We have data from 1 hundred record this data I see its small

If we split data to 70 and 30

=over fitting

When we test rhe results is so bad

=>under fitting

Bad results at trai9jngning and testing

Data is the problem in both over and under fitting

100 70,30 we will learn from seventy and test from 30

1000 700,300

K form cross validation you well learn form all data and test using all data

K=5

Take 1/5 from the data to test and 4/5 for train test this for all 1/5 is the test data

Each time I will return the acurassy take the average of accuracy

In big data it will be deficlut to use this way so we will not use it else we have over fitting in the data

We will split the data to x and y go to x at index 1 then go to y to see the result

For train\_index, test\_index in kforl.split(x):

Print(test\_index)

X\_train, x\_test=x.ilo[train\_index,:],x.iloc[test\_index,:]

X\_train,y\_test=y[train\_index],y[test\_index]

Break;

#print(x)\_test)

Method iloc for take data from location x.iloc[ row,column]

Row can be train index

X test x.iloc[test\_index,0:0]

From row 3 take all data

We have 4 parameters x for train and test and ytrain , y test

Notes

We have two classes one positive and one negative we target the positive class

The cycle is the data that i see that is positive

Precision how many selected items are relevant tp/(tp+fp) positive from what I chose

Sensitivity: how many relevant items are selected tp/(fn+tp) positive form what I chose and the reality

Not all what choose that is true is true (from last section)

F1 measure

🡺both precision and recall

Do the task is the task