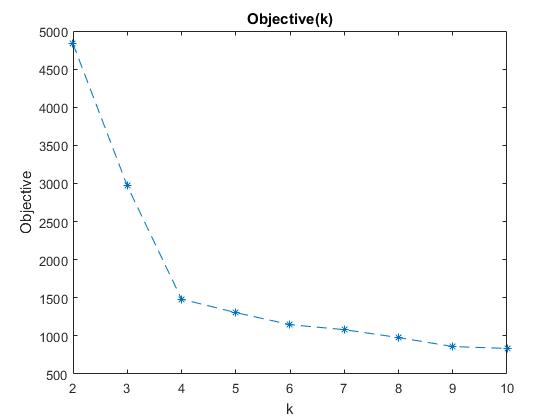
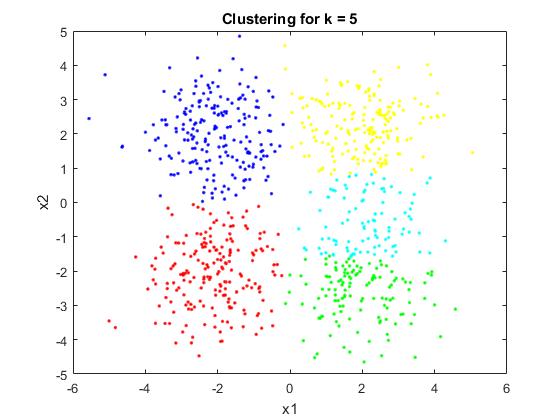
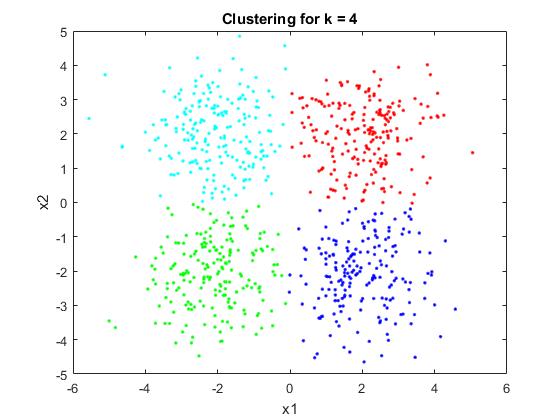
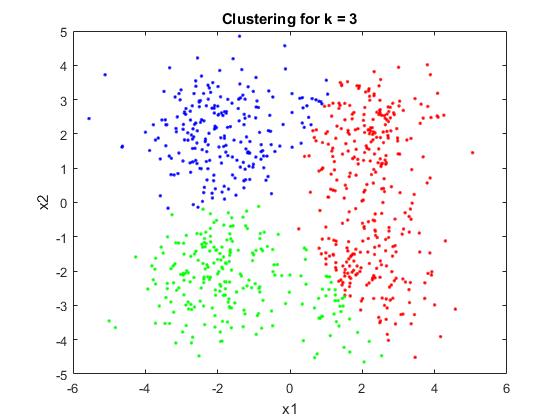
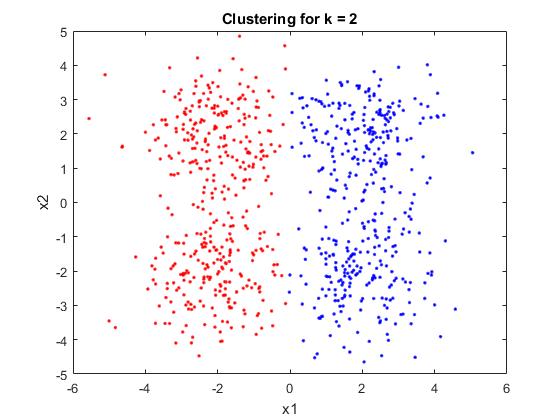
1.b.



The objective decreases in k. This is logical if we look at the objective formula:

As k increases, each example has a center that is closer to it, so the summation of all distances will be smaller in the end.

1. c.



