

Assignment-1

Name: Gohil Happy kanaiyalal

Roll No: 21IM30006

NOTE

This report contains 4 sections which are Experiment-1, Experiment-2, Experiment-3 and Observations. In experiment sections contains tables of Mean and variance of performance value(Precision, recall, accuracy, F1) collected over a K-fold

Experiment-1

Procedure:

- Trained dataset-1 with 2 to 10 splits and validated
- Then made a 80-20 train test split of data counted mis-classified instance and plotted it using matplotlib

1.1 K-fold cross validation

K	Precision	recall	accuracy	F1
2	0.996992	0.995992	0.993	0.996485
3	0.997992	0.996994	0.994998	0.997491
4	0.999	0.997	0.996	0.997996
5	0.996995	0.997995	0.995	0.997491
6	0.997998	0.997998	0.996008	0.997995
7	0.995983	0.993978	0.990003	0.994959
8	0.997976	0.993976	0.992	0.995968
9	0.998999	0.995004	0.994003	0.996983
10	0.999	0.997	0.996	0.997985

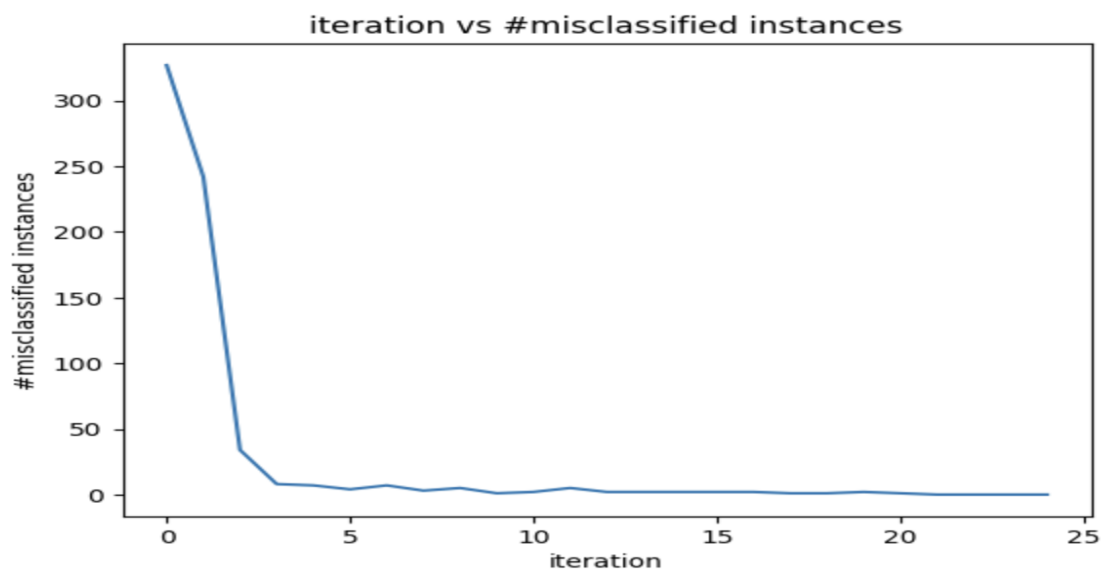
Table - 1: Mean values of performance measures for different value of K

K	Precision	recall	accuracy	F1
2	0.000004	0.00005	0.000081	0.000021
3	0.000018	0.000008	0.000038	0.00001
4	0.000027	0.000008	0.000011	0.000003
5	0.000014	0.000016	0.000026	0.000007
6	0.000005	0.000021	0.00002	0.000005
7	0.000026	0.000026	0.000056	0.000014
8	0.000007	0.000015	0.000032	0.000008
9	0.000036	0.000037	0.00009	0.000023
10	0.000024	0.000016	0.000044	0.000011

Table - 2: variance of performance measures for different value of K

Fig:1 experiment 1

1.2 80-20 train-test split performance



Precision	0.99497486
Accuracy	0.99
Recall	0.99497487
F1	0.994974

Table: Performance value for experiment 1

Experiment-2

Procedure:

- Trained dataset using k-fold cross validation for k in range from 2 to 10
- Then made a 80-20 train test split of data counted mis-classified instance and plotted it using matplotlib

Below is the tables and results

2.1 K-fold cross validation

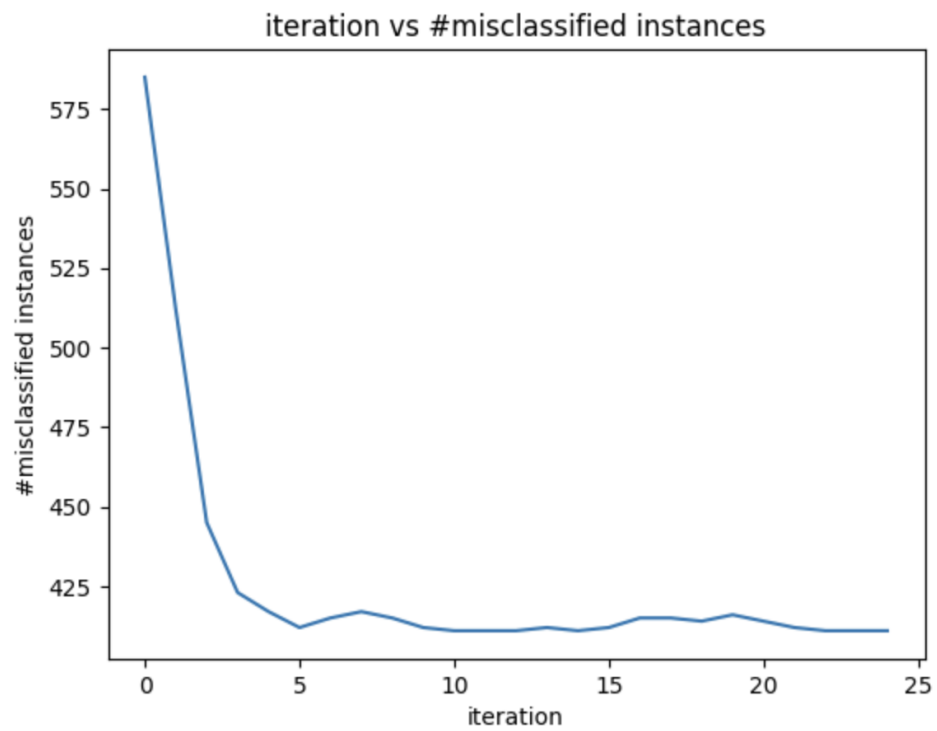
K	Precision	recall	accuracy	F1
2	0.993951	0.993952	0.988	0.993947
3	0.997992	0.995993	0.993997	0.996988
4	0.999	0.997	0.996	0.997994
5	0.995995	0.997995	0.994	0.996987
6	0.999002	0.994998	0.994	0.996988
7	0.996982	0.995976	0.993	0.996466
8	0.997984	0.995984	0.994	0.996976
9	0.996997	0.996997	0.993994	0.996974
10	0.99798	0.99498	0.993	0.996467

Table - 3: Mean values of performance measures for different value of K

K	Precision	recall	accuracy	F1
2	0.000004	0.000037	0.000064	0.000016
3	0.000008	0.000002	0.000006	0.000002
4	0.000003	0.000011	0.000008	0.000002
5	0.000014	0.000006	0.000014	0.000004
6	0.000005	0.000017	0.000012	0.000003
7	0.000012	0.000055	0.000084	0.000022
8	0.000012	0.000032	0.00006	0.000015
9	0.000036	0.000036	0.000054	0.000014
10	0.000016	0.000046	0.000081	0.000021

Table - 4: variance of performance measures for different value of K

2.2 80-20 train-test split performance



Precision	0.994949
Accuracy	0.985
Recall	0.9899
F1	0.9924

Table: Performance value for experiment 1

Experiment-3

Procedure:

- Trained dataset-1 with 2 to 10 splits and validated
- Then made a 80-20 train test split of data counted mis-classified instance and plotted it using matplotlib

