

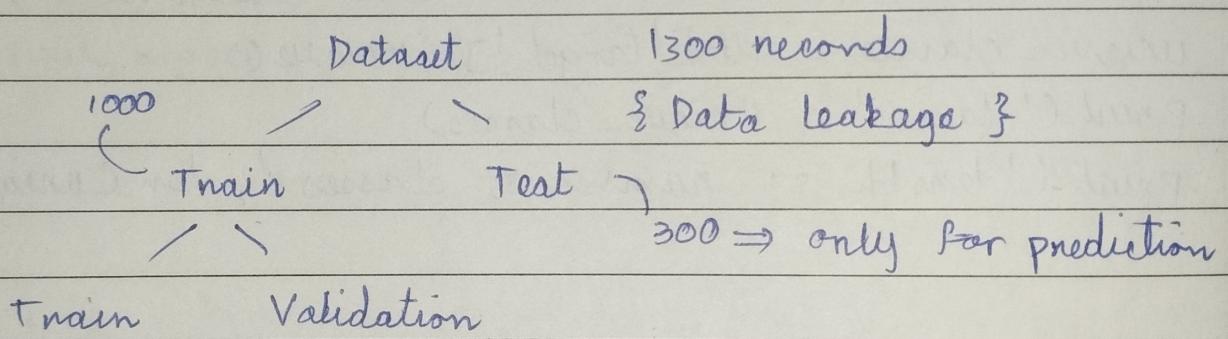
Cross Validation Techniques

- cross validation is a technique used to evaluate the performance of a machine learning model by testing how well generalizes to unseen data

Types of cross validation

1. LOOCV (Leave one out cross validation)
2. Hold out cross-validation
3. k-fold cross-validation
4. stratified k-fold cross validation
5. Time series cross-validation

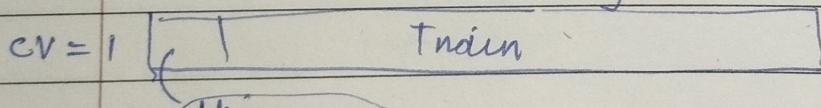
- ① LOOCV (Leave one out cross-validation)



Disadvantage

- ① Time-consuming task

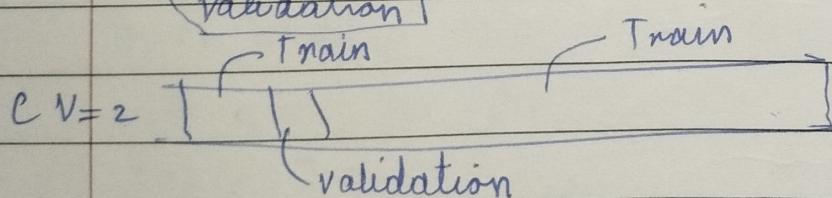
1000 records for eg



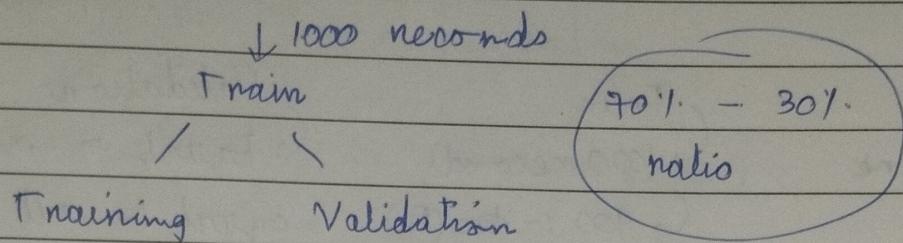
- ② Overfitting

(Low bias high variance)

this
will be
validation



② Hold out cross-validation



Training data

should be always greater
than validation

&

we use random_state = 42

↳ 87%

84%

} accuracy

Minimum = 82%.

Average = 87%.

Maximum = 97%.

89%.