By looking at the dataset, we can observe that the points generally divide into four groups. Therefore, k=4 is the most logical clustering (this can also be clearly seen in the figure for k=4).   
This is not the clustering with the smallest value of the k-means objective.

1.d.

By looking at the formula for the k-means objective, we can see that with high probability, the objective decreases in k regardless of the dataset. However, the most logical clustering is not always the one with the most number of clusters.

1.e.

|  |  |  |
| --- | --- | --- |
| Cluster | Most common label | Percentage of common label in Cluster (%) |
| 1 | 7 | 91.0448 |
| 2 | 1 | 66.0606 |
| 3 | 6 | 71.1712 |
| 4 | 0 | 86 |
| 5 | 4 | 50 |
| 6 | 9 | 40.2878 |
| 7 | 8 | 54.8077 |
| 8 | 3 | 54.7445 |
| 9 | 0 | 91.8367 |
| 10 | 2 | 89.1892 |

1.f.

The classification error on the sample resulting from a classifier derived from the above clusters

Is: 0.357 this calculation was done by the following matlab calculation:

percentages = zeros(1, 10);

modes = zeros(1, 10);

Ypredict = zeros(size(C));

for i = 1:10

labels = Y(C==i);

modes(i) = mode(labels);

percentages(i) = sum(labels==modes(i)) / size(labels,1) \* 100.0;

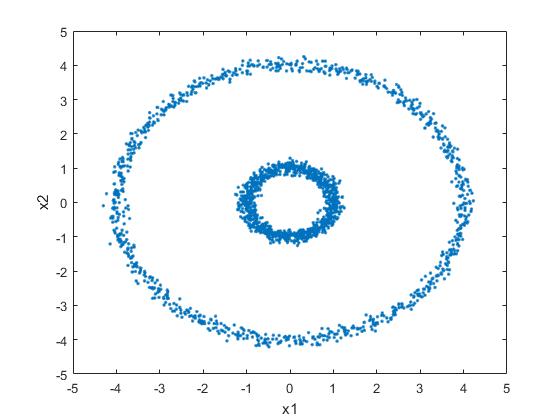
Ypredict = Ypredict + (C==i)\*modes(i);

end

errs(j) = mean(Y ~= Ypredict);

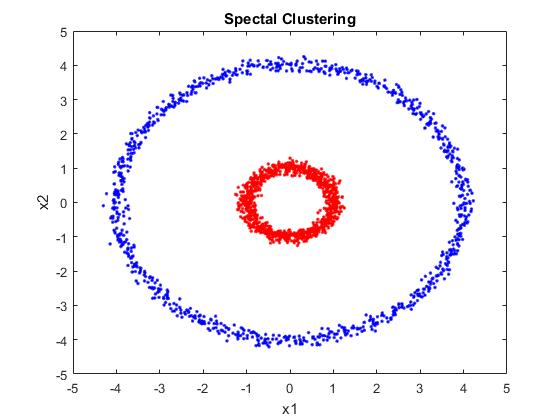
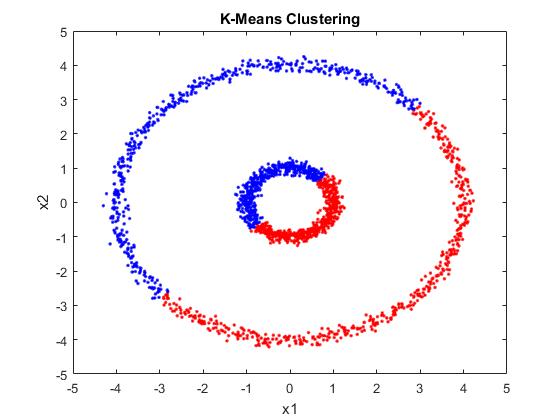
with C being the clusters array returned from the K-means, And Y being the original labels.

2.b.



