CS 545 Machine Learning
Homework 5: K-Means Clustering
Haomin He

Experiment 1:

For experiment 1, K = 10, I obtain 10 final clusters. I repeat K-Means Clustering run for 5 times. I choose the run out of 5 that yields the smallest average mean-square-error (Average mean-square-error: 645.6238769482709). For this best run, I calculate mean-square-separation, and mean entropy (Mean-square-separation: 1319.09848419583), (Mean entropy: 0.740501141) of the resulting clustering on the training data. I calculate the accuracy on the test data (Accuracy: 0.735671). Below is a test data results confusion matrix. I visualize the resulting cluster centers by drawing the corresponding digit on an 8 x 8 grid.

Confusion matrix:

176	0	0	0	2	0	0	0	0	0
0	62	21	1	0	0	4	0	94	0
1	2	150	8	0	0	0	3	13	0
0	0	0	165	0	1	0	7	8	2
0	4	0	0	150	0	0	3	8	16
0	0	0	27	1	151	1	0	0	2
1	1	0	0	1	0	176	0	2	0
0	0	0	0	0	0	0	145	3	31
0	6	1	32	0	2	2	1	120	10
0	3	0	144	0	3	0	2	1	27

Visualization results: [8, 2, 3, 0, 1, 5, 4, 9, 7, 6]. Yes, the visualized cluster centers look like their associated digits.

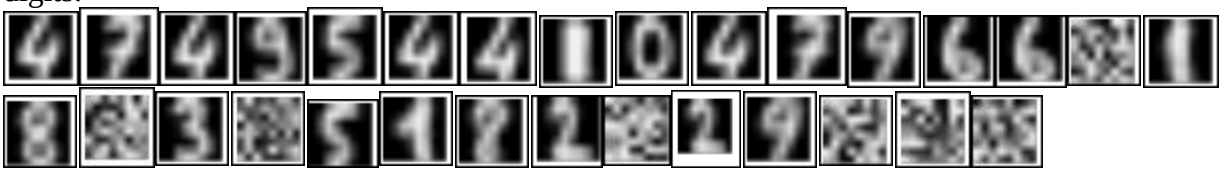
Experiment 2:

For experiment 2, K = 30, I obtain 30 final clusters. I repeat K-Means Clustering run for 5 times. I choose the run out of 5 that yields the smallest average mean-square-error (Average mean-square-error: 510.005680121353). For this best run, I calculate mean-square-separation, and mean entropy (Mean-square-separation: 1513.2148875487783), (Mean entropy: 0.735540447) of the resulting clustering on the training data. I calculate the accuracy on the test data (Accuracy: 0.89705064). Below is a test data results confusion matrix. I visualize the resulting cluster centers by drawing the corresponding digit on an 8 x 8 grid.

Confusion matrix:

176	0	0	0	2	0	0	0	0	0
0	153	20	1	0	0	3	0	2	3
1	5	156	0	0	0	0	3	12	0
0	0	2	146	0	3	0	4	12	16
0	5	0	0	169	0	0	0	3	4
0	0	0	0	1	166	1	0	0	14
0	1	0	0	1	1	176	0	2	0
0	0	0	0	1	0	0	165	5	8
0	21	1	1	0	1	1	1	140	8
0	3	0	3	0	4	0	1	4	165

Visualization results: [4, 7, 4, 9, 5, 4, 4, 1, 0, 4, 7, 9, 6, 6, None, 1, 8, None, 3, None, 5, 1, 8, 2, None, 2, 9, None, None, None]. Yes, the visualized cluster centers look like their associated digits.



Compare the results of Experiments 1 and 2:

	Experiment 1	Experiment 2
Mean-square-square error	645.6238769	510.0056801
Mean-square-separation	1319.098484	1513.214888
Mean entropy	0.740501141	0.735540447
Accuracy	0.735671	0.89705064