
Programming 3

Samuel Cuthbertson

SAMUEL.CUTHBERTSON@COLORADO.EDU

1. Analysis

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

Number of training points has a direct correlation to accuracy, at least while using this data set. The more the training points, the higher the accuracy. This is because the more training points we have, the higher our likelihood of having more training points that are very similar (close) to a given test point are.

1.2. What is the role of k to accuracy?

k has a complicated effect on accuracy. This is most clearly seen with a low number of training points, as shown in the table below on the left, which was generated with 50 training points. The table on the right was generated with 5000. For both, there's a range of k 's around 3 which produce the best results for this dataset. Any larger or smaller and accuracy falls.

k	Accuracy (%)	k	Accuracy (%)
1	0.643300	1	0.938800
3	0.574400	3	0.941000
5	0.546800	5	0.939000
7	0.503100	7	0.933700
9	0.463700	9	0.933500
11	0.433700	11	0.931100

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

	0	1	2	3	4	5	6	7	8	9
0	979	1	2	1	0	1	4	2	0	1
1	0	1058	3	0	1	0	0	1	1	0
2	14	26	907	5	2	0	3	19	7	7
3	2	4	12	969	1	18	0	3	15	6
4	2	18	0	0	895	0	3	4	1	60
5	6	7	1	31	3	813	22	0	16	16
6	11	1	0	0	2	2	949	0	2	0
7	0	19	3	0	4	0	0	1037	0	27
8	6	25	4	18	2	17	4	7	904	22
9	6	6	0	8	14	4	0	19	5	899

The table above shows the confusion matrix generated using 5000 training points and $k = 9$. The most common wrong guess for each number is highlighted in red. For example, 4's are most commonly wrongfully identified 9's.