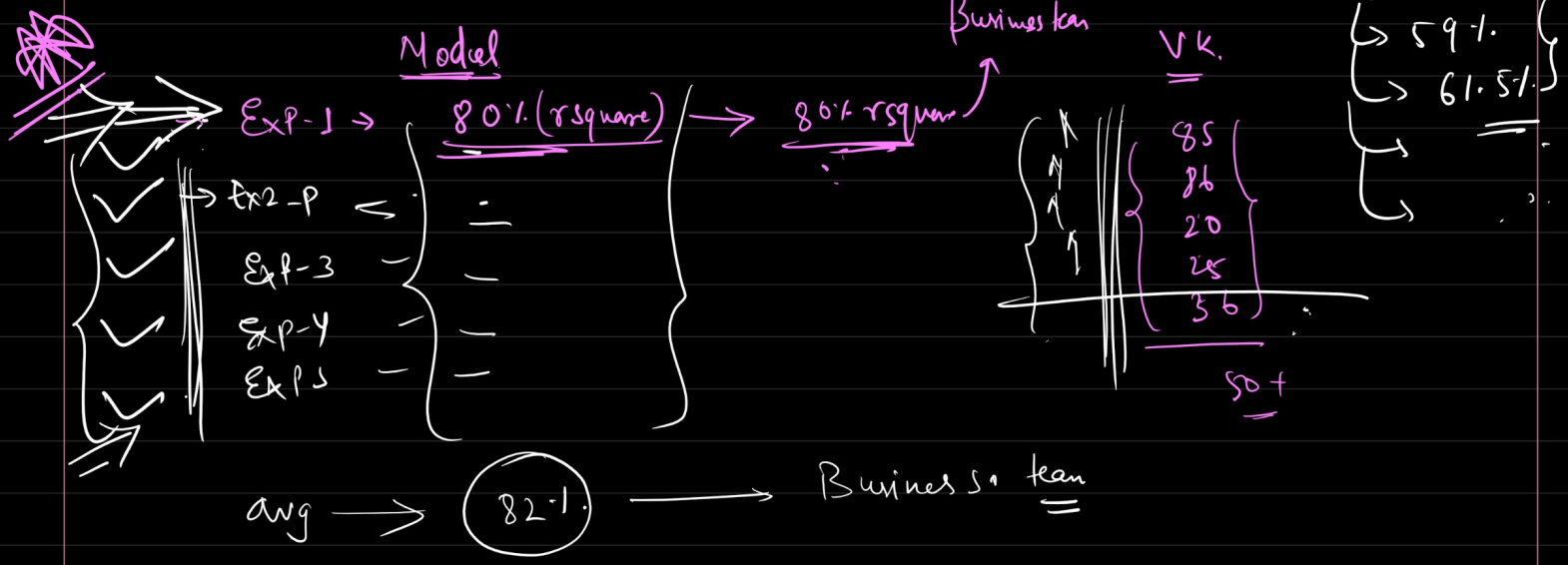
Agenda

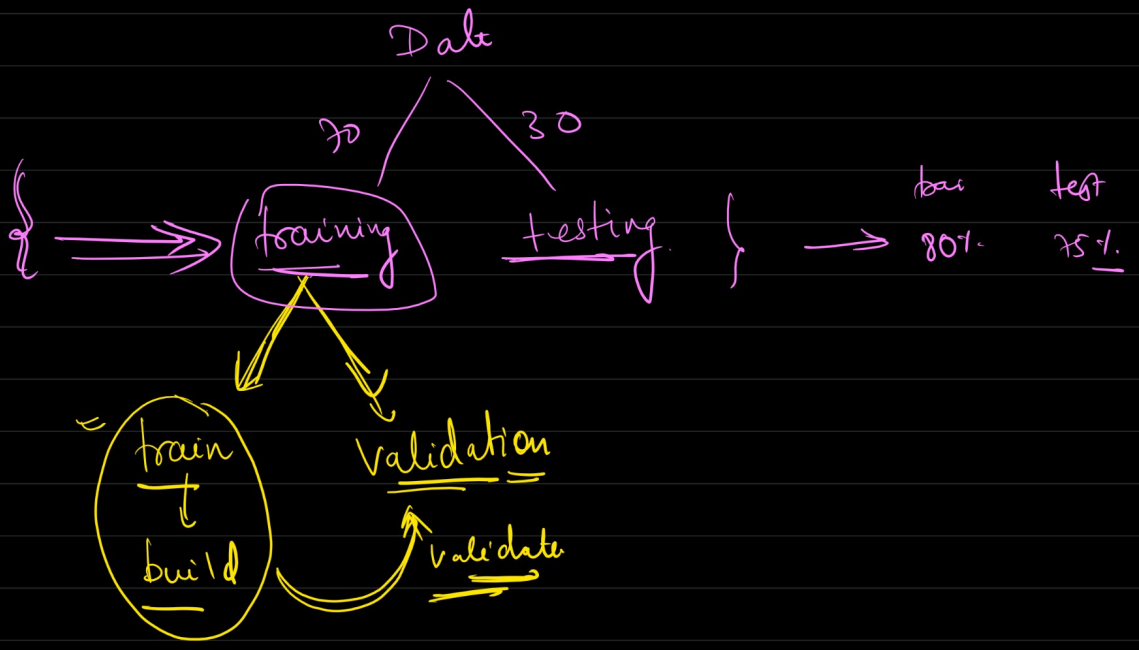
- → Cross validation & hyper parameter tuning.
- → Logistic Regression.
- → indepth Log reg.
- → Evaluation metrics
- → Practical Log.
- → Pract Cross validation / hyper parameter tuning
- → Multi class classification
- → " " implementation.