3.1 K-fold cross validation

K	Precision	recall	accuracy	F1
2	0.994933	0.990933	0.986	0.992925
3	0.996988	0.99599	0.992996	0.996482
4	0.999	0.996	0.995	0.997493
5	0.997	0.996	0.993	0.996484
6	0.99899	0.996988	0.996002	0.997986
7	0.998002	0.996996	0.994998	0.99749
8	0.997	0.997	0.994	0.996984
9	0.996988	0.994995	0.992001	0.995969
10	0.997	0.997	0.994	0.996985

Table - 5: Mean values of performance measures for different value of K

K	Precision	recall	accuracy	F1
2	0.000009	0.00005	0.0001	2.57E-05
3	0.000018	0.000002	0.000014	3.57E-06
4	0.000003	0.000008	0.000003	7.59E-07
5	0.000016	0.000024	0.000016	4.04E-06
6	0.000005	0.000021	0.000044	1.12E-05
7	0.00001	0.000012	0.00001	2.52E-06
8	0.000015	0.000031	0.000028	7.10E-06
9	0.000036	0.000038	0.000062	1.58E-05
10	0.000021	0.000021	0.000024	6.06E-06

Table - 6: variance of performance measures for different value of K

2.2 80-20 train-test split performance

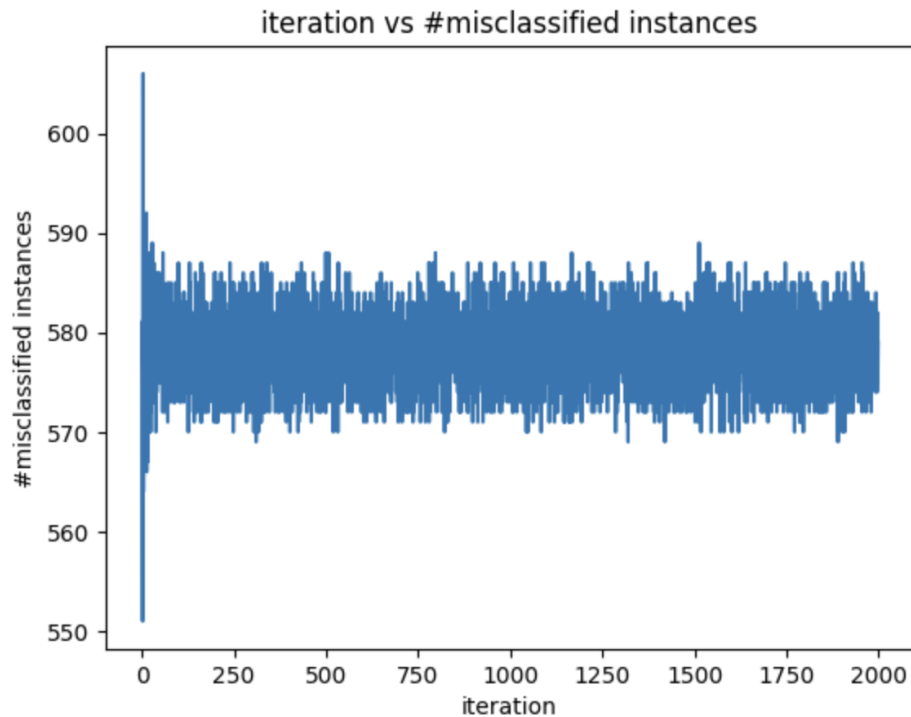


Figure 3

Precision	0.66
Accuracy	0.52
Recall	0.7027
F1	0.68421

Table: Performance value for experiment 1

Observations:

- 1) As per the plot it is observed that **data set 3 is not linearly separable data** as its plot is fluctuating over the iterations
- 2) In comparing data set 1 and 2. Data set 1 is more quickly converging to the 0
- 3) From performance value on test data set it is observed that
 - a) Data 3 has lowest values in all performance values (precision, accuracy, recall, F1)
 - b) In comparing data 2 and 1 there is minute difference that data 2 is less accurate predicted then data 1.