DATA MINING HW-4

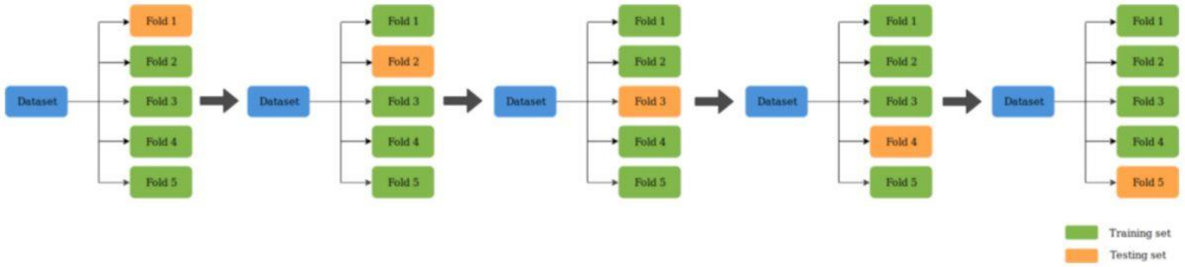
ALPER YASAR – 151044072

Implemented naïve bayes classification model with using k-cross validation made from me and get f1 score.
Bayes’ Theorem:

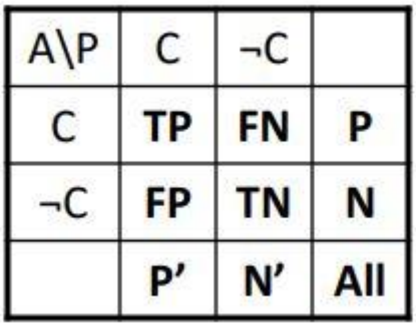
$$P(H|\mathbf{X}) = \frac{P(\mathbf{X}|H)P(H)}{P(\mathbf{X})} = P(\mathbf{X}|H) \times P(H) / P(\mathbf{X})$$

- Let **X** be a data sample (“*evidence*”): class label is unknown
- Let H be a *hypothesis* that X belongs to class C
- Classification is to determine P(H|**X**), (i.e., *posteriori probability):* the probability that the hypothesis holds given the observed data sample **X**
- P(H) (*prior probability*): the initial probability
- P(**X**): probability that sample data is observed
- P(**X**|H) likelihood): the probability of observing the sample **X**, given that the hypothesis holds

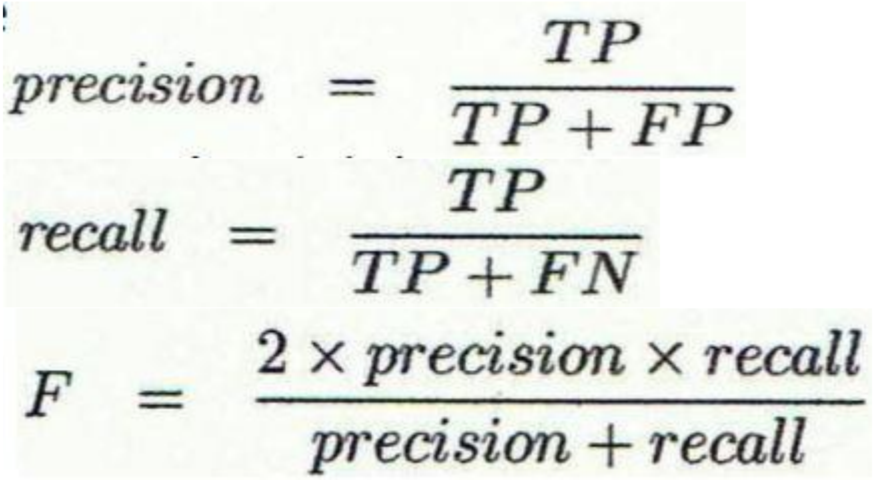
Firstly I read datas from file and separete them to 2 array. First array holding first 4 value and second array hold last values.
I create a k cross validation method for calculate scores mean.
K-Fold CV is where a given data set is split into a ***K*** number of sections/folds where each fold is used as a testing set at some point.



Calculate each f1 scores for each fold and taking mean for f1 average value. It’s average of bayes. When run my own naive bayes class i take f1 score:
k=5 => 59.22
k=3 => 43.59
k=6 => 60.16



A\P	C	¬C	
C	**TP**	**FN**	**P**
¬C	**FP**	**TN**	**N**
	P’	**N’**	**All**



$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F = \frac{2 \times precision \times recall}{precision + recall}$$

I used PCA technique using a data mining tools with k validation and get f1 score is :
k=5 => 67.78.

k=3 => 67.77
k=7 => 66.19
I used LDA technique using a data mining tools with k validation and get f1 score is :
k=5 => 56.39.

k=3 => 63.37
k=7 => 58.57

Which technique has given better results in terms of f1 score? (PCA or LDA)?
PCA give better result than LDA. All time is give approximately same results 67. When train data change but LDA give approximately 60 score.
I can not add featuren selection to my project. So i can not give answer for other questions.