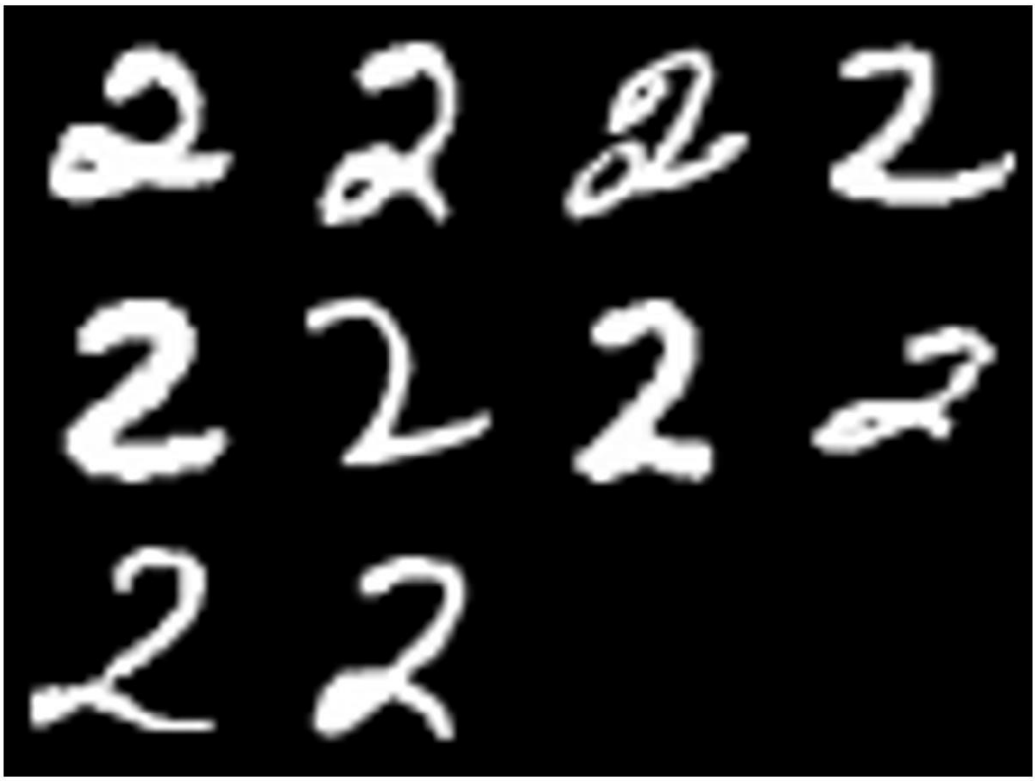
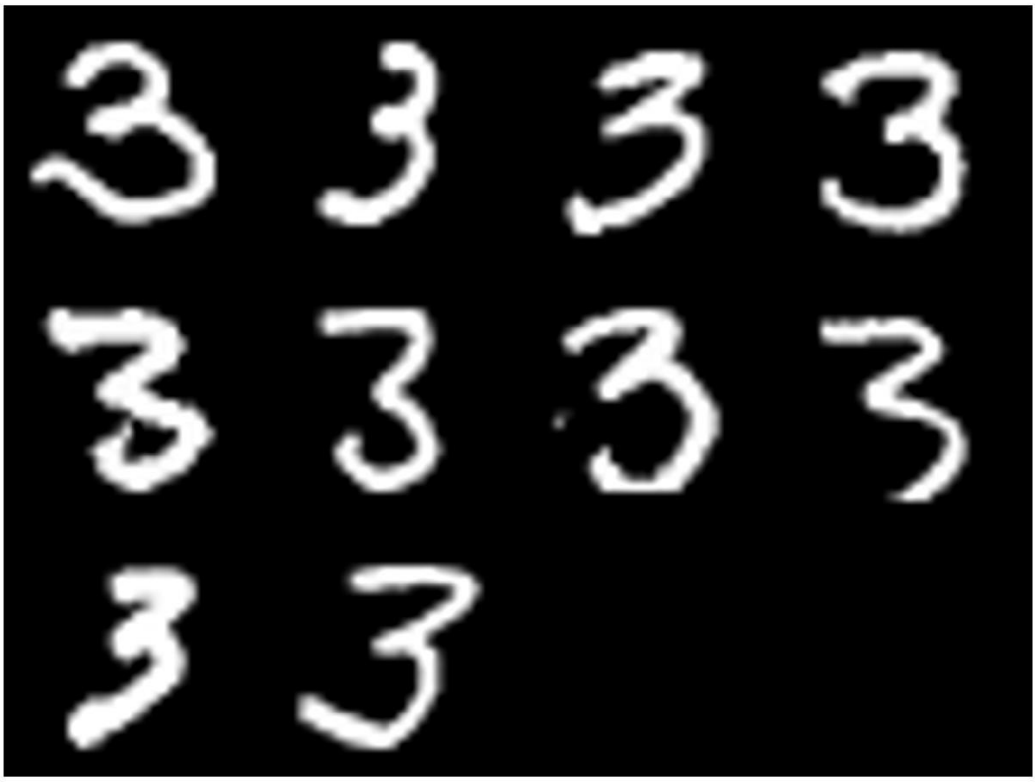
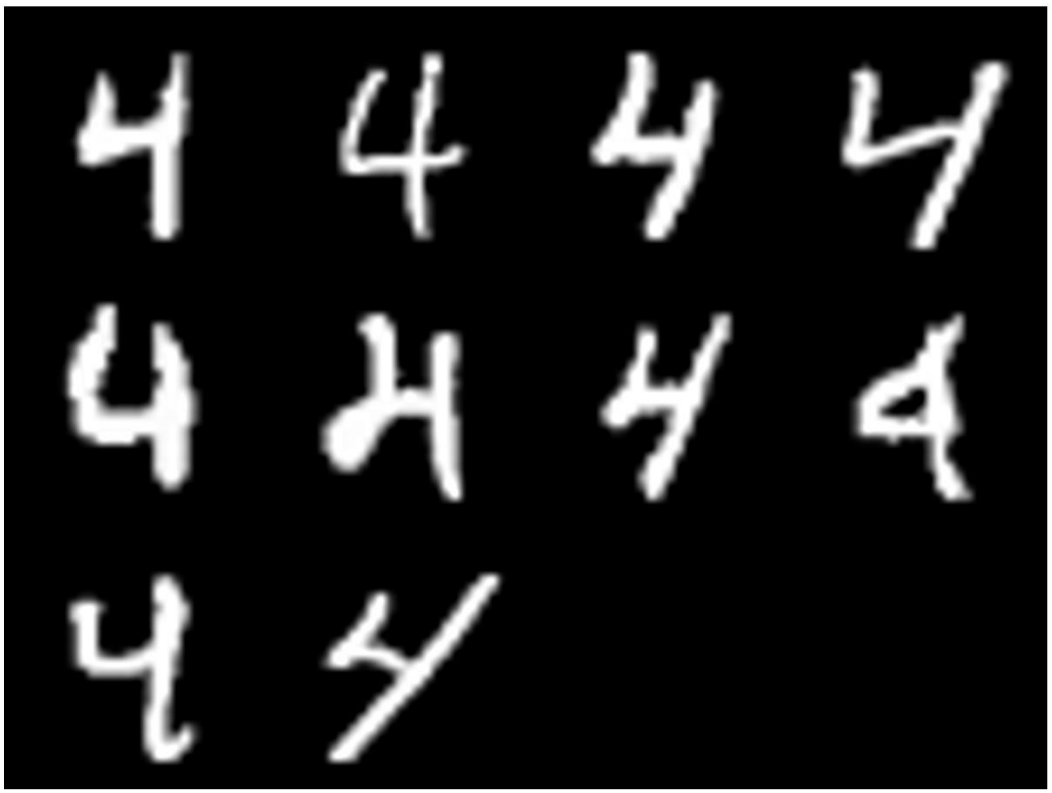
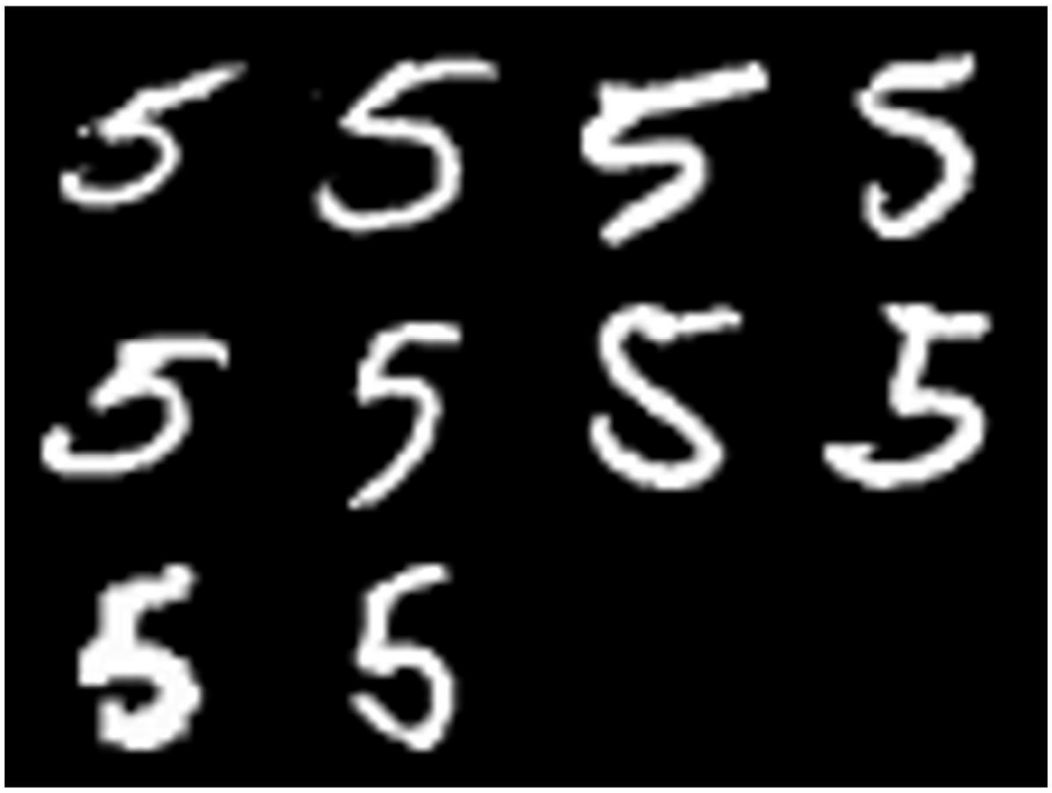
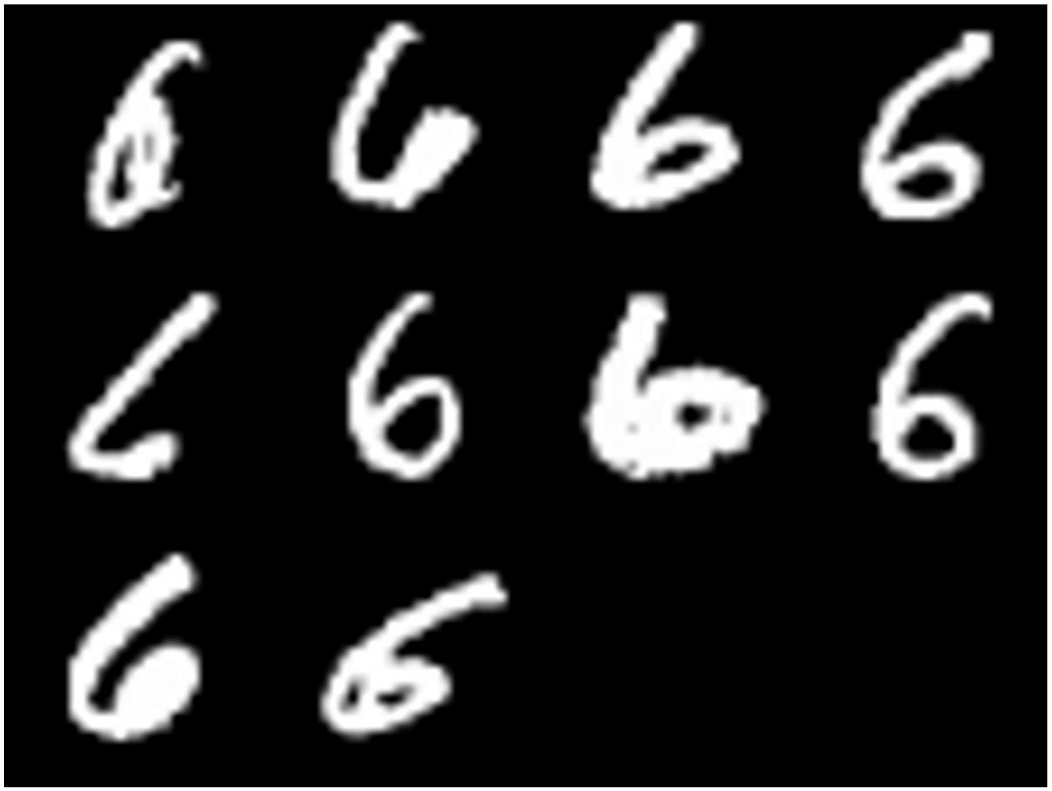
Task 1.1