↳ This doesn't

work in

Imbalanced dataset

Major disadvantage

is doesn't work with Imbalanced dataset

③ K-fold cross validation

for eg :- k = 10

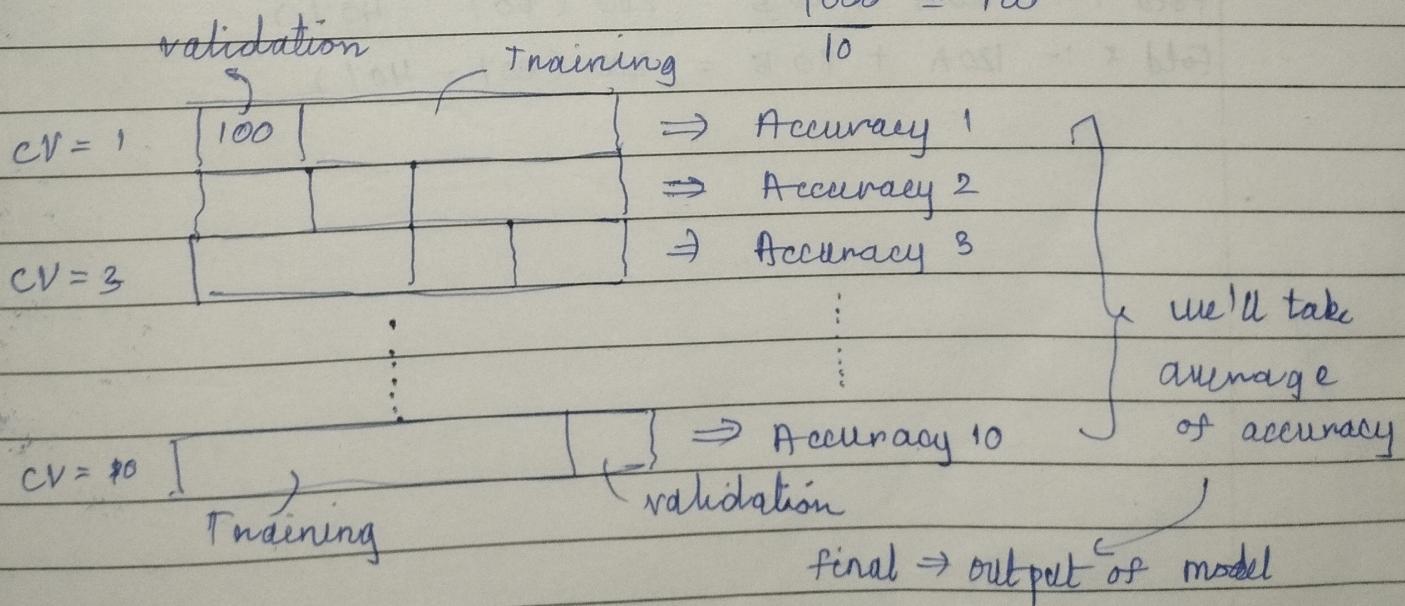
which is k-fold CV

$k=10 \Rightarrow$ this called as

T_{train} \Rightarrow 1000 records

[Experiments]

$$\frac{1000}{10} = 100$$



④ Stratified K-fold Cross Validation

used for hyper parameter tuning

This will work with imbalanced dataset up to some extent

Train	Validation
→ 1000 records	→ Equal or more approx equal to all the categories of output
→ 600 :- (A) (A)	
100 :- (B)	

Eg class A : [600 samples]

class B : [400 samples]

Total : 1000 samples

$$\text{Proportion A} = 600 / 1000 = 0.6 (60\%)$$

$$\text{Proportion B} = 400 / 1000 = 0.4 (40\%)$$

for k = 5 folds :-

$$\text{class A} \rightarrow 600 \div 5 = 120 \text{ per fold}$$

$$\text{class B} \rightarrow 400 \div 5 = 80 \text{ per fold}$$

$$\text{Fold 1} :- 120A + 80B = 200 (60\% - 40\%)$$

$$\text{Fold 2} :- 120A + 80B = 200 (60\% - 40\%)$$

$$\text{Fold 3} :- 120A + 80B = 200 (60\% - 40\%)$$

$$\text{Fold 4} :- 120A + 80B = 200 (60\% - 40\%)$$

$$\text{Fold 5} :- 120A + 80B = 200 (60\% - 40\%)$$

→ Not for Exam purpose just learning

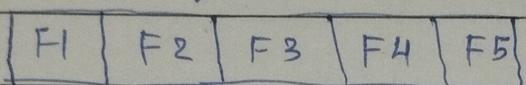
K-fold :- Works Purpose

Example :- 5-Fold cross validation

Training data (70% of original)



Divide into 5 folds



Iteration 1 : Train on F2, F3, F4, F5 ⇒ Validate on F1

Iteration 2 : Train on F1, F3, F4, F5 ⇒ Validate on F2

Iteration 3 : Train on F1, F2, F4, F5 ⇒ Validate on F3

Iteration 4 : Train on F1, F2, F3, F5 ⇒ Validate on F4

Iteration 5 : Train on F1, F2, F3, F4 ⇒ Validate on F5