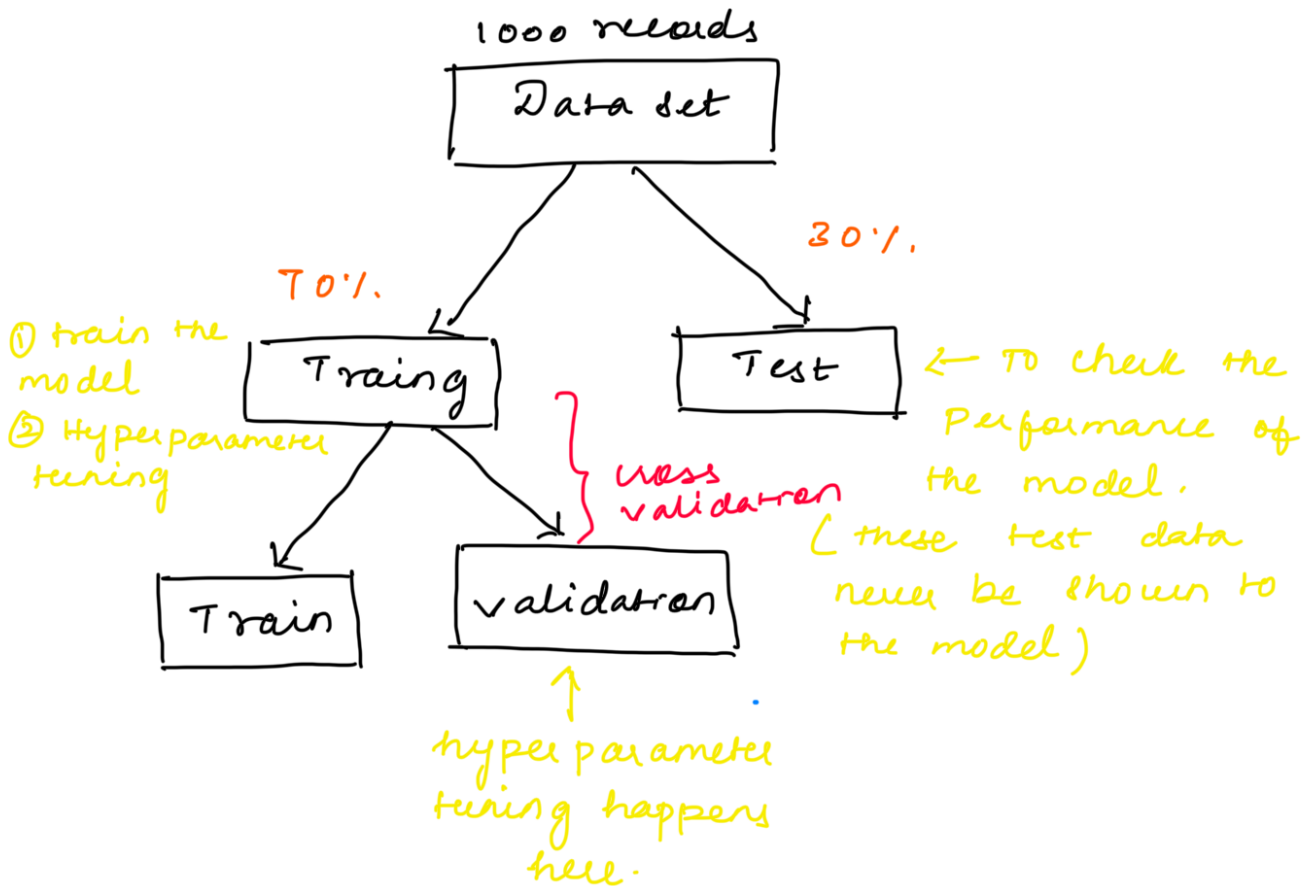
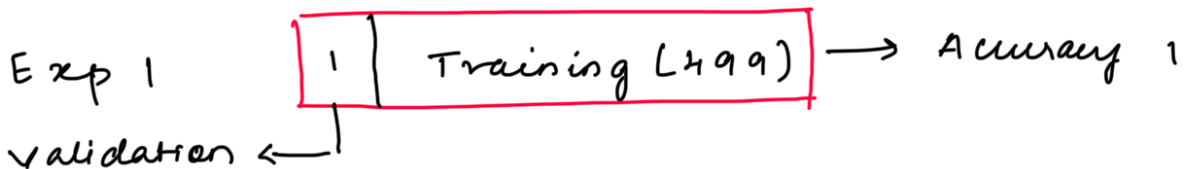
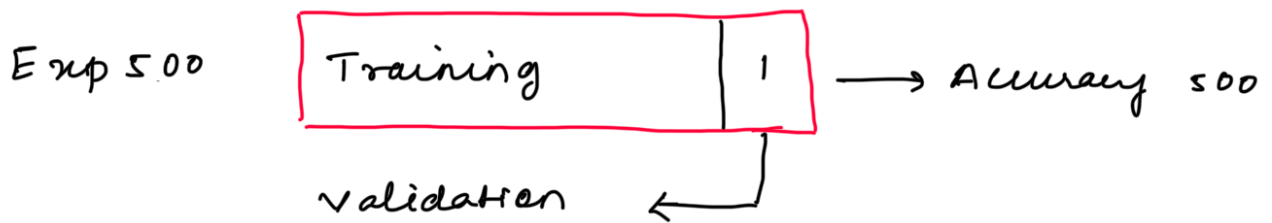


Types of cross validation



① Leave one out cross validation (LOOCV)





Draw back

Dataset increases $\uparrow\uparrow$ complexity in training the model also increase $\uparrow\uparrow$

→ In the above example, we have 500 records and we had performed 500 experiments - and had to calculate the average of accuracy.

- In recent times this technique is nowhere used.

→ This leads to overfitting.

| |
|--|
| Training accuracy $\uparrow\uparrow$ validation accuracy $\downarrow\downarrow$ |
|--|

↓
Testing with new data
results in low performance
(accuracy $\downarrow\downarrow$)

② Leave P out cross validation (LPOCV)

This is very similar to LOOCV, but here P is the variable it can take any value

$P=10$, $P=5$, $P=30$

The rest of the processing remains same as LOOCV.

③ K Fold cross Validation

Let say,

total dataset is 500

$$K = 5$$

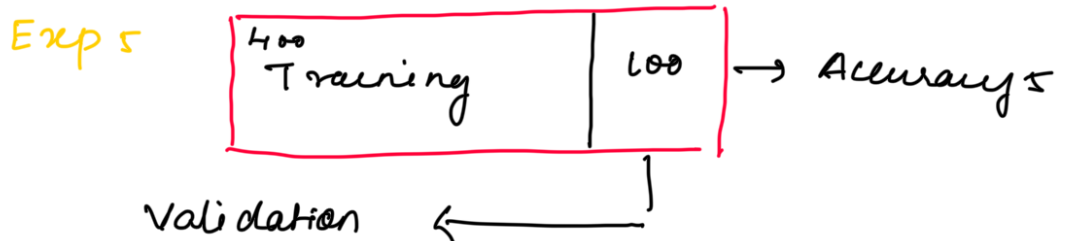
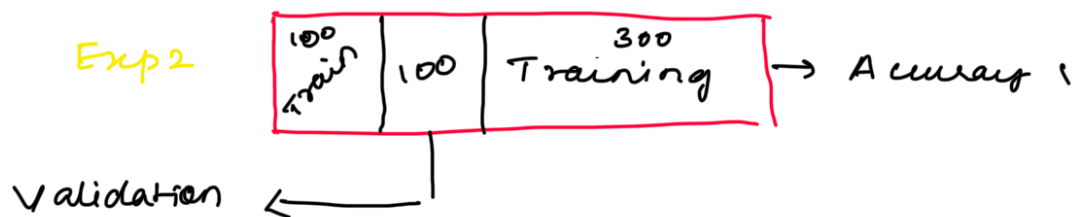
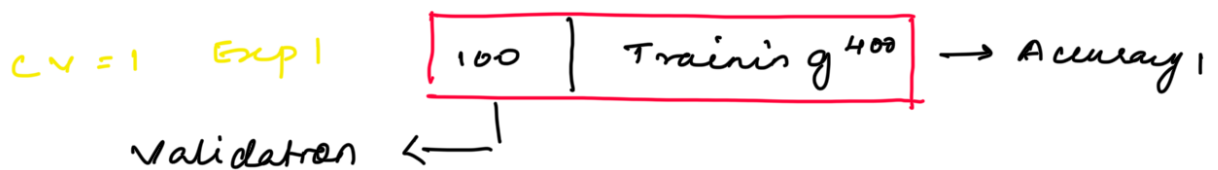
∴

$$n = 500$$

$$K = 5$$

therefore,

$$\text{Test Size} = \frac{500}{5} = 100$$



Finally, take the average of all the accuracy.

④ Another K Fold cross Validation

④ Stratified K Fold cross validation

Why stratified K Fold ?

Say for example, we are dealing with a classification problem, there might be possibilities that only one type of classification is comes in the validation, so we may not get the correct classification. hence the model may not trained properly.

So in Stratified K Fold CV. ensures that number of outputs from validation is evenly distributed.

Rest of the operations are similar to K Fold CV

⑤ Time Series cross Validation

Validation happens based on time / days

| Training | | | | Validation |
|----------|-------|-------|-------|------------|
| DAY 1 | DAY 2 | DAY 3 | DAY 4 | ... DAY n |

In Time series application, this validation will be used.

example:

→ Product review