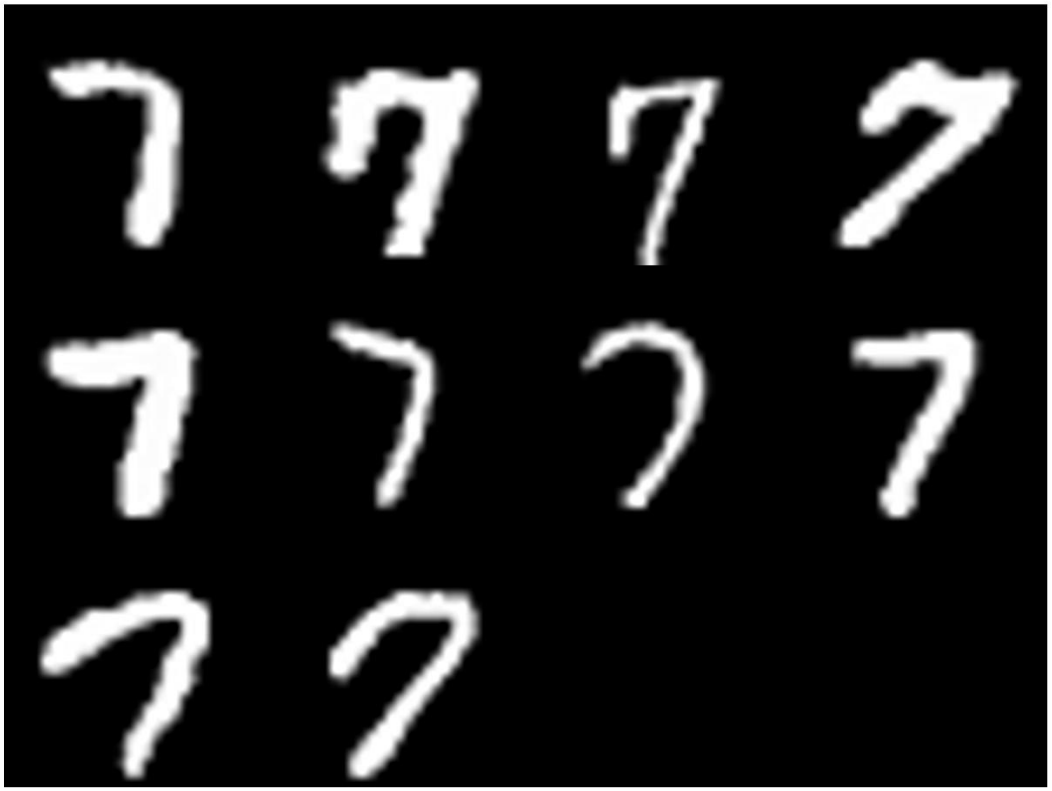


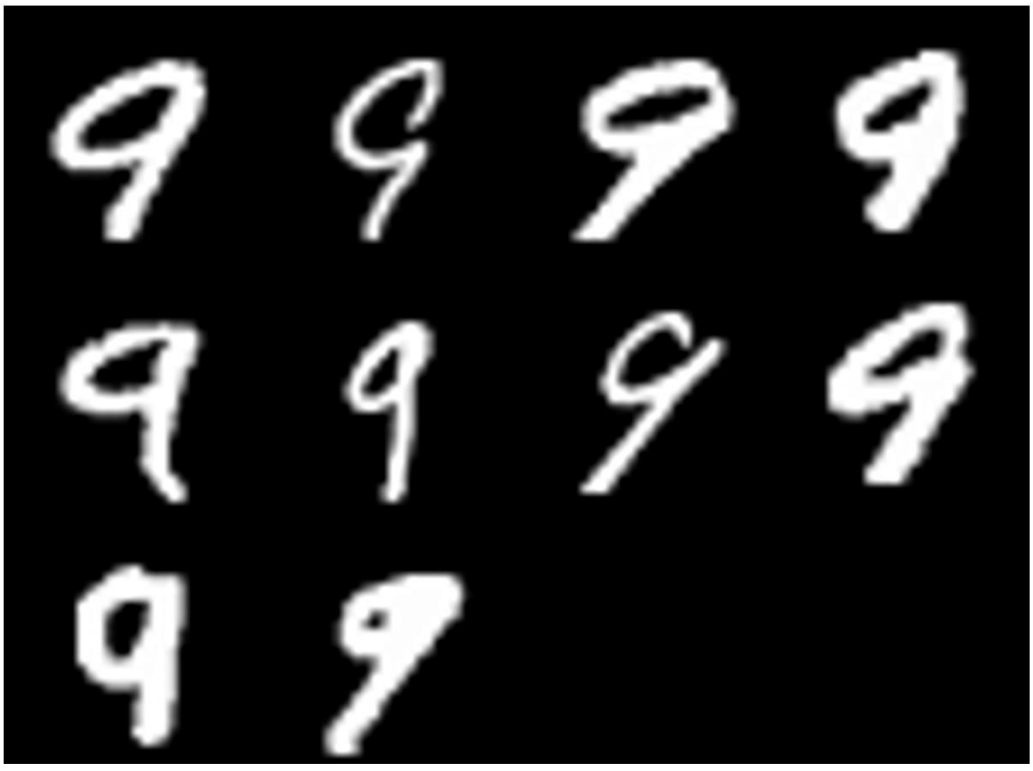




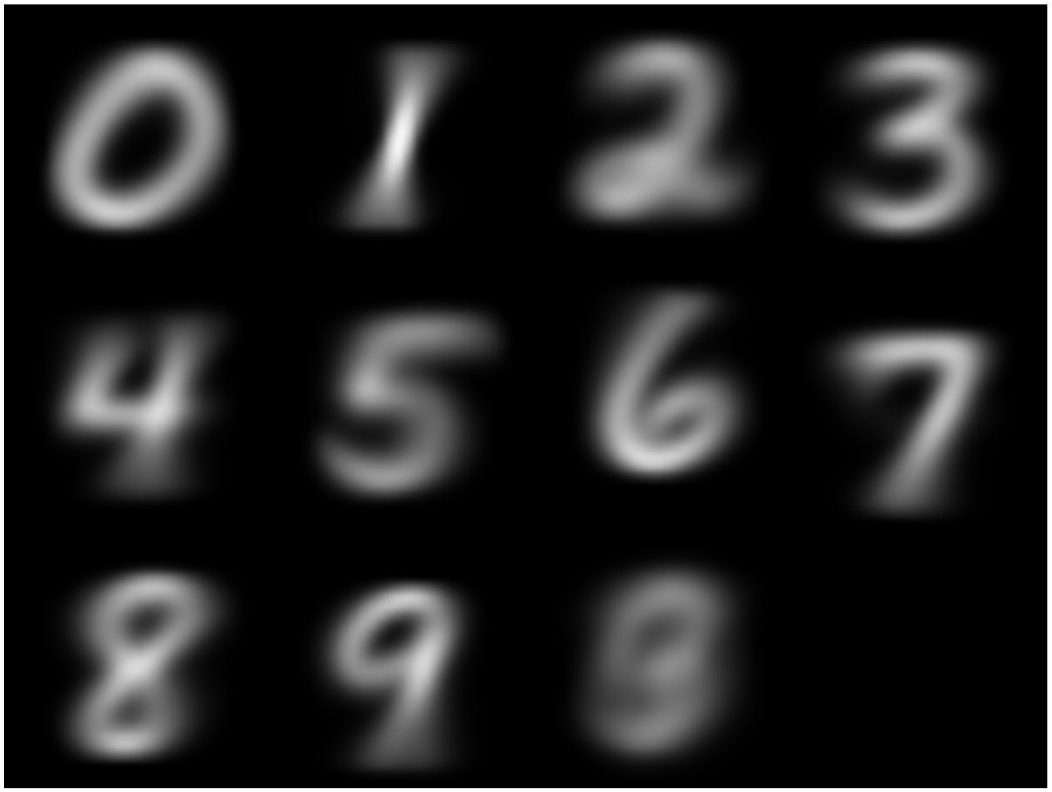




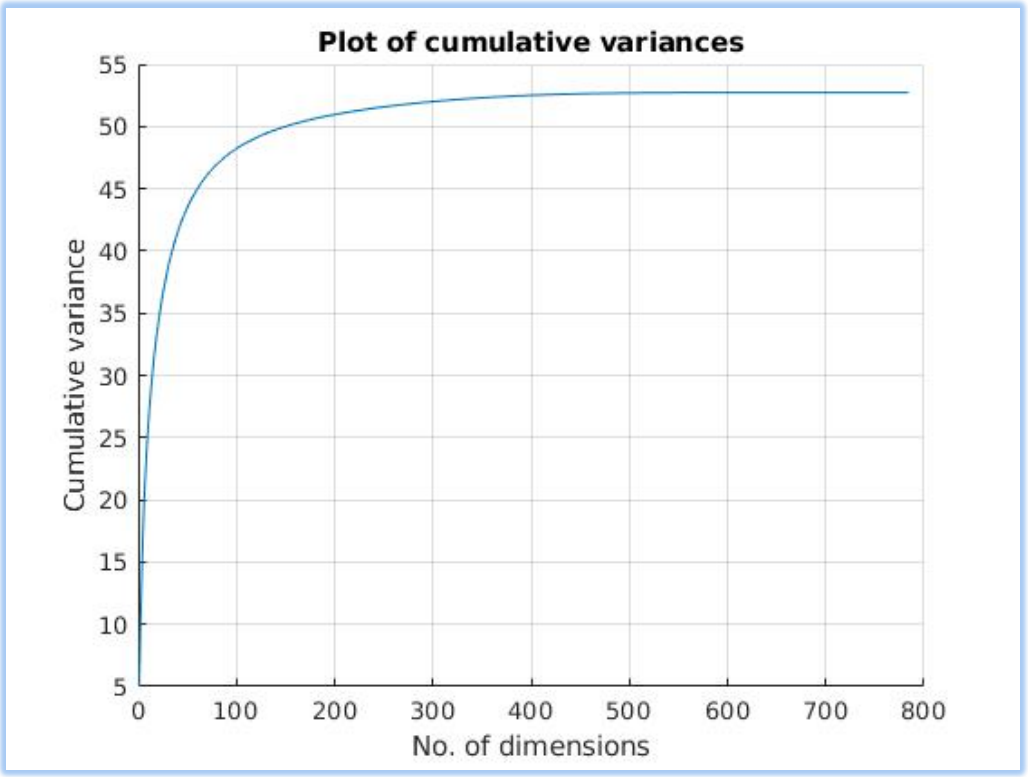




Task 1.2

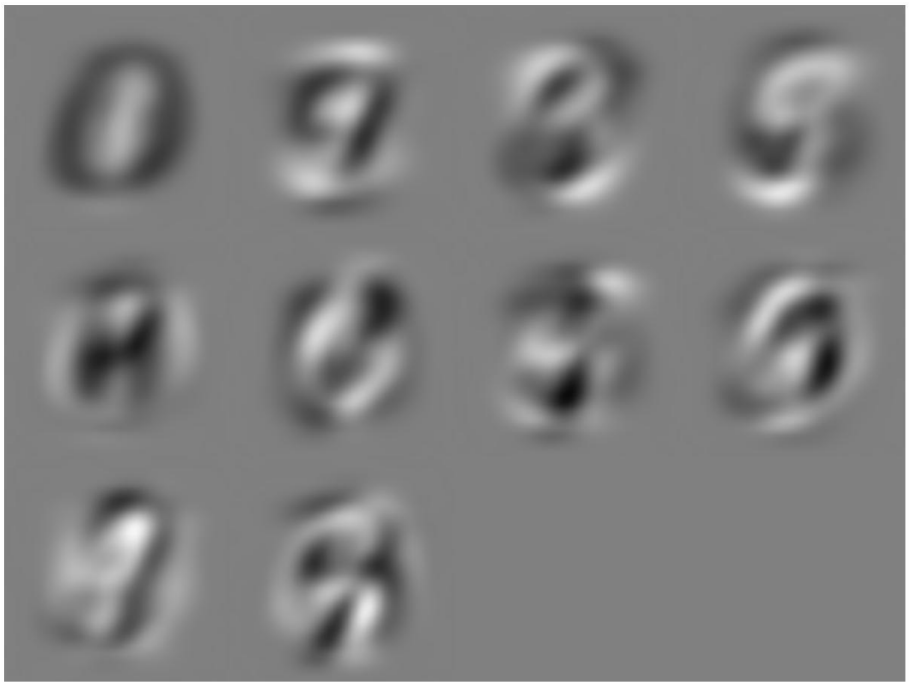


Task 1.3

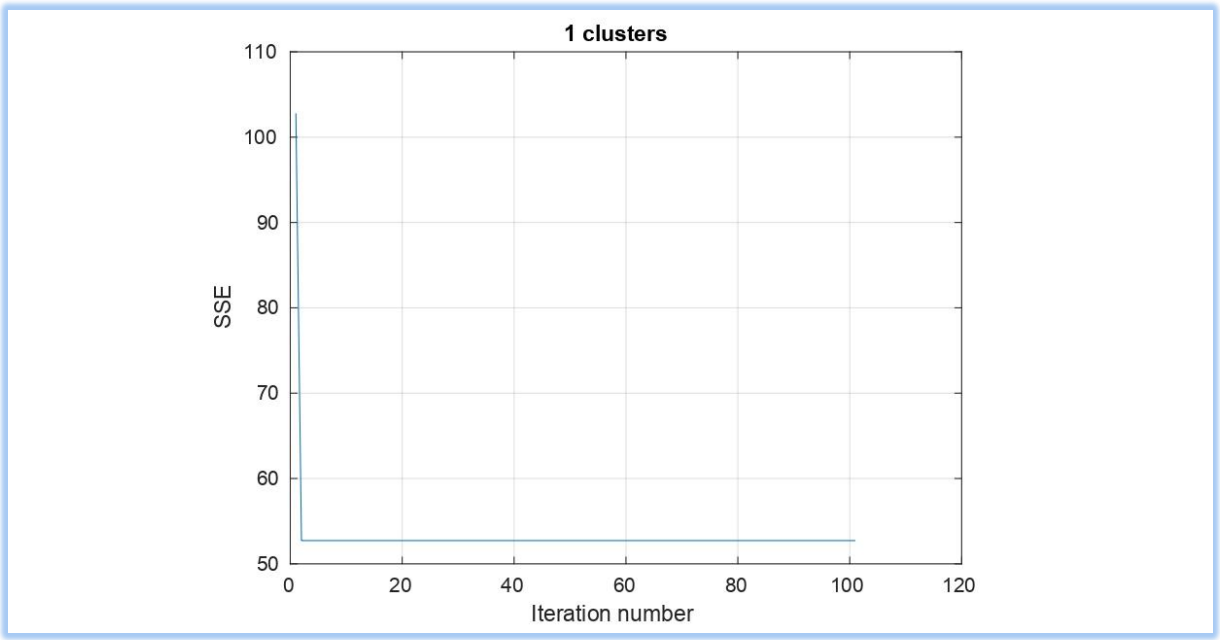


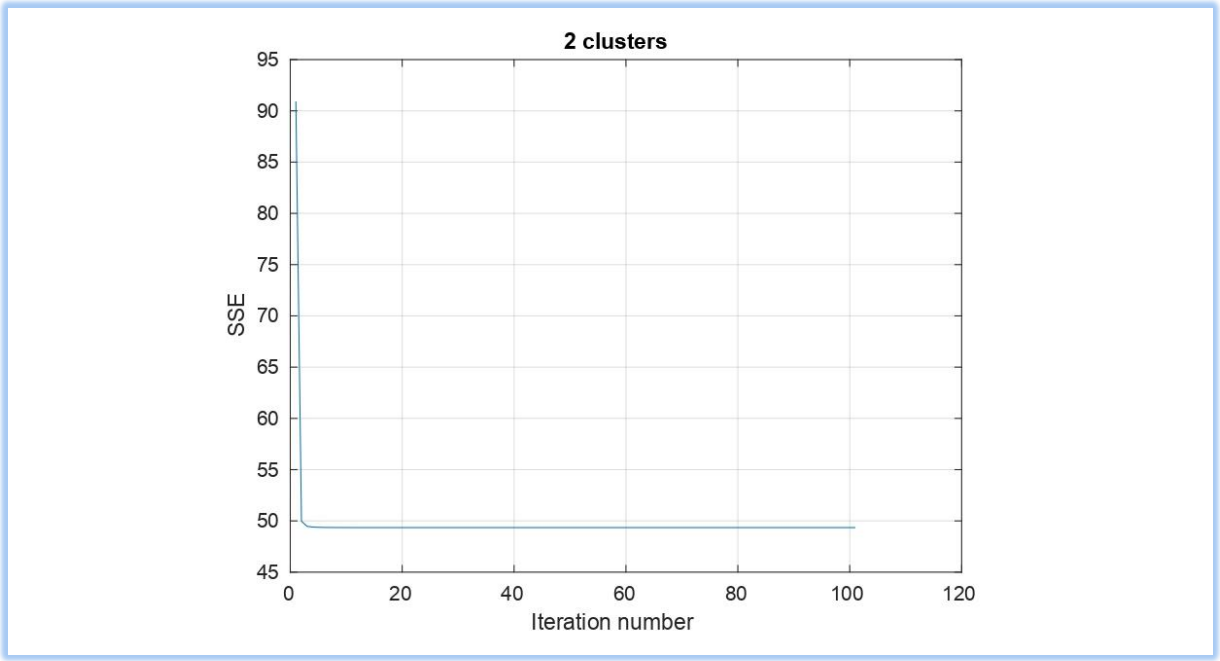
MinDims = [27; 44; 88; 154]