* Cross validation

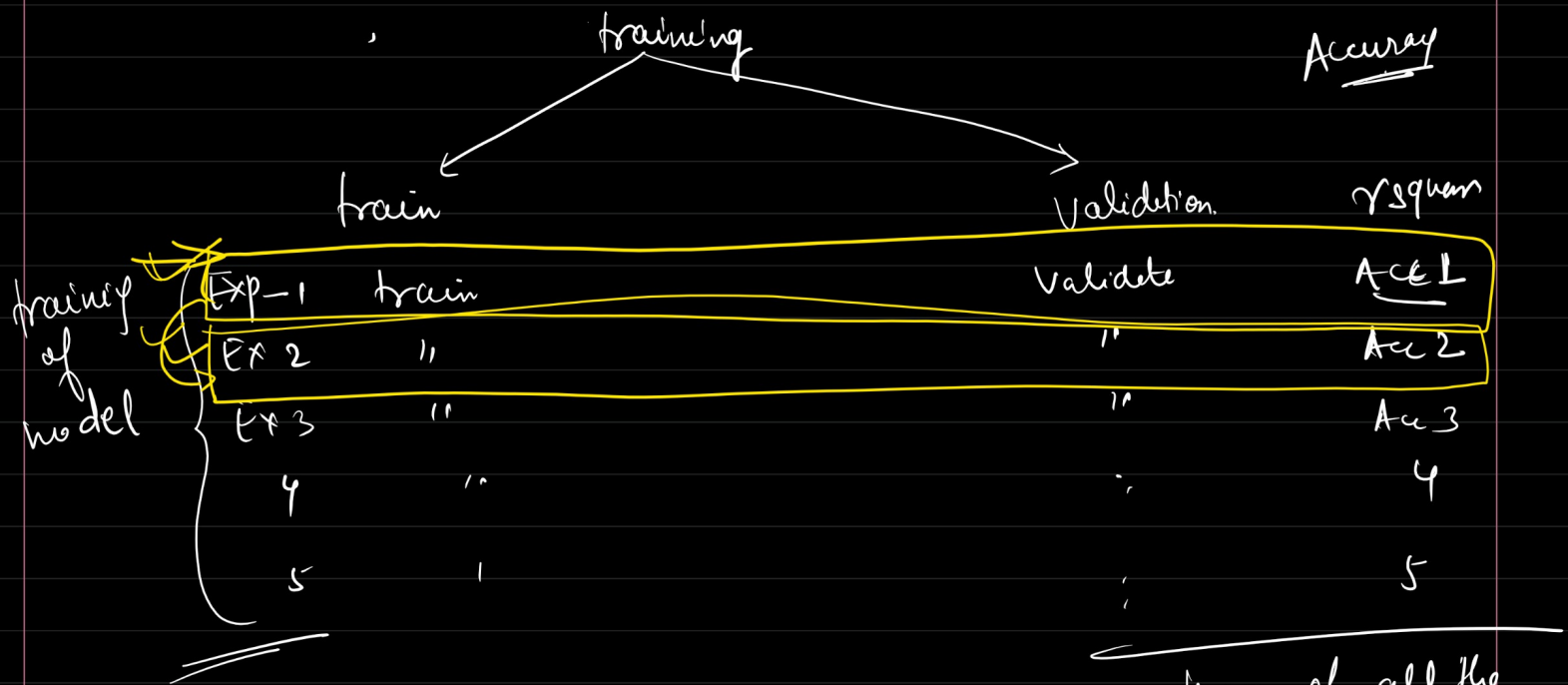
→ if you dont fix random state each time rsquare will change.