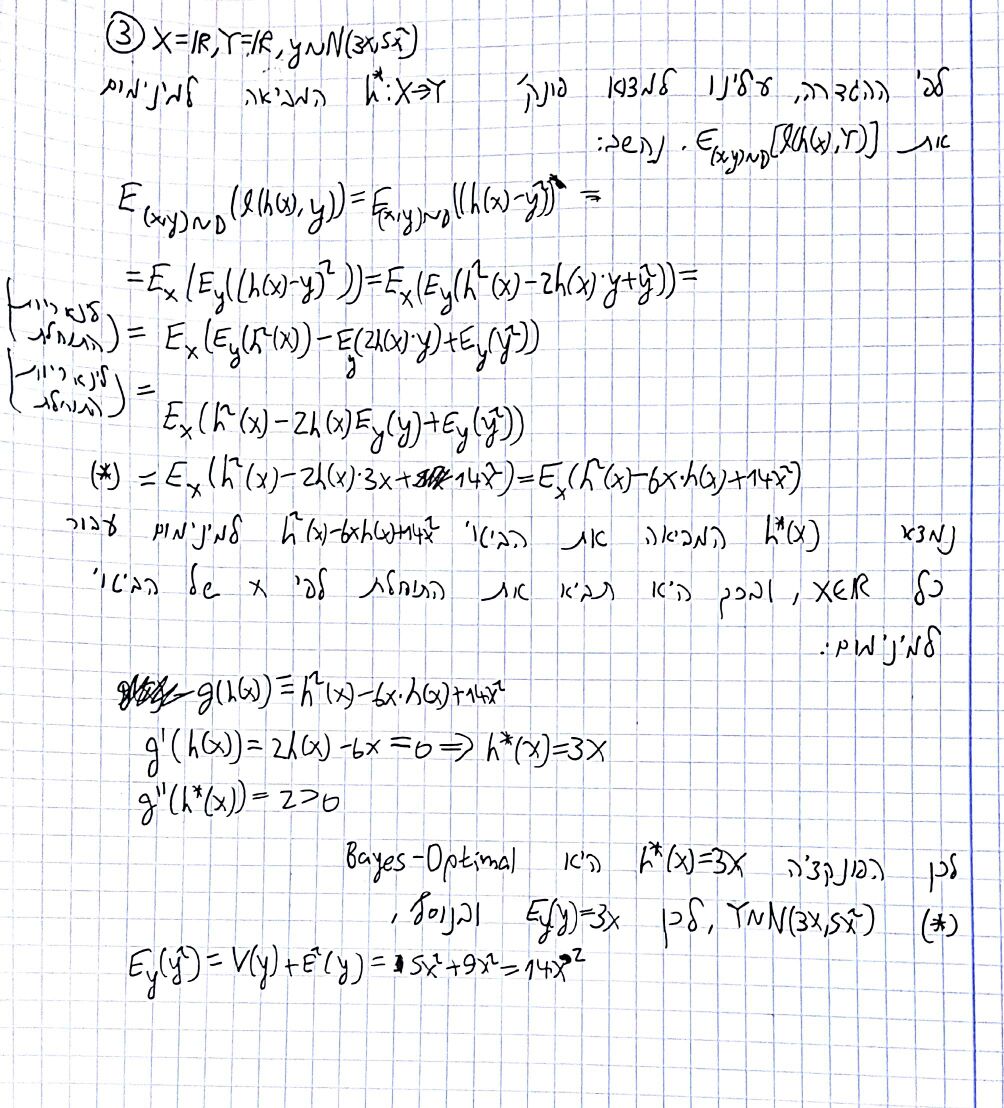
k-means will find the clusters in which the points are closest to their center. In this case this will not output the desired outcome since the outer circle would never be identifies as one cluster.

2. c.