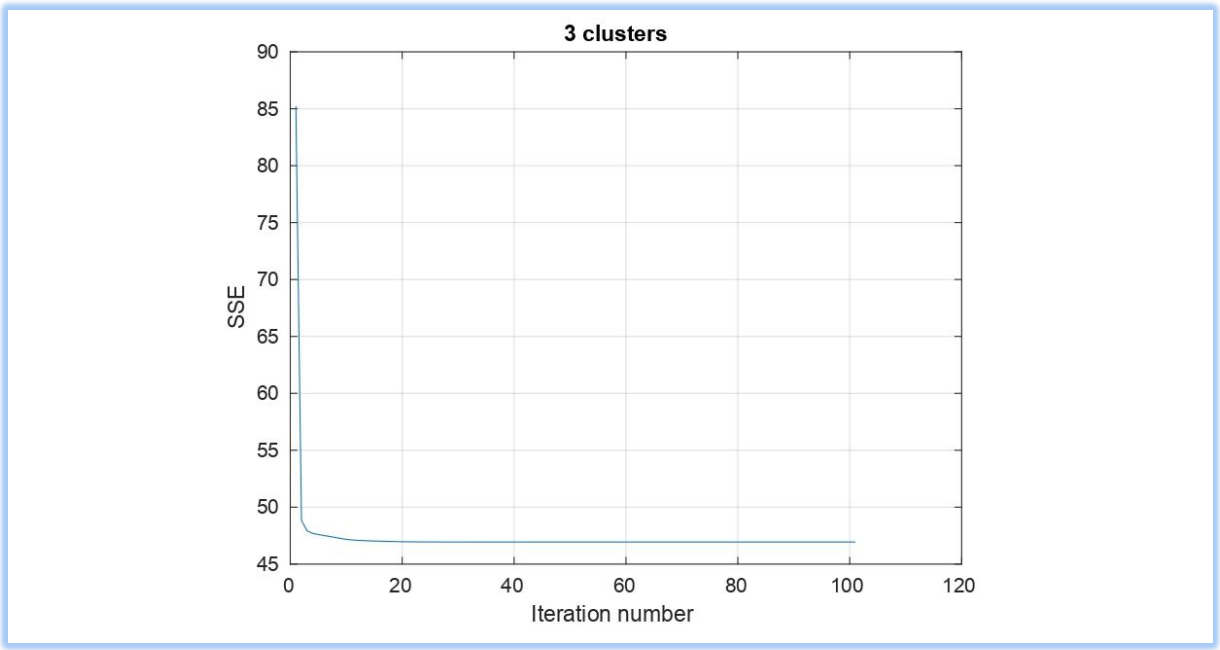
Task 1.4

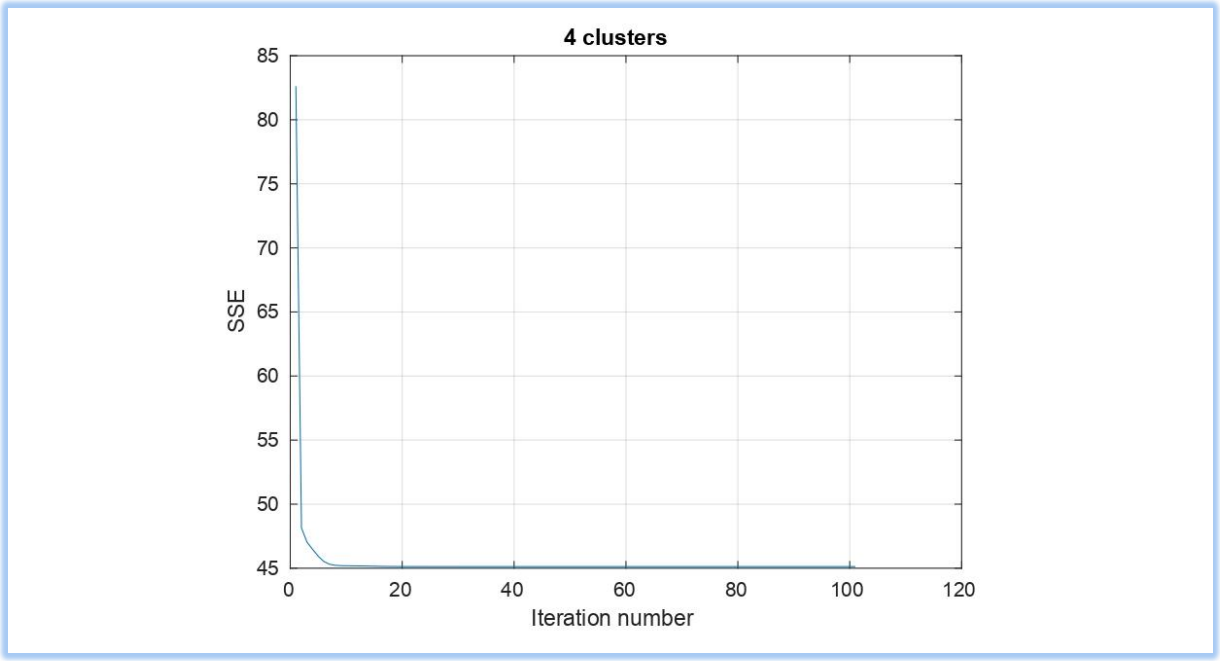


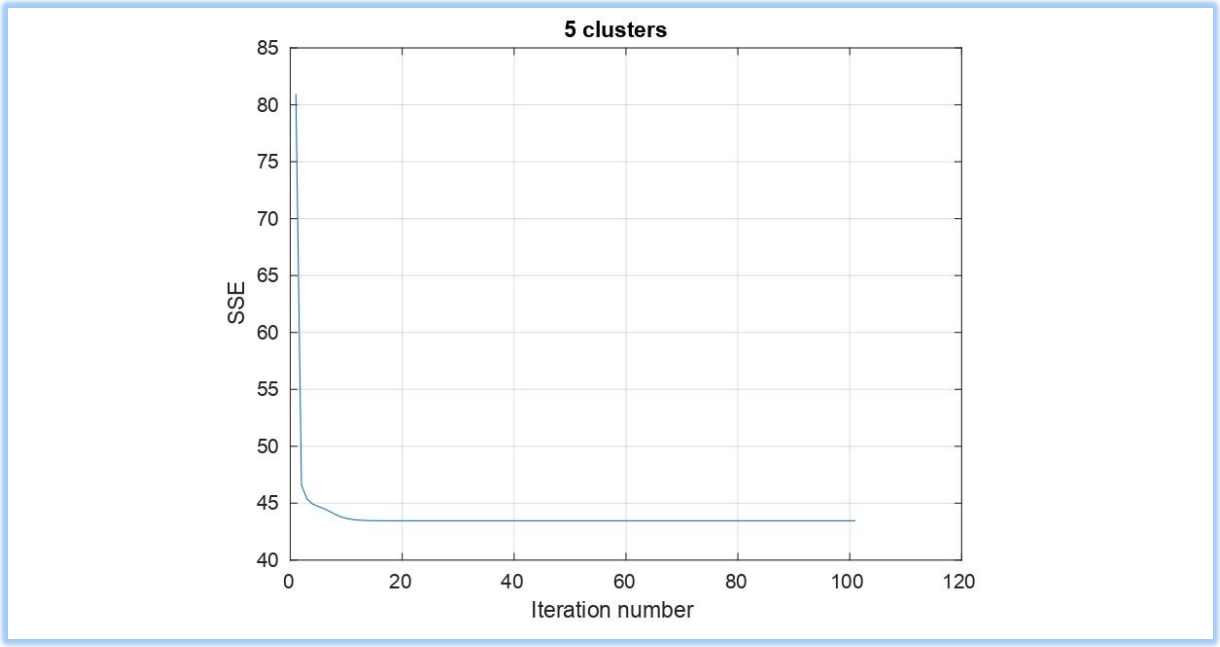
Task 1.5

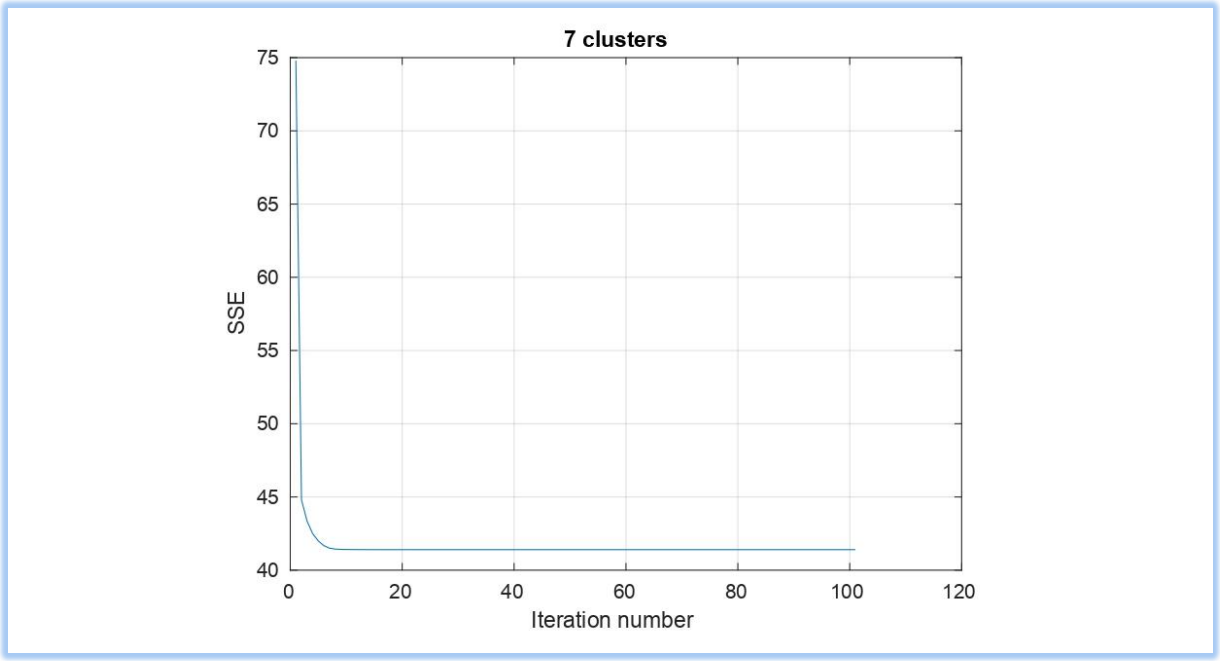


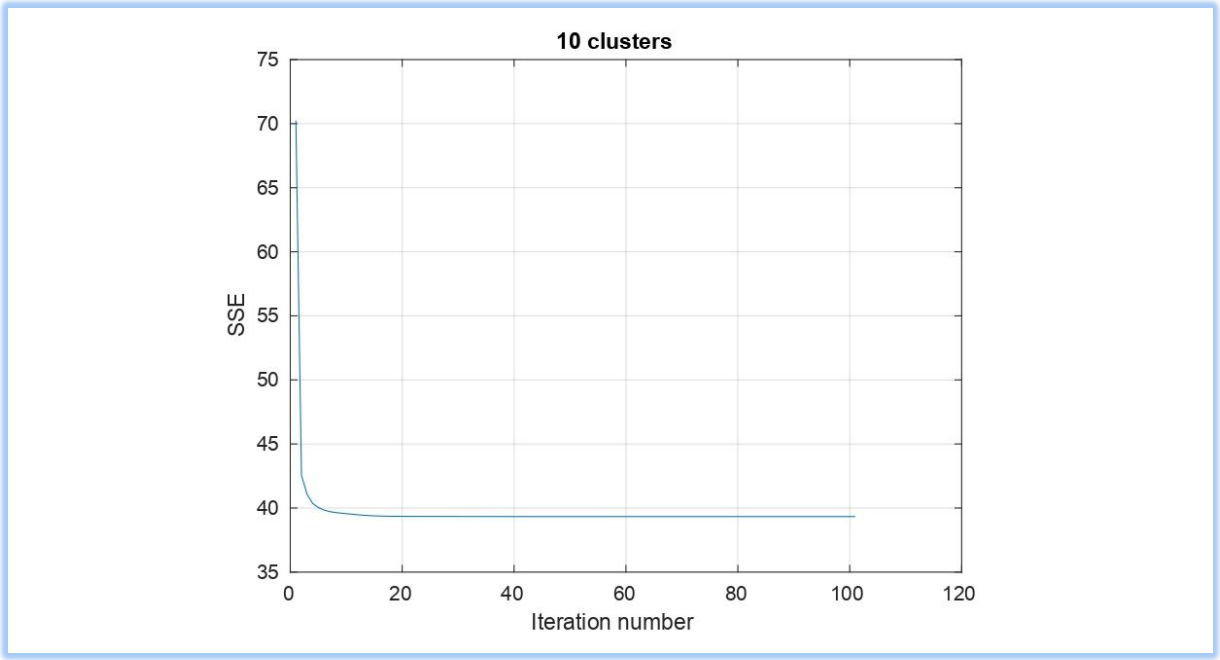

2 clusters
SSE
Iteration number
95
90
85
80
75
70
65
60
55
50
45
0
20
40
60
80
100
120

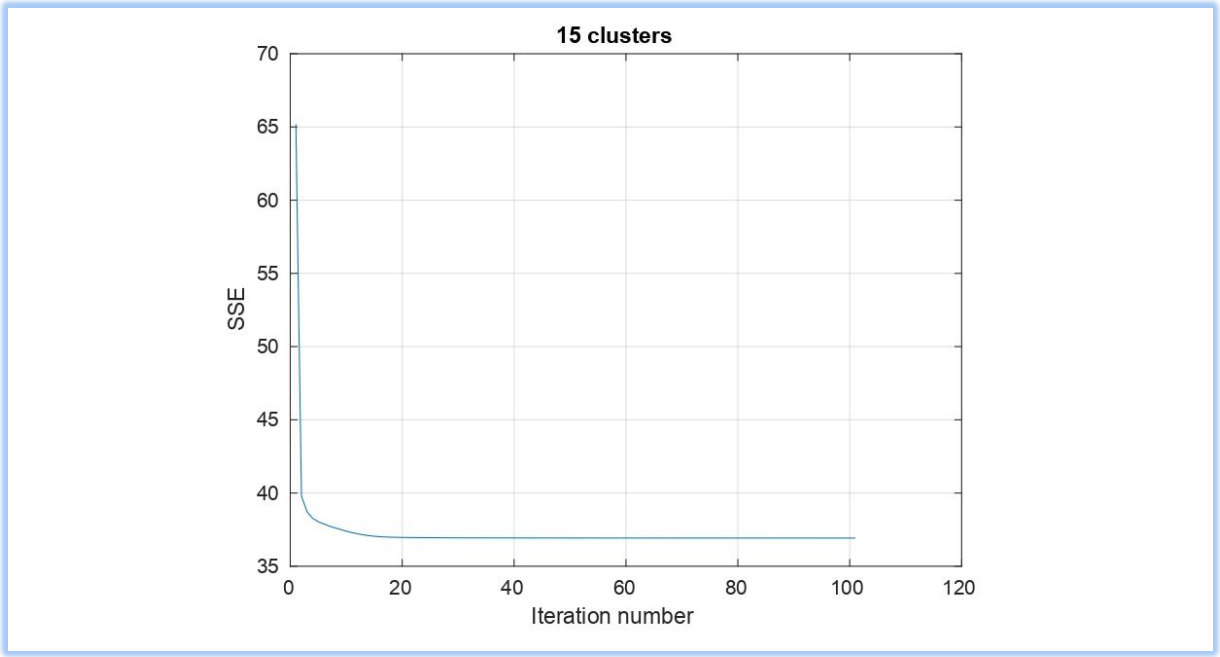

3 clusters
SSE
Iteration number
90
85
80
75
70
65
60
55
50
45
0
20
40
60
80
100
120