* Many experiments and average out the metrics increases the confidence in the model



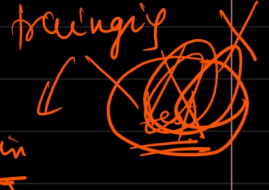
* Cross-validation → Experimenting with different arrangement of same data to build different models of same algorithms.



Arrangement of data

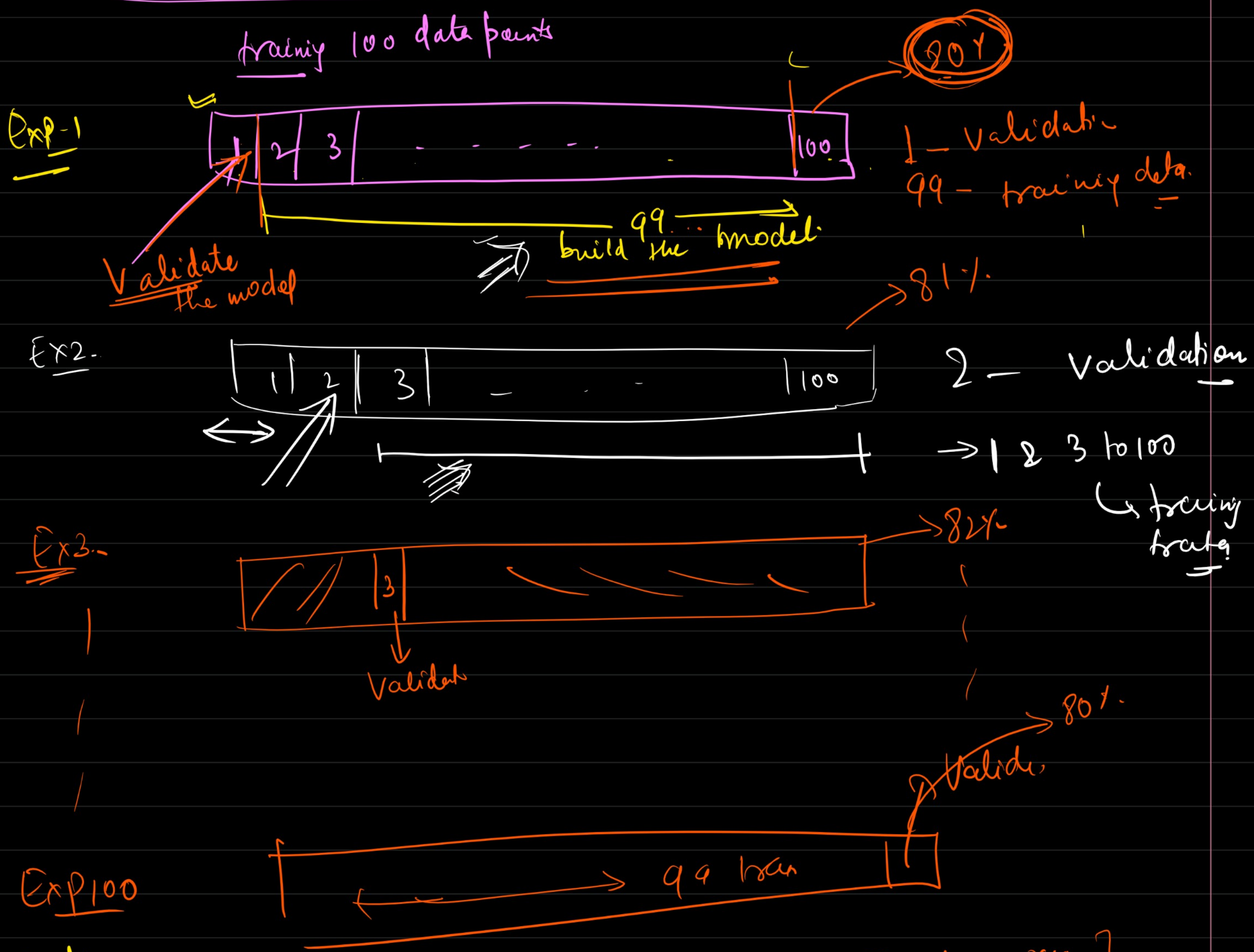
What ?????

