M3L4

1. Train & Test (from sklearn)

test size = 0.33 KNN classifier

↳ K → # of nearest neighbours

how change in K ⇒ how accuracy changes.

some samples gone to test, some gone to train.

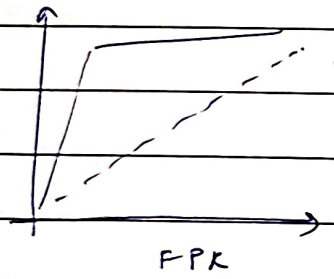
Plot can be diff

Q. How do we know our model is robust?

also dataset has values that may be in 1000's or even fractions.

2. AUC - ROC curve.

Probability → degree of seperability.



← Diagonal line always tells

class^n above red line : good

→ It tells how much the model is capable of distinguishing b/w classes.

Higher area ⇒ Higher the classification

the better the model is by classifying

random_state: 1079

max accuracy is acheived.

we divide training data into n folds.

ex: 5 folds

in split 1 : validate on split 1
train it on remaining splitting.

each

we train 5 models

⇒ we have diff validation sets.

k-fold split:

train KNN classifier.

x-axis : # of neighbours

y-axis : accuracy. (average)

for every value of k (nearest model)

10 models ⇒ 10 accuracys