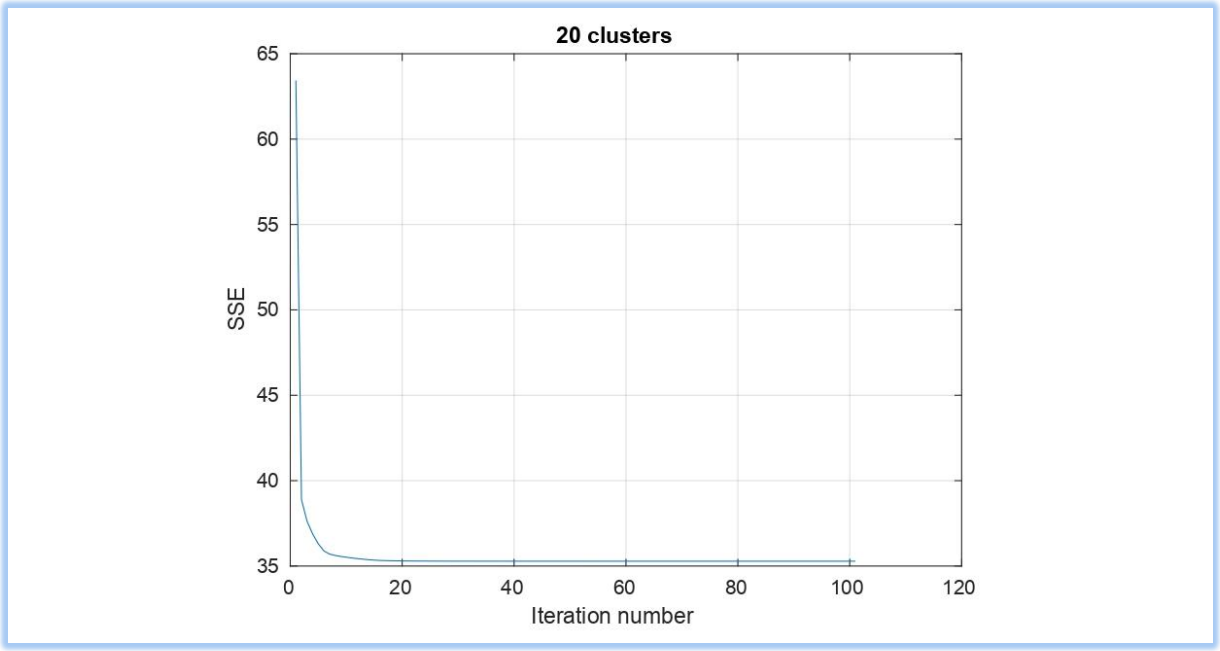

4 clusters
SSE
Iteration number
85
80
75
70
65
60
55
50
45
0
20
40
60
80
100
120


5 clusters
SSE
Iteration number
85
80
75
70
65
60
55
50
45
40
0
20
40
60
80
100
120


7 clusters
SSE
Iteration number
75
70
65
60
55
50
45
40
0
20
40
60
80
100
120


10 clusters
SSE
Iteration number
75
70
65
60
55
50
45
40
35
0
20
40
60
80
100
120


15 clusters
70
65
60
55
50
45
40
35
SSE
0
20
40
60
80
100
120
Iteration number


20 clusters
65
60
55
50
45
40
35
SSE
0
20
40
60
80
100
120
Iteration number

Timings for each k:

k	Time [seconds]
1	28.765397
2	30.597430
3	33.692614
4	33.937943
5	34.580180
7	36.479291
10	36.095470
15	37.558576
20	39.313538

