Cross Validation & its types (Week-3 Assignment)

CV is one of the techniques used to test the effectiveness of a ML model., it is also a Re-Sampling procedure to evaluate a model if we have a limited data.

In CV, we need to keep aside a portion of data (to be used for testing / validation).

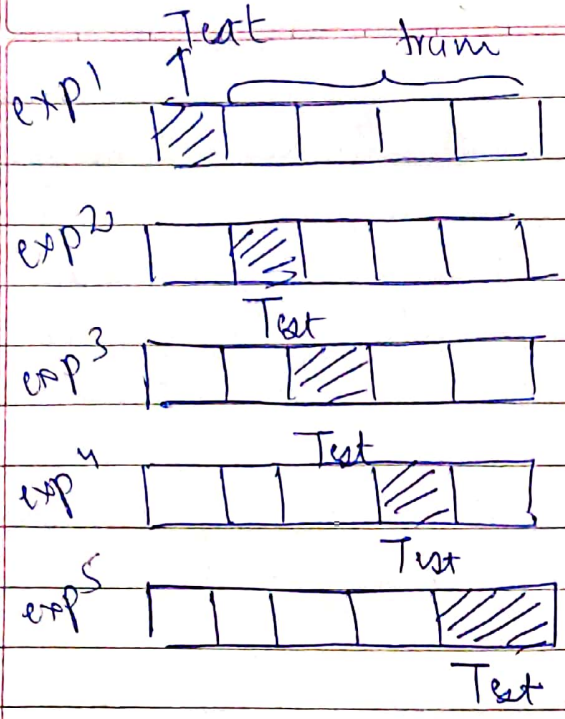
Steps in CV :-

1) Keep aside some portion of dataset
2) use rest dataset for training
3) Test / validate model using (kept aside) / ~~Remained~~ initially reserved portion of dataset.

1) Leave One Out CV (LOOCV)

Say we have 5 data ~~points~~ points

how our ML model very likely to make errors on test data



Closed vantage

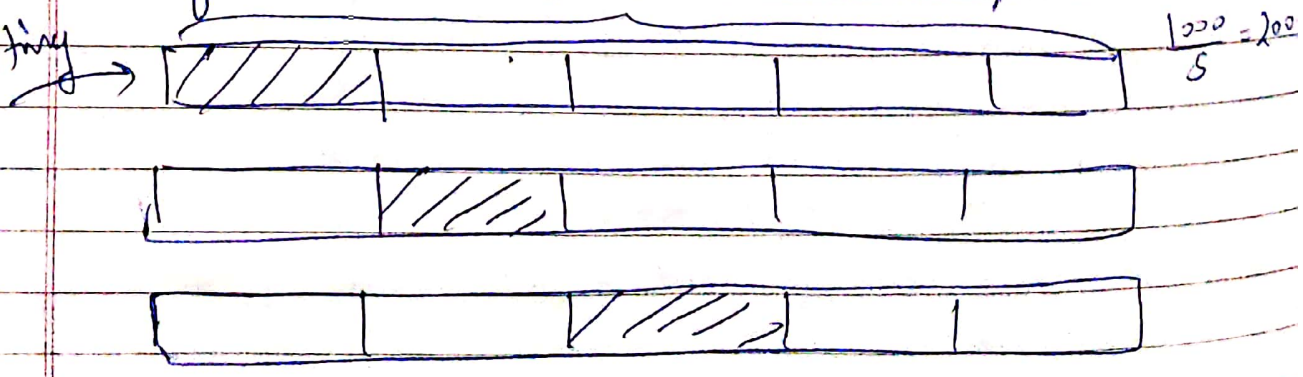
1) lots of iterations
2) ML is low biased
ie ML model incorporates fewer assumptions about target function.
3) high variance - if test data happens to be outliers

Resampling :- It is process of extracting new samples from a data set in order to get more accurate results.

2) K-Fold Cross Validation (Practically used)

→ Here you split ~~data into~~ input data into k-subsets of data (also known as folds)

→ Train ML model over k-1 folds & test the model ~~over~~ over kth fold. Repeat this process k times; with a diff. subset reserved for evaluation each time.



→ for each fold, our model gives some score. After k iterations, we have k scores, we take average of all these k scores. k scores, that will be performance metric for the model

Imp → If we have smaller dataset, we can go for k-fold CV. If we have large dataset, we can go for simple train-test split

→ CV gives less much better measure of model quality compared to train-test-split

3) Stratified k fold

If we are working on classification / multi class classification its better to use stratified k fold because because while making folds it maintains percentage of samples for each class in every fold.

ie in every fold we have equal no of data points from each class.

PAGE NO.
DATE / /

4) Time Series CV

We use it with time series data. Ex:- stock prediction

↓ we predict

Day 1	Day 2	Day 3	Day 4	Day 5
Day 2	Day 3	Day 4	Day 5	Day 6
Day 3	Day 4	Day 5	Day 6	Day 7