Atanas Delevski
ECE 407 HW #7
4/15/2020

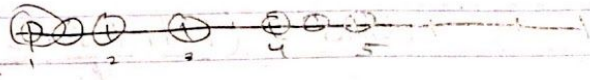
Problem 1: (I did on paper [2 pages]) (black box is covering a mistake)

ATANAS DELEVSKI ECE 407 HW #7
67354175 3
4/15/2020

1) $X = \{1, 2, 1.5, 1, 3, 4, 4.5, 5\}$

↳ cluster to 2 clusters w/ k-means

Solution:



→ By observation, I will initialize my first center at 2 and my second center at 4.5.

[redacted]

Iteration 1	Center 1	Center 2
$x=1 \Rightarrow$	$2-1=1$ ✓	$4.5-1=3.5$
$x=2 \Rightarrow$	$2-2=0$ ✓	$4.5-2=2.5$
$x=1.5 \Rightarrow$	$2-1.5=0.5$ ✓	$4.5-1.5=3$
$x=1 \Rightarrow$	$2-1=1$ ✓	$4.5-1=3.5$
$x=3 \Rightarrow$	$3-2=1$ ✓	$4.5-3=1.5$
$x=4 \Rightarrow$	$4-2=2$	$4.5-4=0.5$ ✓
$x=4.5 \Rightarrow$	$4.5-2=2.5$	$4.5-4.5=0$ ✓
$x=5 \Rightarrow$	$5-2=3$	$5-4.5=0.5$ ✓

⇒ For cluster 1 ⇒ $3+2+1.5+1+1 = 8.5/5 = 1.7$

⇒ For cluster 2 ⇒ $5+4.5+4 = 13.5/3 = 4.5$

Once I got to this point I realized that the point $x=3$ is right in the middle between $x=2$ and $x=4$, and so depending on how you initialize both centers, the point $x=3$ will converge to the closer one and wont change after more iterations. →

For example, if I had picked my first center to be $x=1$, then the point $x=3$ would have gone to the second cluster, and the centers would be:

$$\text{center 1} \left(\frac{1+1+1.5+2}{4}\right) = 1.375$$

$$\text{center 2} \left(\frac{3+4+4.5+5}{4}\right) = 4.125$$

And so the point would have stayed in the second cluster.

Therefore there are two possible solutions to this question.

My solution was:

Cluster 1: 1, 1, 1.5, 2, 3 w/ C=1.7

Cluster 2: 4, 4.5, 5 w/ C = 4.5

Problem 2:

As in the previous homework, I decided to once again use Python for this problem. I used the same libraries as last time; **pandas** for data manipulation and **matplotlib** for data visualization.

At first I tried to program the K-means algorithm by myself but even after looking at multiple resources online, I found it way too complicated to turn into code even though I understood the algorithm.

Therefore, after hours of trying, I decided to just use the **scikit-learn** library in order to use it's k-means function. I wish I could have gone without using it but it turned out to be too hard for this homework assignment.

After using the k-means clustering function from scikit-learn and setting the number of clusters to 4, I got these centers: (It is worth mentioning that K-Means algorithm has a downfall of falling into local minima in terms of optimization. This means that every time I ran the program, I got different centers. This is just one out of many different solutions depending on how the program first initializes the centers.)

	X	Y
Center 1 (red)	4.75	3.15
Center 2 (green)	6.352	2.92
Center 3 (blue)	5.2833	3.70833
Center 4 (yellow)	5.52	2.62

Part B: Given the test vector, (4.9, 6.2), this instance of the algorithm would classify that vector as Class 3 (blue).