Task 1.6

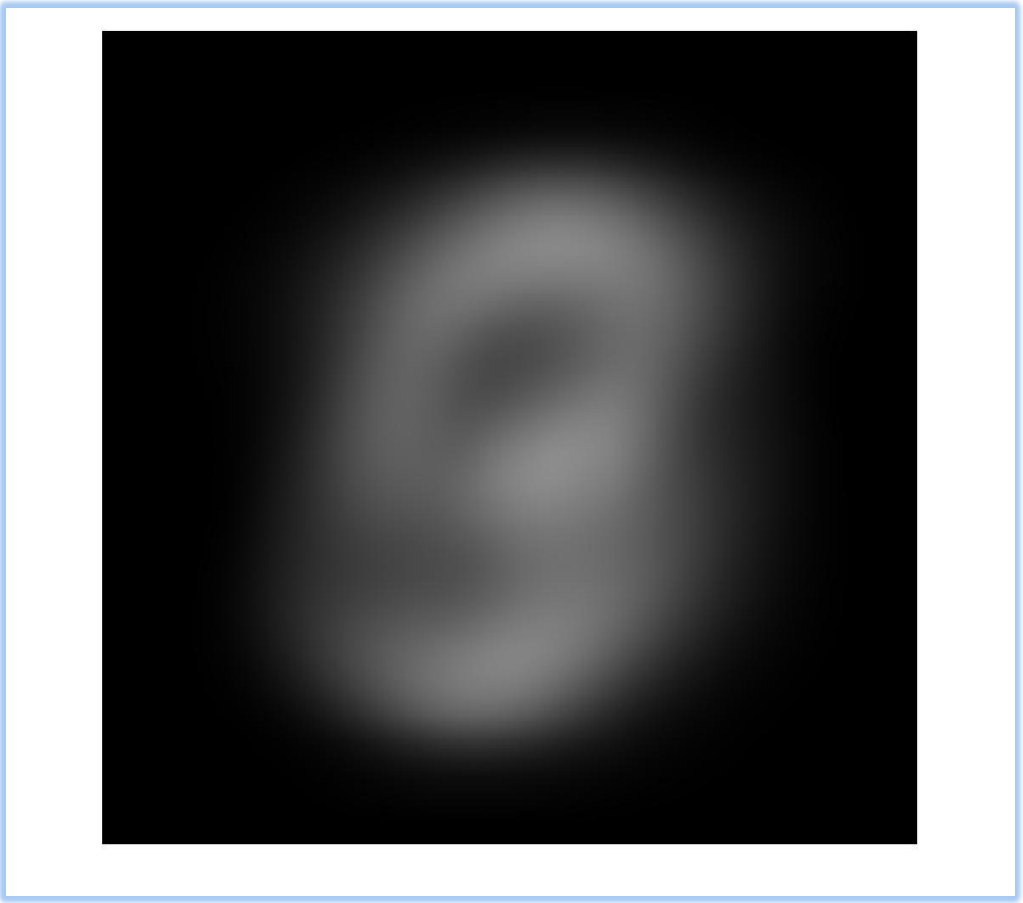


Figure 1: 1 cluster

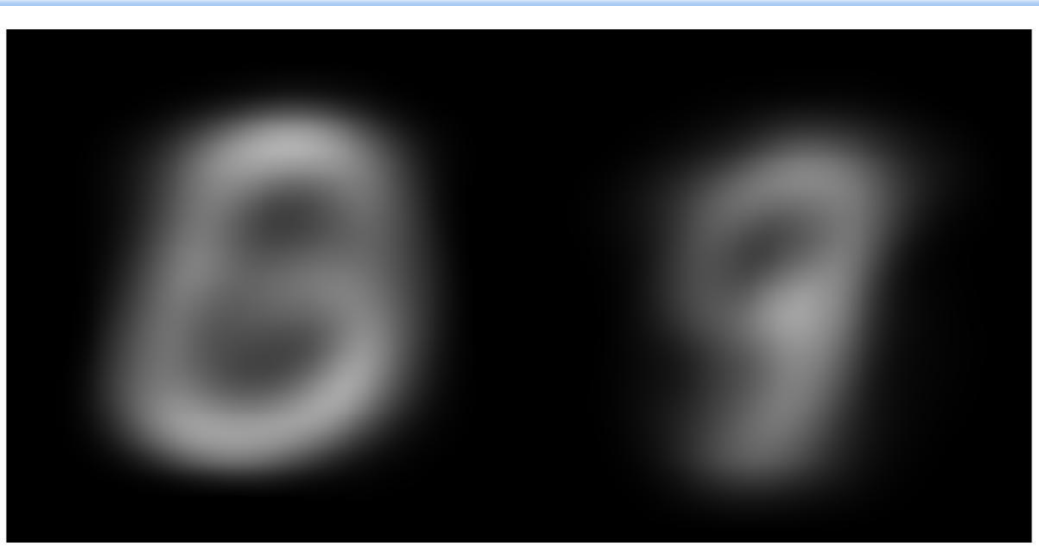


Figure 2: 2 clusters

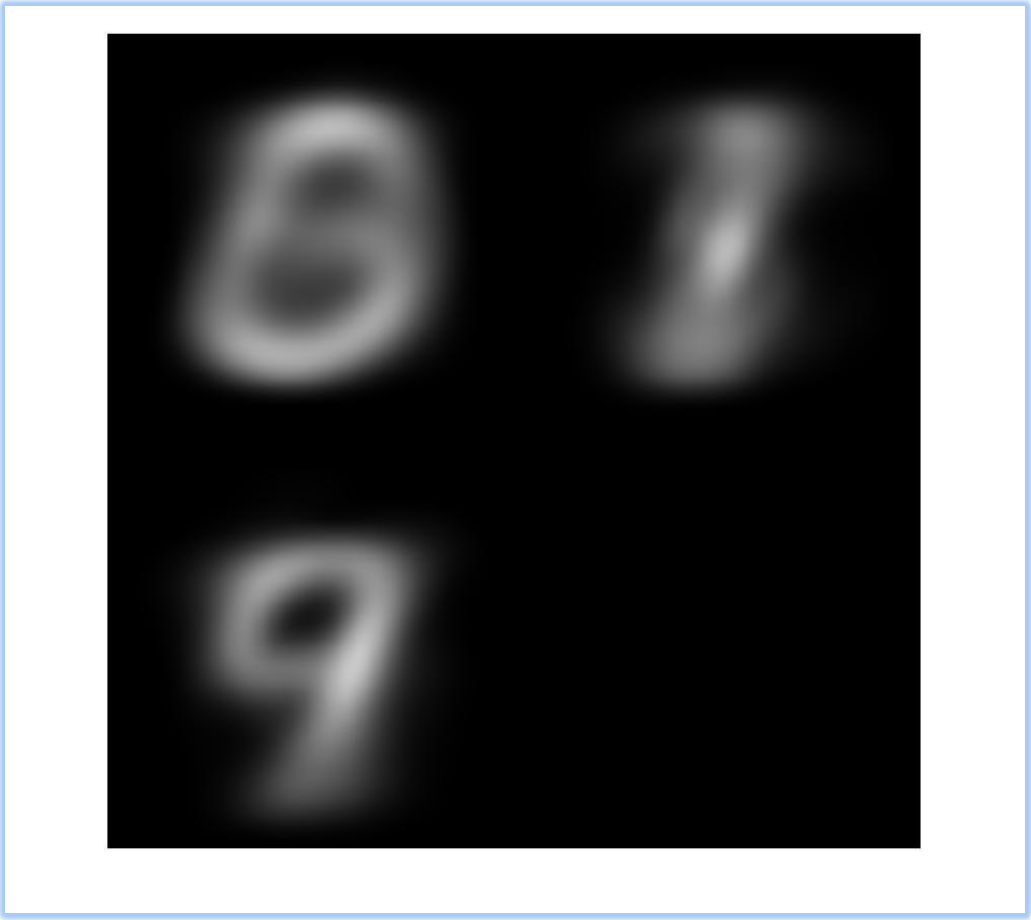


Figure 3: 3 clusters

Figure 4: 4 clusters

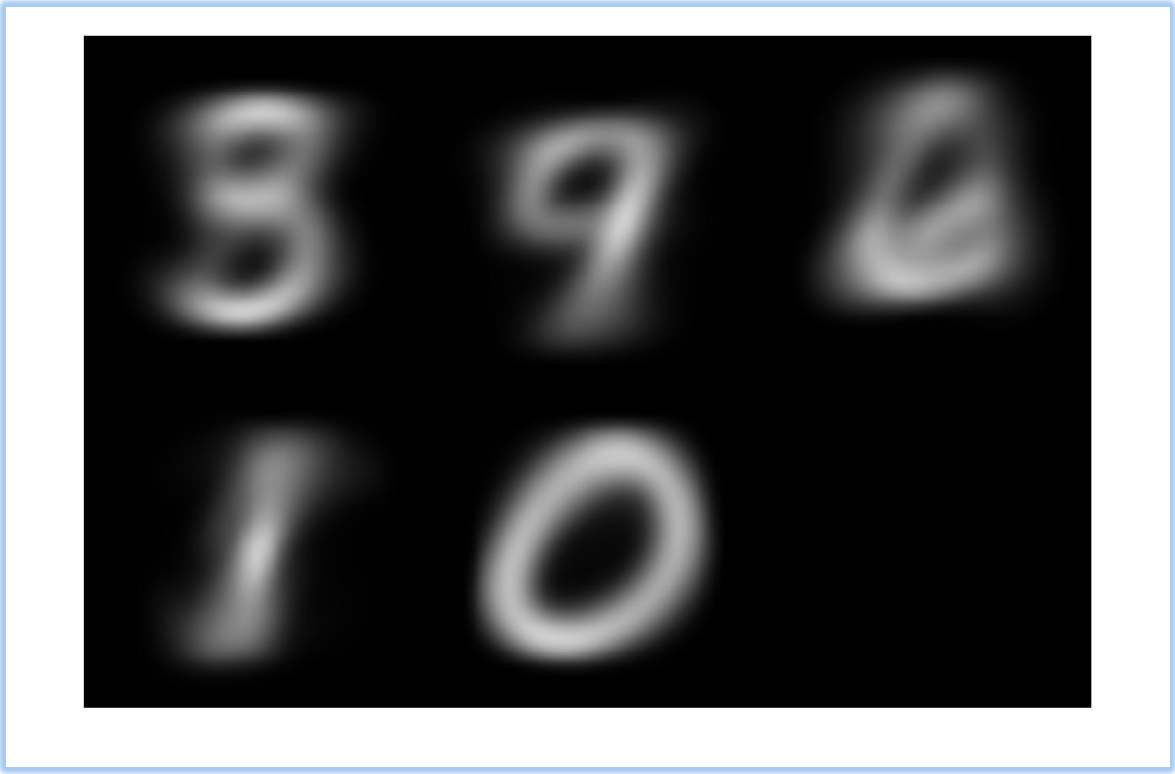


Figure 5: 5 clusters

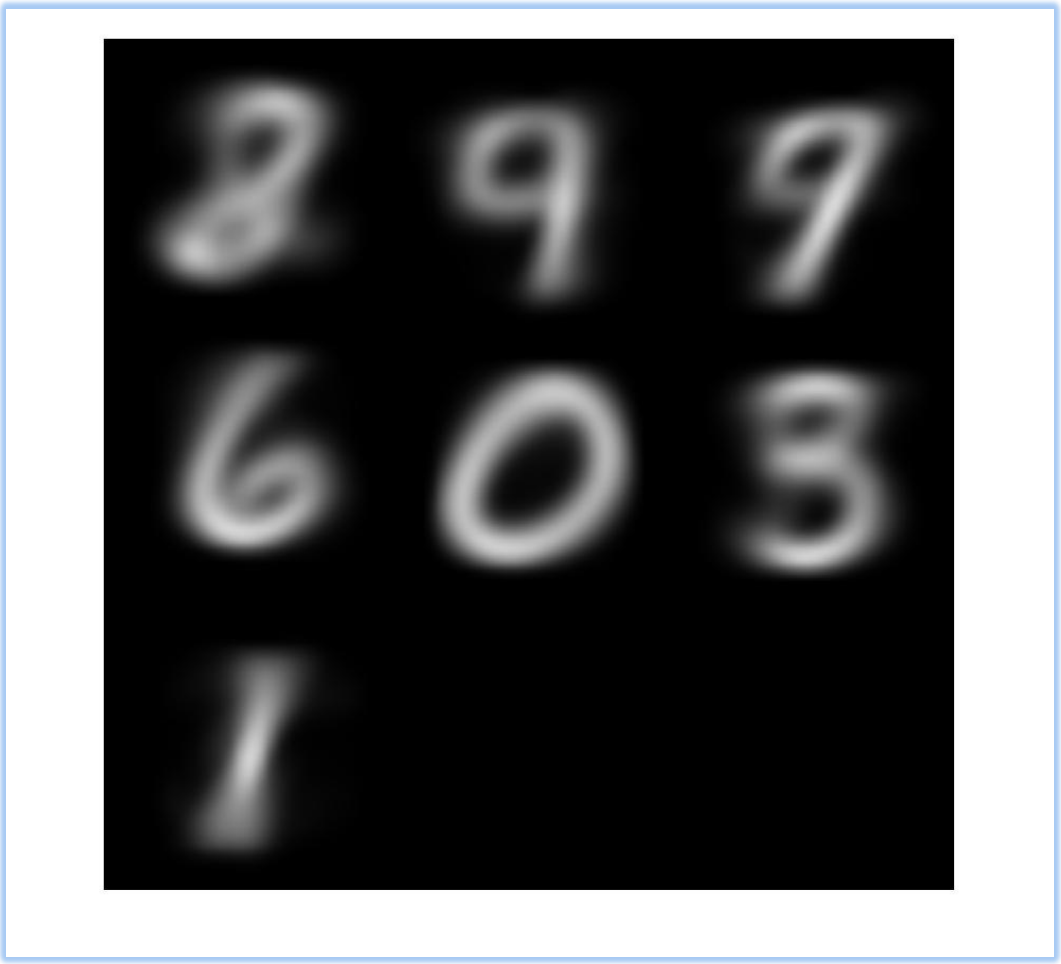


Figure 6: 7 clusters

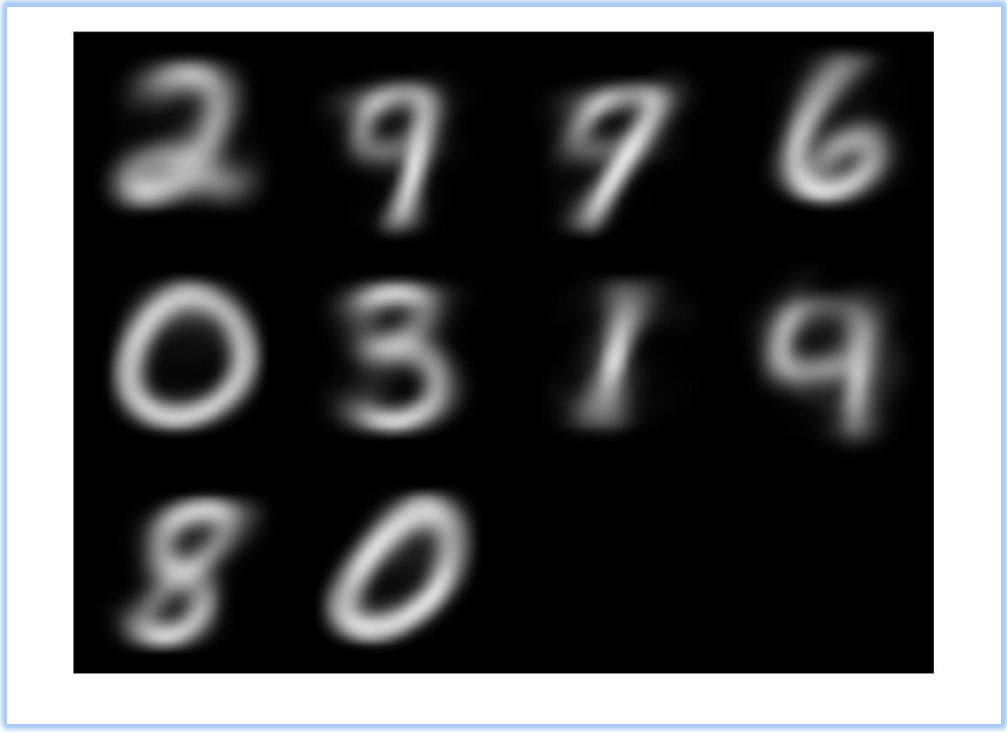


Figure 7: 10 clusters

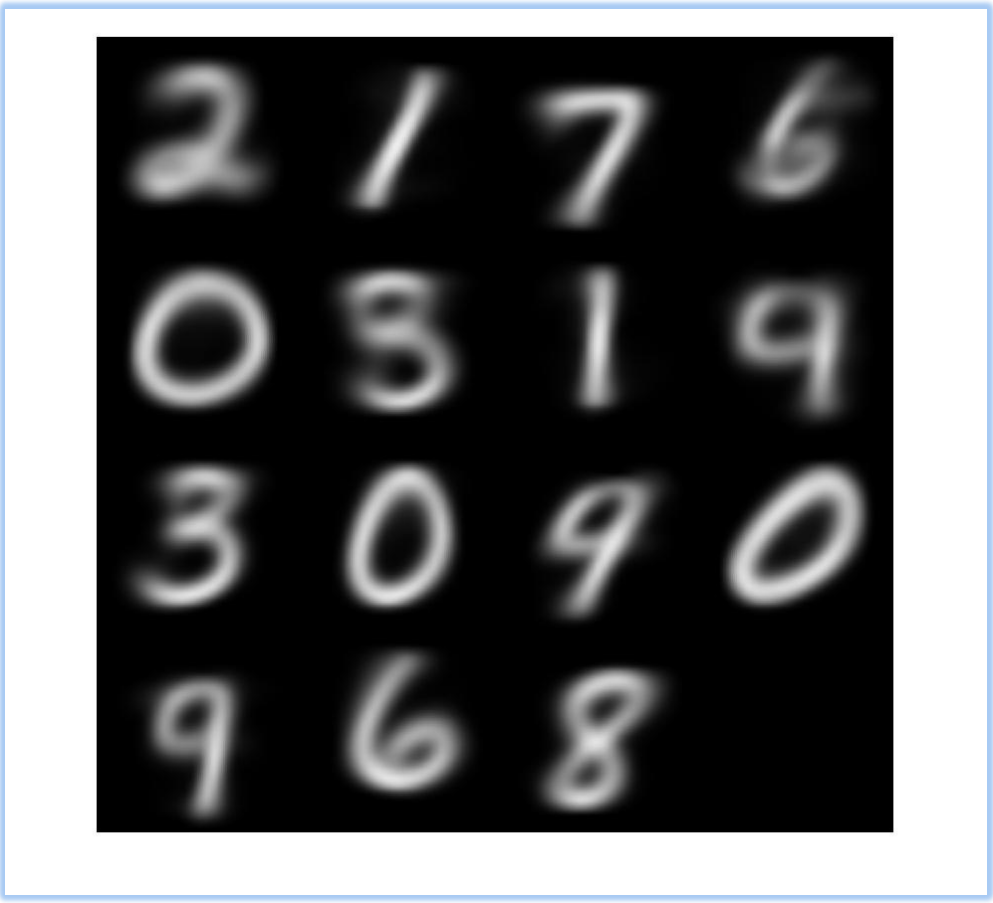


Figure 8: 15 clusters

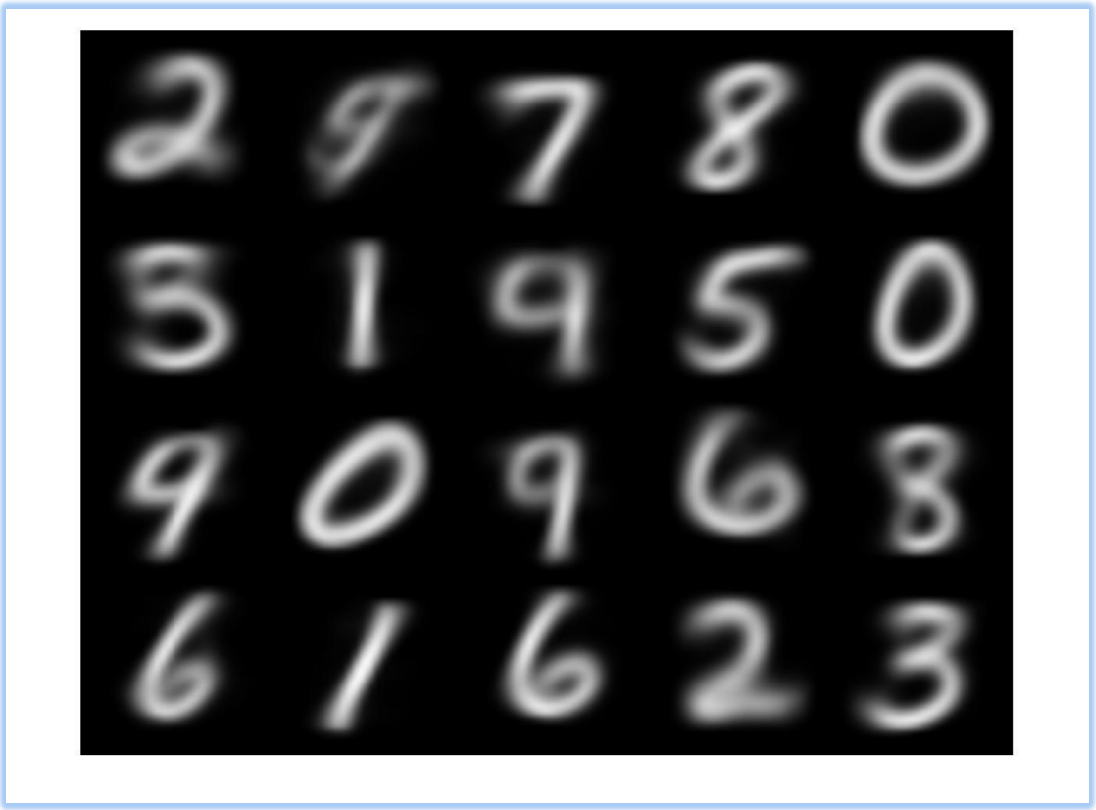


Figure 9: 20 clusters

Task 1.7

First step in this task was to translate the original, 784 dimensional data to just 2 dimensions, using Principal Component Analysis. This was essential to plot the cluster assignments on a 2D graph. Having done that, next thing was to compute the plotting range, which was supposed to be $m \pm 5\sigma$. For the mean, the 2D translation of the mean vector of entire data set was used. For standard deviation, 2 first eigenvalues were used, as they represent variances. Finally, the plane was divided into 200x200 points, creating a grid. Then, distances from all clusters were computed for each point in a grid, so that all of them were assigned a cluster, represented by its index. In the end, each point in the grid was coloured with a colour assigned to its cluster index.

