

Zhanchong Deng
A51491777
1/29/2020

HW1

1. For $k = 1, 5, 9$ and 15 , build k -nearest neighbor classifiers from the training data. For each of these values of k , write down a table of training errors (error on the training data) and the validation errors (error on the validation data). Which of these classifiers performs the best on validation data? What is the test error of this classifier? [Hint: As a check for your code, the training error for $k = 3$ should be about 0.04 .]

Errors:

k	training_error	runtime	validation_error	runtime	test_error	runtime
1	0	61.77	0.082	31.09	0.094	31.11
5	0.055	62.13	0.095	31.09	0.097	31.56
9	0.0685	62.68	0.106	31.21	0.102	31.71
15	0.0915	63.00	0.106	31.31	0.114	31.70

The classifier with $k=1$ perform the best on the validation data. The test error of this classifier is 0.094 .

2.

[Use projection]

How is the classification accuracy affected by projection? How does the running time of your program change when you run it on projected data?

k	training_error	runtime	validation_error	runtime	test_error	runtime
1	0	55.91	0.32	27.61	0.314	27.52
5	0.196	56.00	0.298	27.91	0.293	28.06
9	0.229	56.33	0.29	28.23	0.287	28.24
15	0.257	57.01	0.287	28.49	0.306	28.74

The accuracy is decreased dramatically. It is expected because projection blurs the data thus the model can become inaccurate. However, that definitely saves time as most of the prediction takes a few seconds less time consistently. This is also expected because there are less operation conducted due to less dimension of the vectors.