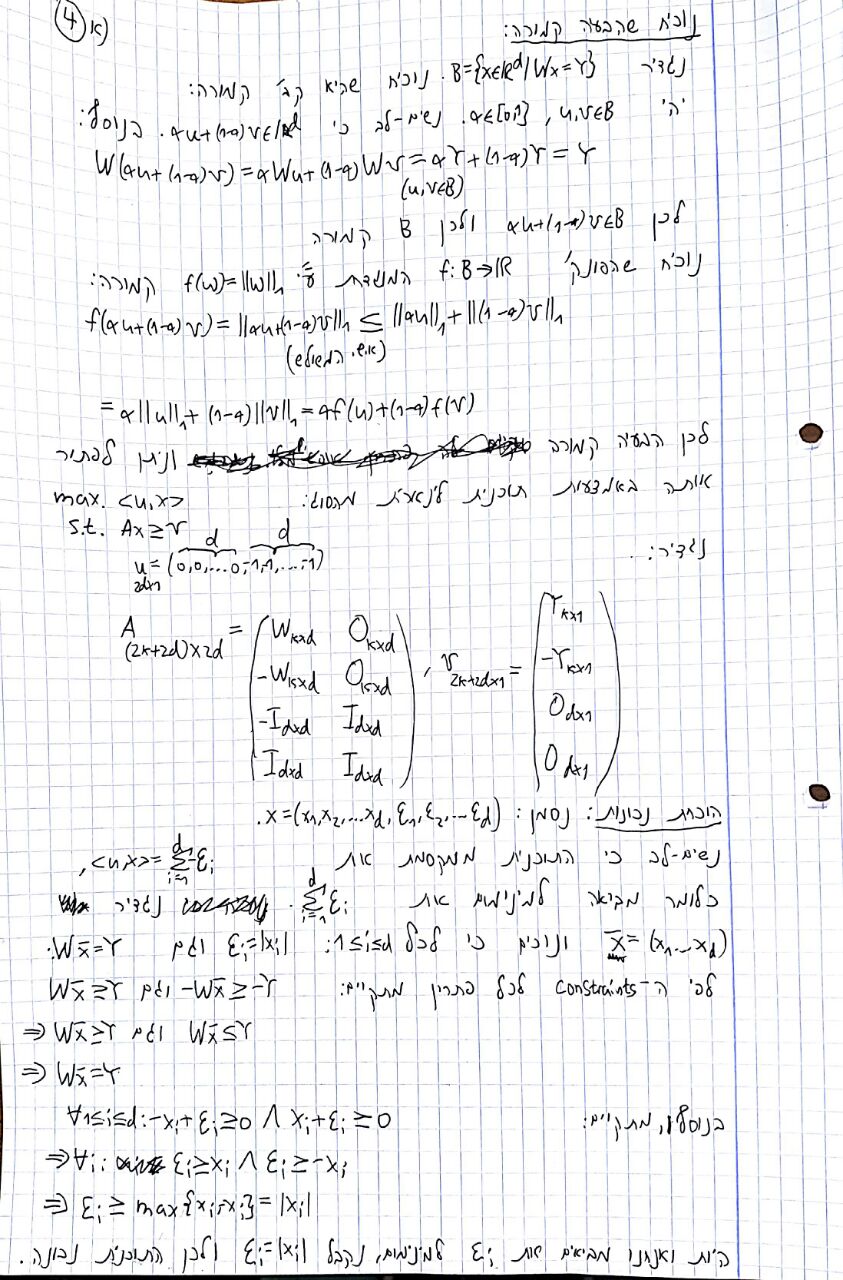


K-means Clustering tries to minimize the distances from each point to its cluster’s center. Therefore, it “splits” the image in 2.  
Spectral Clustering tries to minimize the variance of its clusters. Therefore, it identifies each circle as a cluster.

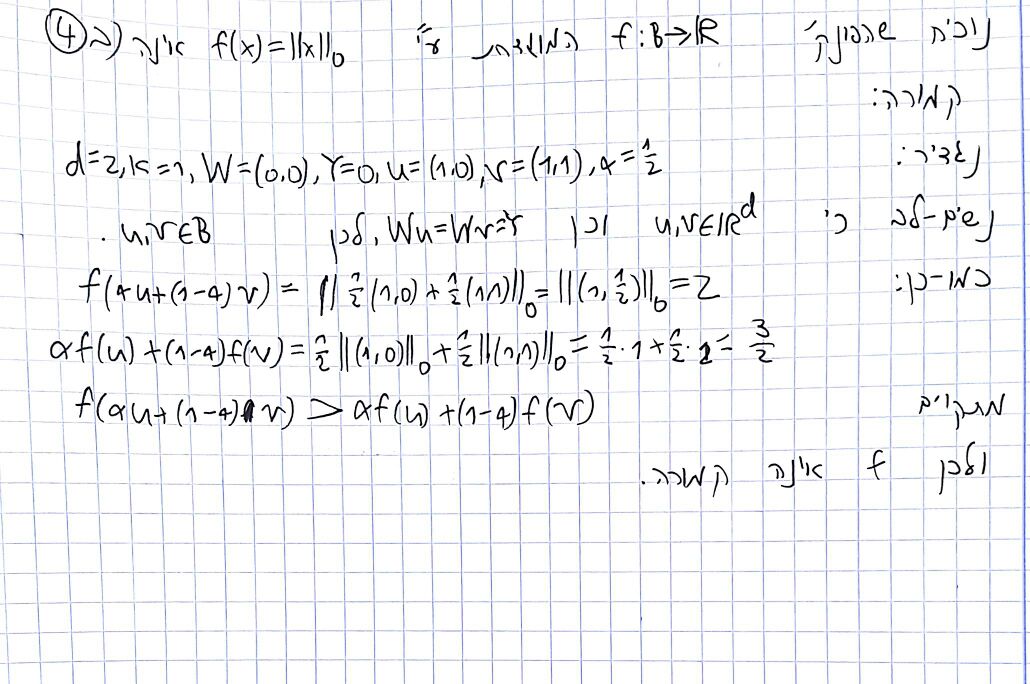
3.



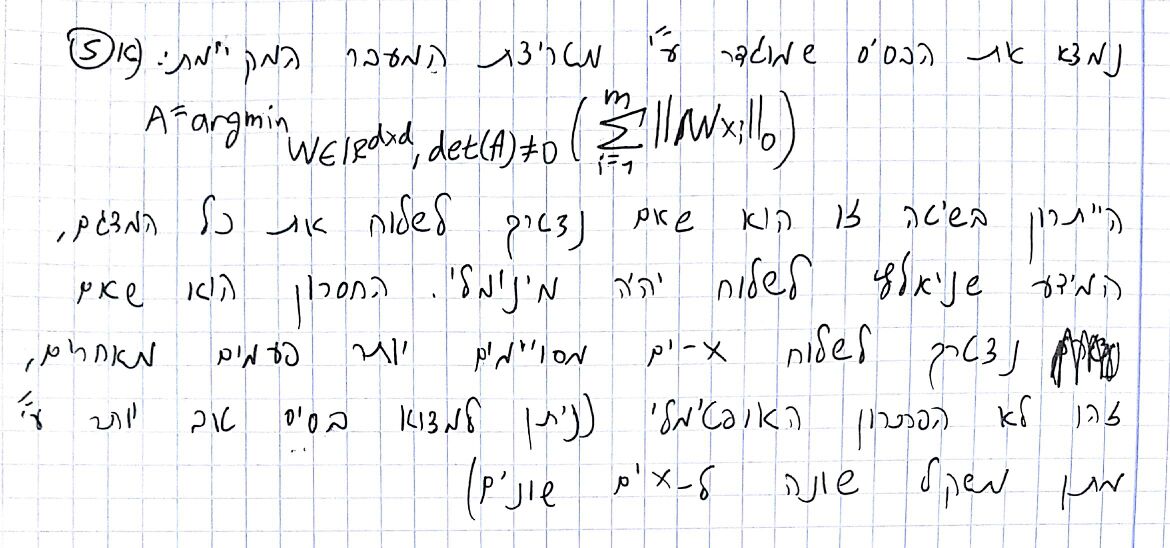
4.a.