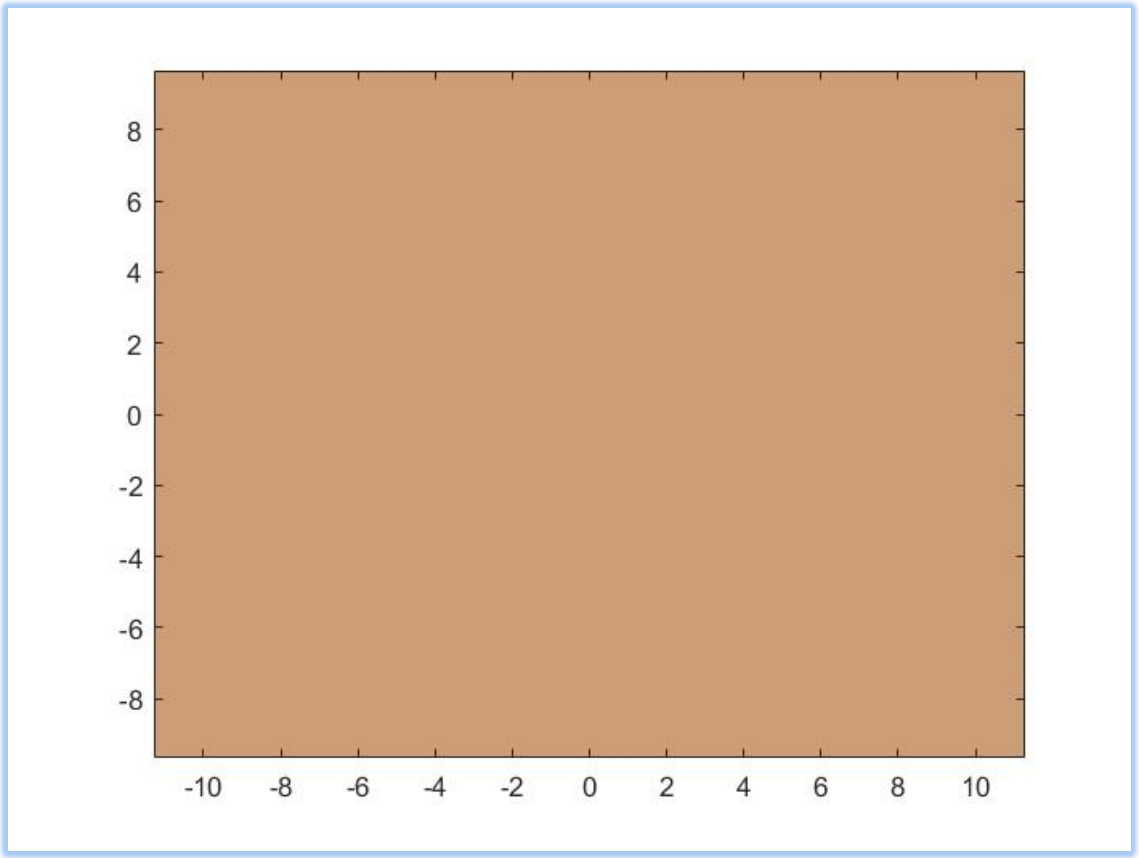


Figure 10: 1 cluster

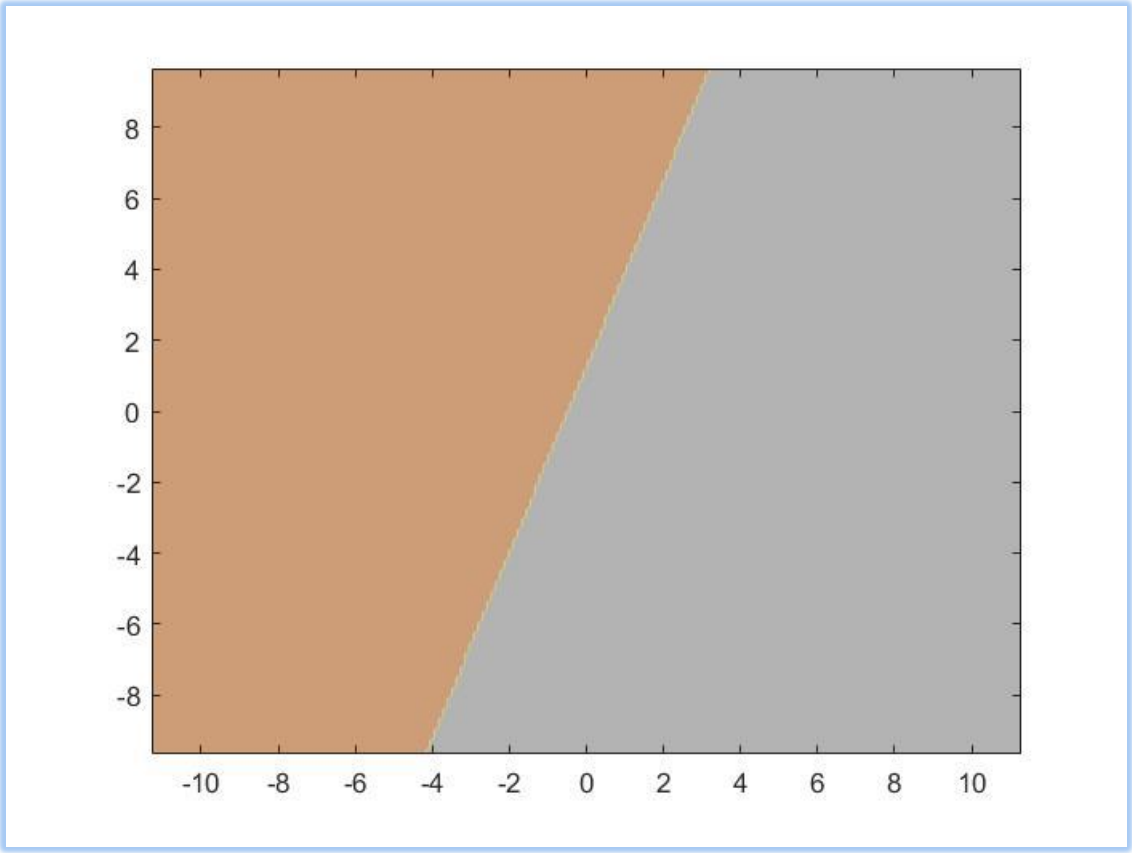


Figure 11: 2 clusters

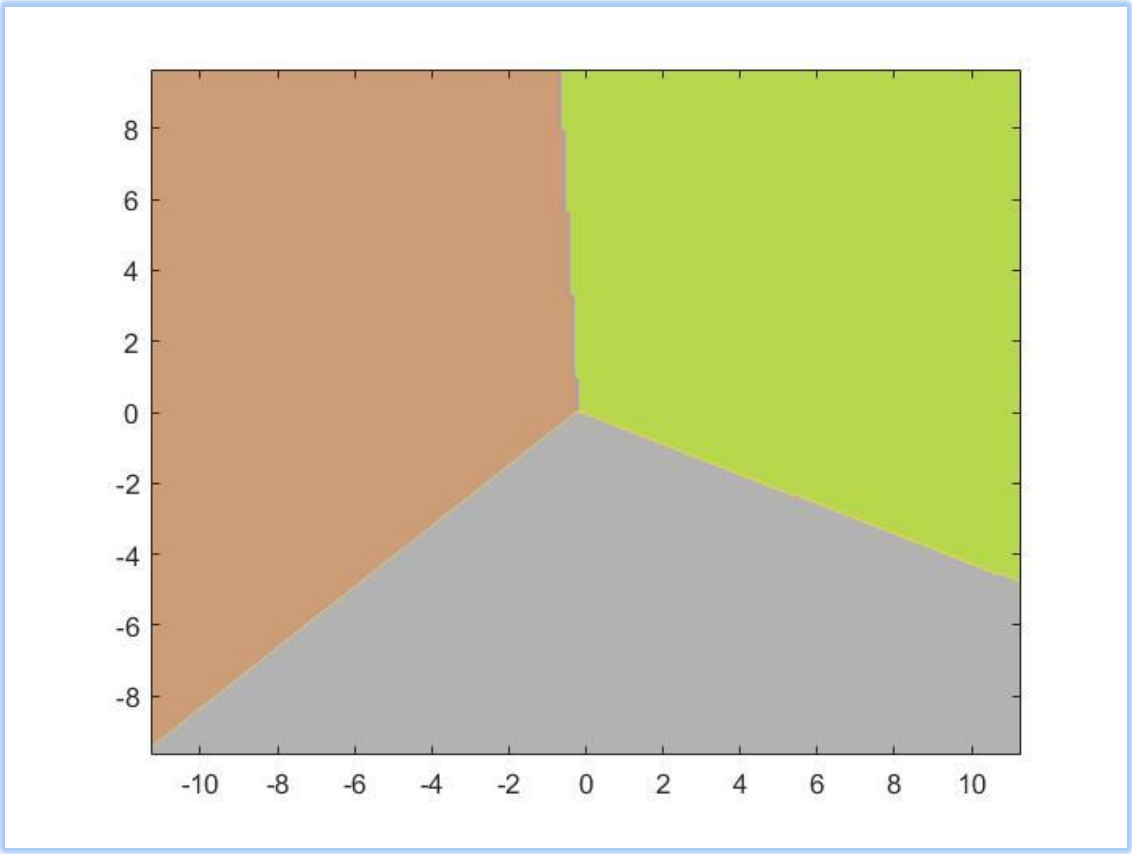


Figure 12: 3 clusters

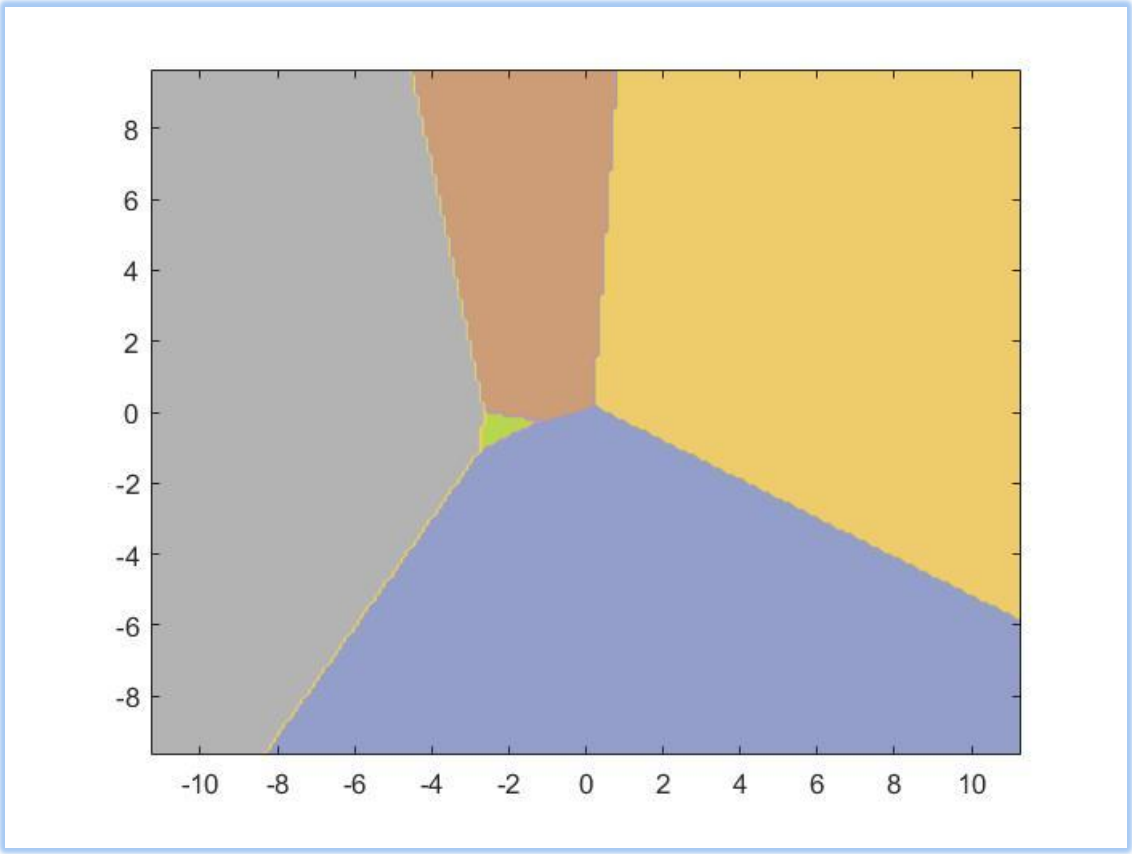


Figure 13: 5 clusters

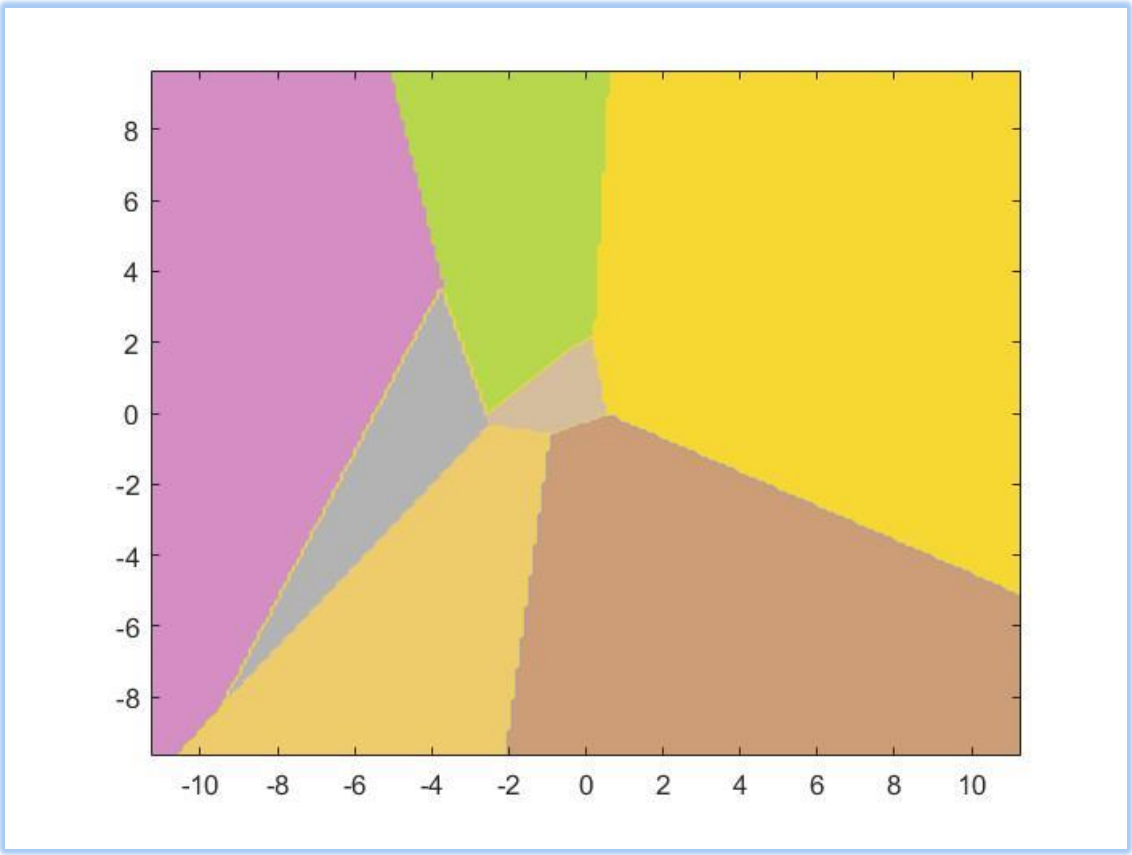


Figure 14: 10 clusters

Task 1.8

To investigate the effect of initial cluster centres on K-means, a total of 100 experiments was ran. For each number of K initial clusters, 10 sets of K initial centres were picked from the Xtrn data set. K-means function was then called on every set and final cluster centres, as well as final SSE for each experiment were recorded and stored in appropriate variables. In a table below, all the calculated SSEs are presented. Row number signifies what was the initial number of clusters, whereas number of column tells which experiment gave the resultant value.

Number of initial clusters	Number of experiment									
	1	2	3	4	5	6	7	8	9	10
1	52.726	52.726	52.726	52.726	52.726	52.726	52.726	52.726	52.726	52.726
2	49.347	49.347	49.347	49.347	49.347	49.347	49.347	49.347	49.347	49.347
3	46.941	46.941	46.941	46.941	46.941	46.941	46.941	46.941	47.438	46.941
4	45.129	45.199	45.199	45.199	45.129	45.199	45.129	45.129	45.129	45.201
5	43.455	43.455	43.455	43.455	43.455	43.455	43.455	43.455	43.956	43.455
6	42.233	42.233	42.233	42.233	42.233	42.233	42.233	42.233	42.233	42.233
7	41.339	41.339	41.339	41.339	41.863	41.339	41.339	41.339	41.339	41.339
8	40.764	40.567	40.567	40.733	40.622	40.567	40.567	40.716	40.716	40.567
9	39.960	39.938	39.937	39.960	39.937	39.858	39.937	39.937	39.937	39.937
10	39.248	39.458	39.249	39.248	39.324	39.249	39.414	39.336	39.251	39.418

Since values in the same row are very close to each other, sometimes all of them being actually the same, it can be seen that different initial clusters do not affect the final value of Sum Squared Error significantly. It is the number of clusters, not the choice of them which has the most impact on SSE. We notice that the average SSE is inversely proportional to the number of initial clusters. The explanation for that is that the more clusters there are, the more accurately data can be assigned to them. Average SSE for different number of clusters is presented in the graph below: