

ASSIGNMENT 3

NAME: KANIKA MIDHA
ROLL NO: 201301037

RESULTS OF 1 NEAREST NEIGHBOUR CLASSIFIER:

Accuracy: $9540/10000 = 95.4 \%$

Time to run: 24 hrs

KNN	Neural Networks
Easy to code	Complex to code
Lazy classifier-major time is classification time	Eager classifier-Training time is high
High Dimension input features causes knn to take a lot of time	Much Faster than knn even with large amount of data
Knn can be retrained.	Neural networks cannot be retrained. If you add data later, this is almost impossible to add to an existing network

NEURAL NETWORKS

1) FIRST FOLD CONFUSION MATRIX FOR FIVE FOLD CROSS VALIDATION

1077	2	8	6	0	21	58	0	0	3
0	1361	5	3	1	3	6	0	0	1
17	134	808	43	26	2	100	27	9	5
9	59	32	1076	2	21	37	7	0	16
0	40	6	3	942	0	63	1	0	108
21	73	10	269	21	520	113	10	1	58
5	21	5	0	4	14	1112	0	0	1
9	85	25	7	14	0	3	989	0	78
3	277	16	200	26	35	71	5	397	135
19	64	16	29	115	3	6	37	0	929

Accuracy: 76

Sensitivity: 76.76

Precision: 76.76

Specificity: 97.4

Error Rate: 23.23

2) CONFUSION MATRIX FOR SECOND FOLD IN FIVE FOLD CROSS VALIDATION

1143	0	5	15	5	0	16	2	20	0
0	1313	3	7	1	1	3	1	21	1
18	44	898	70	52	0	35	9	49	1
4	24	18	1115	4	2	11	10	36	4
5	15	6	5	1090	0	15	1	26	21
38	28	7	349	39	228	48	5	298	8
20	8	13	4	11	3	1132	0	17	0
12	61	17	26	45	0	0	1046	15	57
13	38	9	130	13	1	19	2	895	7
23	16	11	45	210	1	1	46	77	763

Accuracy: 80.19

Sensitivity: 80.19

Precision: 80.19

Specificity: 97.7

Error Rate: 19.8

3) CONFUSION MATRIX FOR THIRD FOLD IN FIVE FOLD CROSS VALIDATION

1142	0	6	3	2	0	17	6	30	0
0	1305	4	0	1	3	2	3	32	1
19	47	950	9	37	2	23	24	65	0
7	31	37	935	1	52	11	32	116	6
5	15	12	0	1000	3	17	11	38	83
14	23	7	67	16	656	28	17	213	7
11	9	17	0	5	13	1126	2	25	0
10	29	10	1	15	3	0	1174	9	28
11	27	14	18	7	18	21	6	998	7
16	11	4	14	45	4	1	164	73	861

Accuracy: 84.55

Sensitivity: 84.55

Precision: 84.55

Speceficity:98.28
Error Rate: 15.44

4) CONFUSION MATRIX FOR FOURTH FOLD IN FIVE FOLD CROSS VALIDATION

1132	1	14	4	1	1	26	6	7	14
0	1327	5	1	1	1	1	2	8	5
9	54	1016	6	18	2	19	13	19	20
5	50	51	984	0	26	11	29	25	47
0	22	20	0	634	0	20	1	3	484
15	48	29	92	13	668	31	12	63	77
9	26	30	1	2	14	1117	1	5	3
8	49	12	0	4	0	0	1065	2	139
11	104	28	43	3	25	21	3	810	79
11	20	7	13	3	2	2	39	5	1091

Accuracy: 82.033
Sensitivity: 82.03
Precision:82.03
Speceficity:98
Error Rate: 17.96

5) CONFUSION MATRIX FOR FIFTH FOLD IN FIVE FOLD CROSS VALIDATION

1140	0	9	2	1	1	45	2	5	1
0	1317	9	1	1	0	5	2	12	4
17	31	1024	4	11	0	49	10	20	10
8	35	74	955	0	34	28	20	38	36
0	17	15	0	801	1	61	1	5	283
39	28	26	69	8	672	87	6	70	43
9	9	17	0	1	8	1161	0	3	0
14	45	17	0	9	1	4	1054	2	133
19	73	38	28	3	26	51	3	841	45
21	19	12	14	10	2	3	34	6	1072

Accuracy: 83.64 %
Sensitivity: 83
Precision:83
Specificity:98.1
Error Rate: 16.35

RESULTS (Mean of all five folds)

- **Accuracy across five folds = 82 %**
- **Average Error Rate : 18 %**
- **Standard Deviation : 2.7**
- **Sensitivity: 81**
- **Specificity: 98**
- **Precision: 81**

For Test Data

Confusion Matrix:

949	0	4	13	1	0	32	1	1	0
0	1102	5	7	0	1	7	2	2	1
19	35	773	48	25	1	59	15	14	2
4	20	16	937	2	4	16	18	7	8
2	14	6	3	878	0	44	5	2	26
41	33	5	247	18	405	76	7	22	9
6	7	9	2	2	8	980	0	0	0
12	38	10	9	20	2	2	947	2	28
21	82	17	146	9	14	59	6	573	17
25	16	4	33	94	3	6	92	2	703