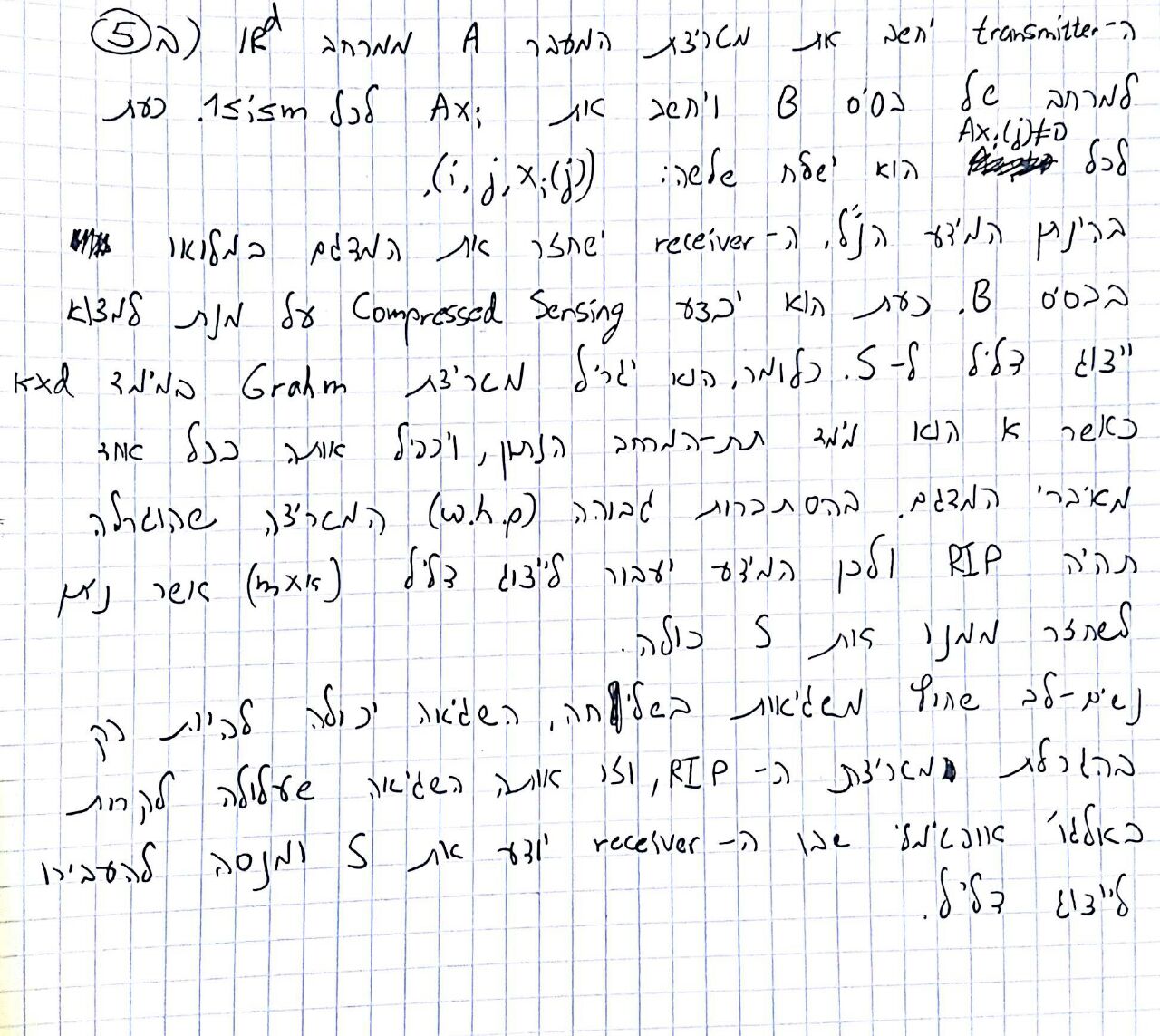
4.b.



5.a.



5.b.



6.a.