→ Model will be robust / Generalised

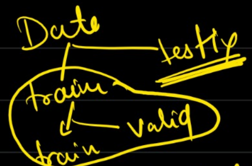


Types :-

(1) Leave One Out cross validation (LOOCV)



Avg Rsq – [All 100 rsquare]



* Validation data which is a part of training acts as test data in model training

* Disadvantage

10 millions

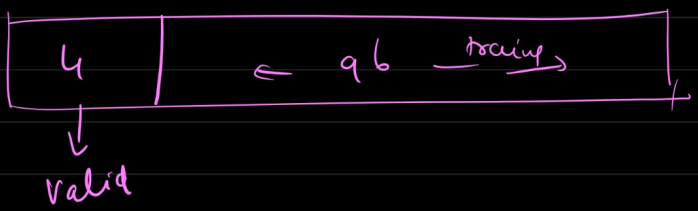
→ Time complexity is huge.

→ Model Overfitting

② leave P out Cross validation

$P \geq 1$: $P = 2, 3, 4, 5, 10, 15$ ...

Exp-1

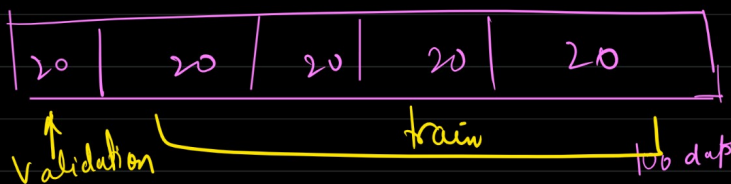


③ K- Fold Cross Validation

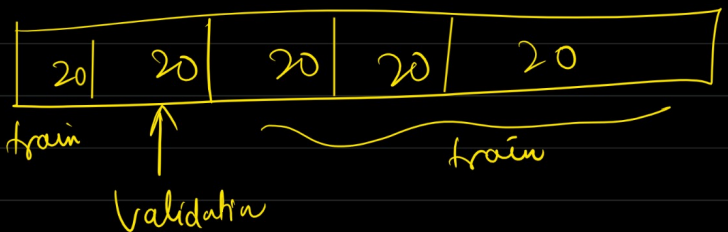
$n = 100 \Rightarrow$ K group $K = 5$ → No of groups that you want to divide the data.

$\frac{100}{5} = 20$ qp

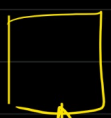
Exp-1



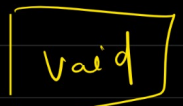
Exp-2



Exp-3



Exp-4



Exp-5

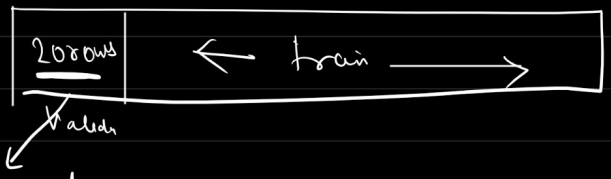
Avg of all Rsquare

④ Stratified k-Fold cross validation (imbalanced data)

K=5, n=100

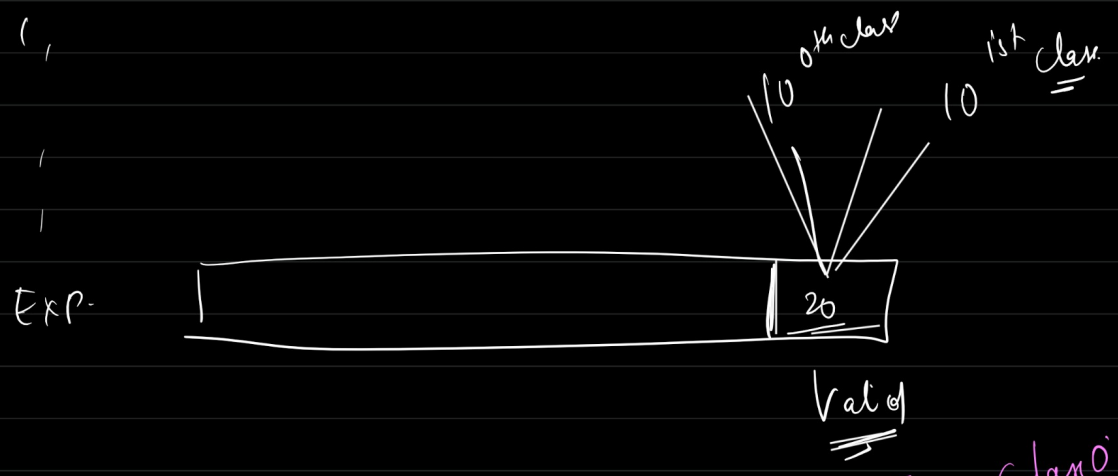
class-1	400 → 90%
class-0	100

Exp-1



- 10 row class 0
- 10 rows class 1

diabetic — 1
non — 0



→ class 0

class 0 class 1

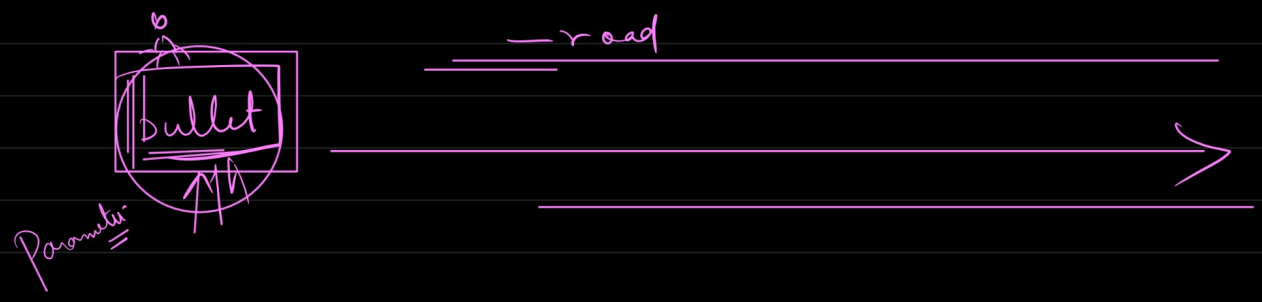
* Hyper parameter.

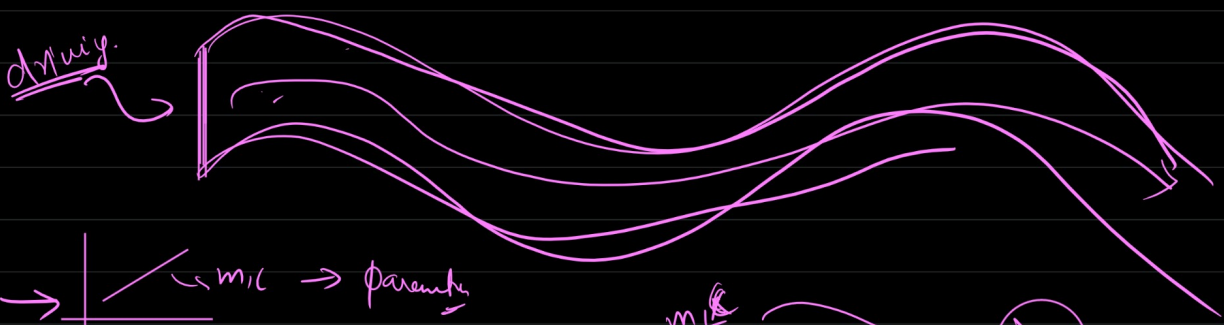
$$\text{Ridge} \rightarrow \left\{ \sum_{i=1}^{n} (y_{act} - y_{pred})^2 + \lambda(\text{slope})^2 \right\}$$

