

## HW1 - K Nearest Neighbors

1. What is the role of the number of training points to accuracy?

For the (figure 1a)  $k = 5$  and 500 training points accuracy is 83.11%. And, (figure 1b) 2000 training points with same  $k$  accuracy is 90.94%. So, when training points increases, accuracy also increases.

	0	1	2	3	4	5	6	7	8	9
0:	910	1	9	1	5	17	26	6	2	14
1:	0	1051	2	2	1	5	0	3	0	0
2:	8	79	741	28	21	12	7	63	21	10
3:	5	17	8	883	4	52	6	10	27	18
4:	0	26	1	0	765	0	7	28	1	155
5:	22	21	1	0	82	13	660	34	14	25
6:	16	34	0	0	28	15	863	1	1	0
7:	11	42	2	10	63	13	49	3	0	937
8:	9	8	2	14	60	6	5	23	30	696
9:	6	8	2	14	60	6	5	23	30	696

Accuracy: 0.831100

(a) Training\_500\_k5

	0	1	2	3	4	5	6	7	8	9
0:	856	3	7	3	0	6	8	5	0	3
1:	0	1053	6	1	2	0	0	2	0	0
2:	9	29	871	14	8	4	6	33	11	5
3:	1	5	11	915	2	46	3	10	29	8
4:	1	24	1	0	857	1	5	8	0	86
5:	8	6	1	32	4	806	27	7	6	18
6:	11	2	1	0	7	941	0	0	0	0
7:	0	23	0	0	9	2	0	1025	1	30
8:	6	40	11	27	8	31	18	20	816	30
9:	5	6	2	10	34	7	1	41	1	854

Accuracy: 0.909400

(b) Training\_2000\_k5

Figure 1

2. What is the role of  $k$  to accuracy?

The parameter  $k$  is important part of the KNN algorithm. For the training example (2000) and  $k = 5, 7, 15$  accuracy are (figure 1b) 90.94%, (figure 2a) 89.39% ,and (figure 2b) 88.05% respectively. Form the result, it is obvious that the accuracy decrease when you increase  $k$  and the computation cost also increases.

	0	1	2	3	4	5	6	7	8	9
0:	938	1	11	4	2	8	20	4	0	3
1:	0	1057	1	2	1	1	0	4	0	0
2:	8	39	824	23	15	10	5	52	9	5
3:	1	8	9	920	1	45	2	13	24	7
4:	0	27	0	0	841	1	8	20	1	85
5:	4	11	29	10	788	27	11	1	19	17
6:	8	8	2	1	5	11	932	0	0	0
7:	0	32	1	0	8	3	0	1021	0	25
8:	6	46	6	48	11	52	23	23	767	27
9:	3	9	1	10	31	11	1	51	3	841

Accuracy: 0.893900

(a) Training\_2000\_k7

	0	1	2	3	4	5	6	7	8	9
0:	920	2	8	6	2	15	28	6	2	2
1:	0	1057	1	1	2	1	0	2	0	0
2:	5	53	803	24	23	10	4	52	11	5
3:	1	12	7	918	13	31	3	18	19	8
4:	0	32	0	0	824	3	8	29	1	86
5:	2	14	1	33	14	776	36	13	9	17
6:	8	8	2	0	17	12	920	0	0	0
7:	0	41	0	0	5	0	1015	0	22	0
8:	4	54	8	43	14	61	26	22	747	30
9:	4	8	1	11	31	8	3	66	4	825

Accuracy: 0.880500

(b) Training\_2000\_k15

Figure 2

3. What numbers get confused with each other most easily?

The most confuse numbers of pair is 4 and 9. It can be proven from highest number of false prediction from confusion matrix.

	0	1	2	3	4	5	6	7	8	9
0:	982	0	3	0	0	0	2	1	1	2
1:	0	1060	1	0	1	0	1	1	0	0
2:	3	6	955	3	1	1	1	18	2	0
3:	0	0	4	1004	0	9	1	3	6	3
4:	0	9	0	0	951	0	0	4	0	19
5:	2	0	1	16	2	871	16	3	1	3
6:	1	0	0	0	0	2	964	0	0	0
7:	0	9	0	0	2	0	0	1073	0	6
8:	2	6	1	12	4	18	6	5	948	7
9:	2	2	0	8	11	5	0	11	3	919

Accuracy: 0.972700

Figure 3: Training\_10e3\_k5