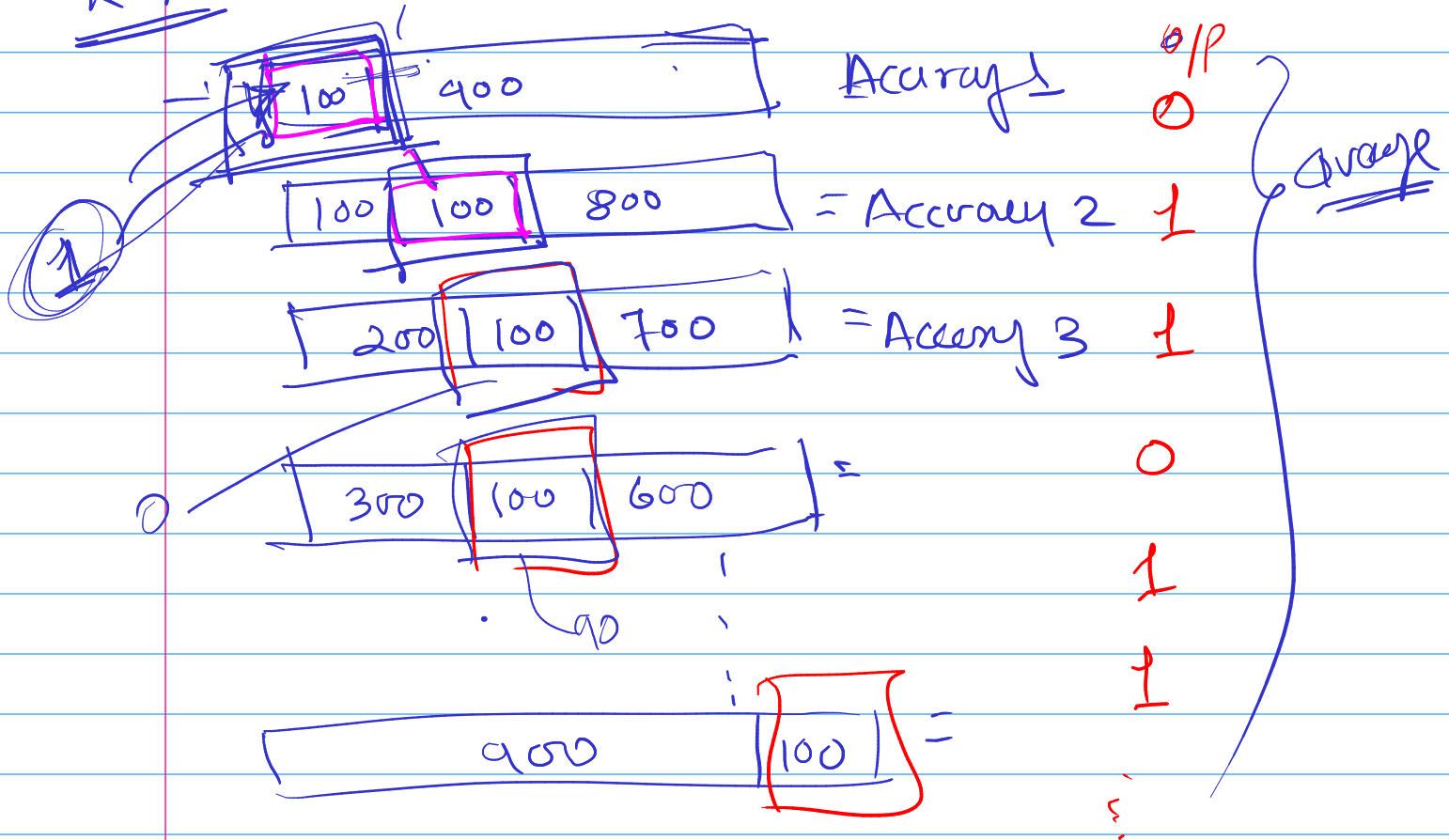
↓
(2) Hold out cross-validation



70% / 30%
↓
Hold cv=50

(3) K-fold cross validation

k=10 $1000 / 10 = 100 \rightarrow$ validation data



disadvantage → not work for imbalanced dataset.

balance
OP

imbalanced
OP

Stratified k-fold Cross Validation

$K=10$ equal & Random proportional
data.

100 → select →

0
1
0
1
0
2
1
0

1
1
1
1
1
1
1
0
0
0