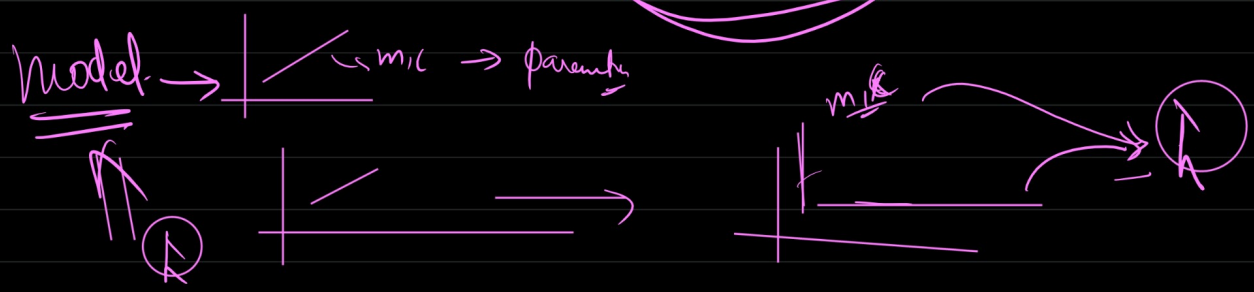
$$(y_{act} - mx_i - c)^2$$

3 Param

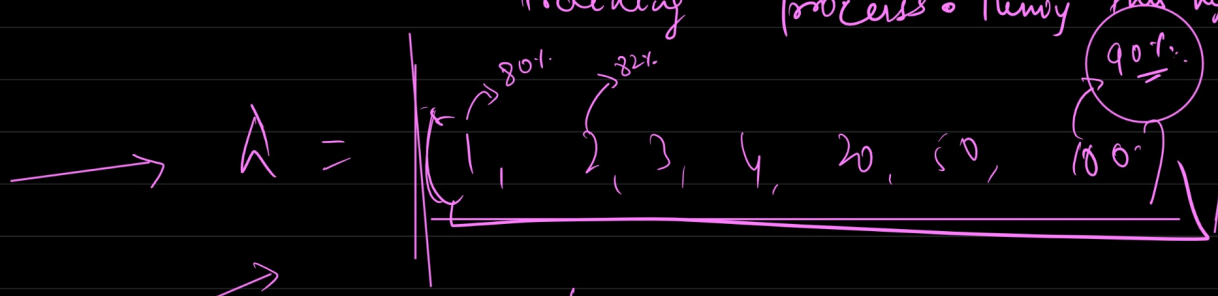
- $m_i c$ → parameter is learned from data → during model training.
- External parameter → Hyper parameter → dictates the behaviour of training algorithm.







* Hyperparmet → External configuration of a model that are not learned from the data but are set prior to training process. Tuning this hyperparmeter is Hyperparameter tuning.

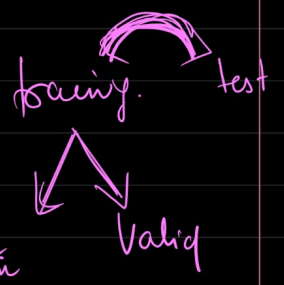


→ train → best performance → Hyperparameter tuned.

Hyper parameter tuning

CV

Hyperparameter

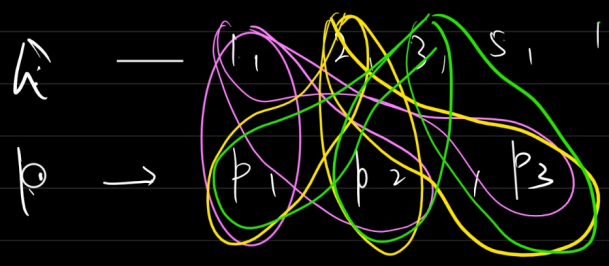


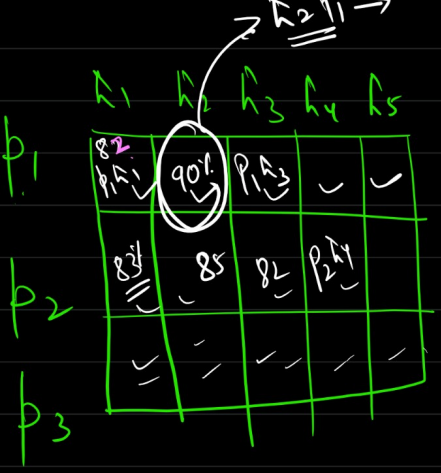
Hyperparameter tuning with Cross Validation

① Grid Search CV.