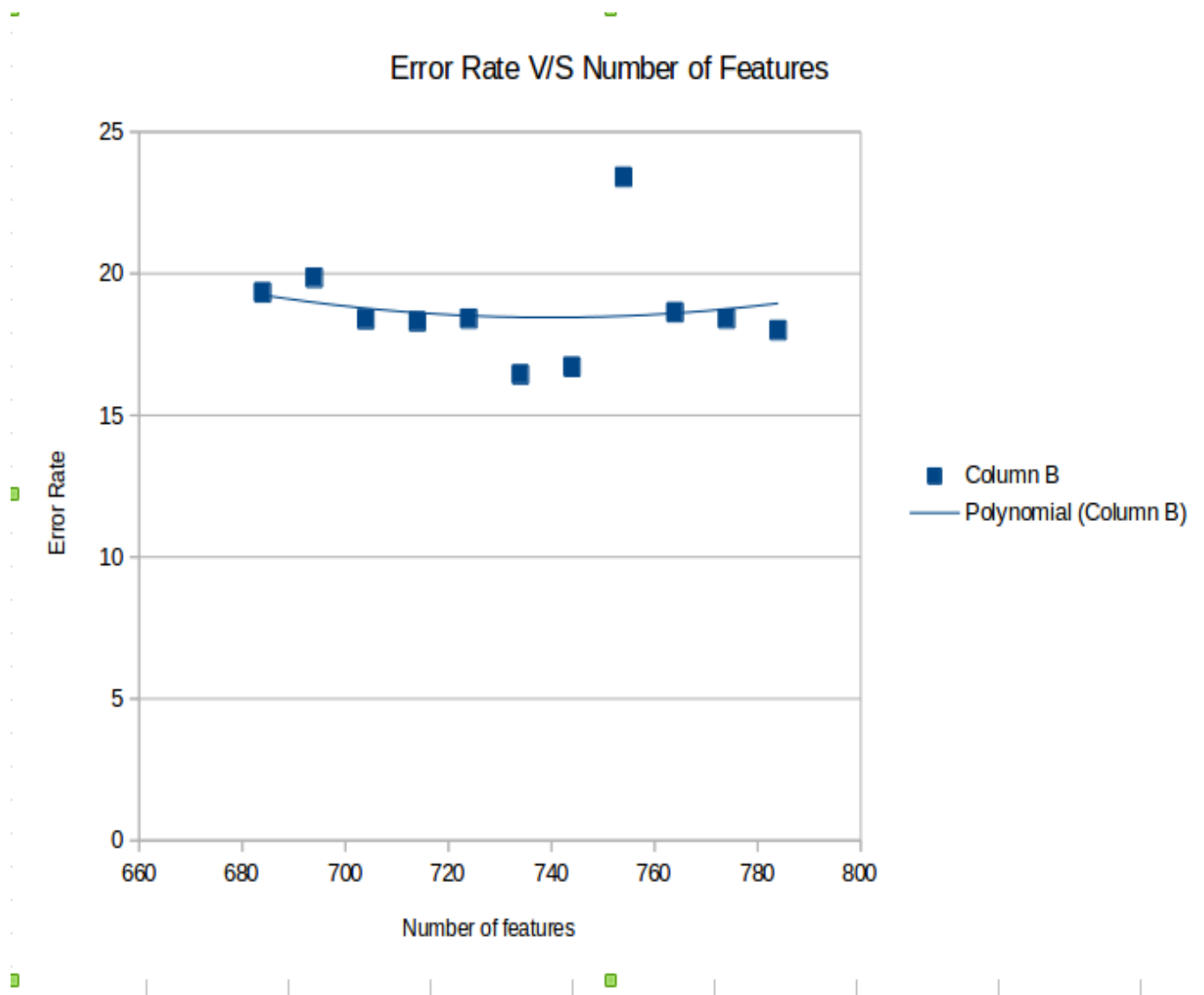
Accuracy : 83.1 %

Precision : 82.47 %
Speceficity: 98.40%
sensitivity: 82.47%
Error Rate: 16.9%

Features used:

28 X 28 pixels sized image = 784 features

Plot of number of Features V/S Error Rates



x = [684,694,704,714,724,734,744,754,764,774,784]

y =

[19.3327777778,19.8505555556, 18.3827777778, 18.3086111111 , 18.4102777778 ,
16.4361111111 , 16.7061111111 , 23.4011111111 , 18.6316666667 , 18]

Adding noise

Adding regularization term of weight decay

Accuracy:

Accuracy for fold 1	84.55
Error Rate for fold 1	15.45
Accuracy for fold 2	83.70
Error Rate for fold 2	17.391
Accuracy for fold 3	80.1
Error Rate for fold 3	19.93
Accuracy for fold 4	83.983
Error Rate for fold 4	17.016
Accuracy for fold 5	85.825
Error Rate for fold 5	15.175

Accuracy for fold	84.5 %
Error Rate for fold	15.5 %