② Randomized Search CV

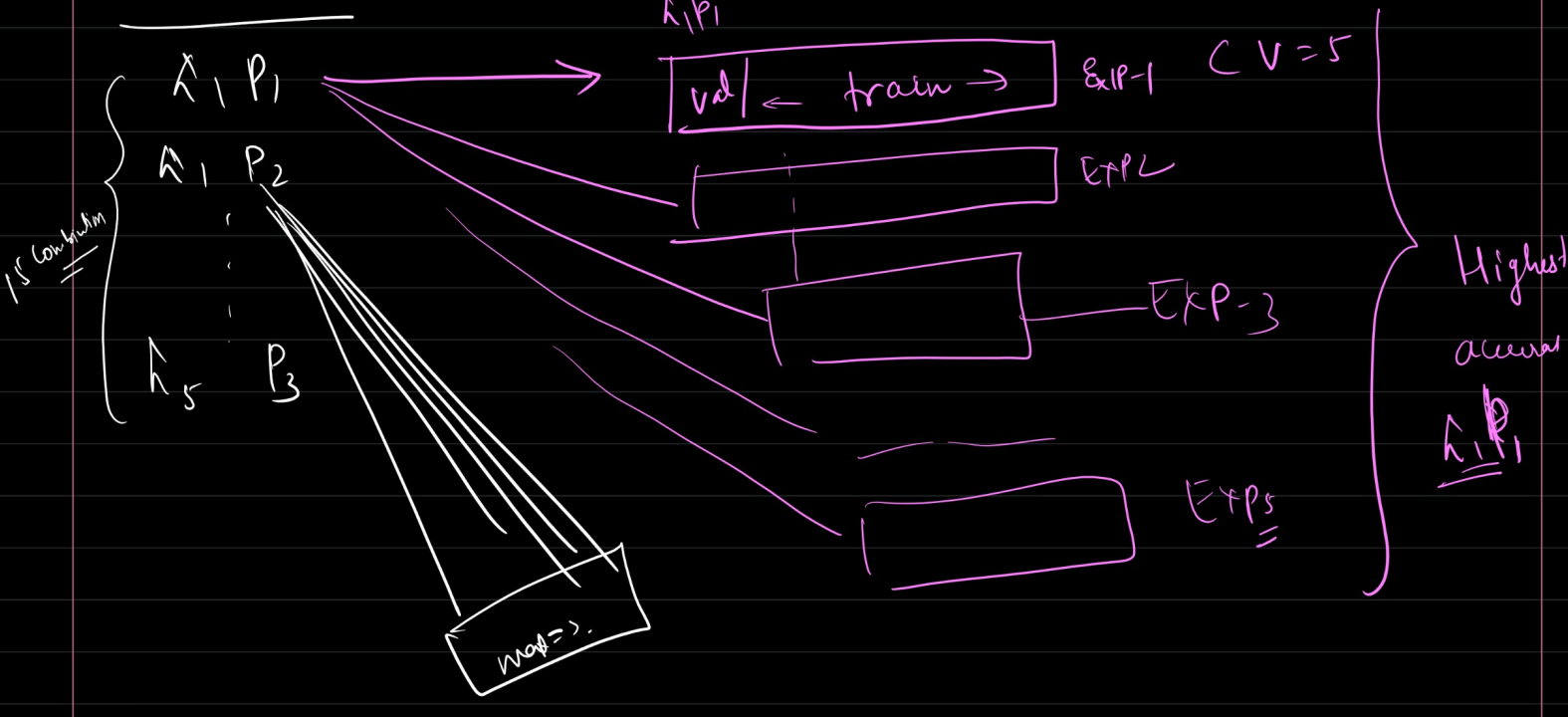
① Grid Search CV (Grid Search + Cross Validation)





→ All possible combination of hyper parameter is taken.

Cross validation



for each combination of K & P
K fold cross validation will happen

3 X 5 X 5

Disadvantage

↳ time complexity increases with huge data for model training.

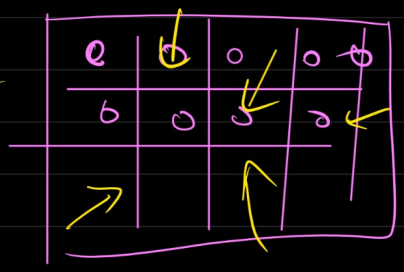
② Randomised Search CV

→ We will not see all possible combinations

→ Out of all the possible combination
Only select some random combination.

K1 P1
K2 P2
⋮
⇒ 15 combs

n_iter = 5



5 X CV = 5 ⇒ 25 models

Advantage — Time complexity decreases.

If data is small ⇒ Instead of train test split → You can go with

K_fold cross Validation

<10000