Alex Iacob, Benson Yan

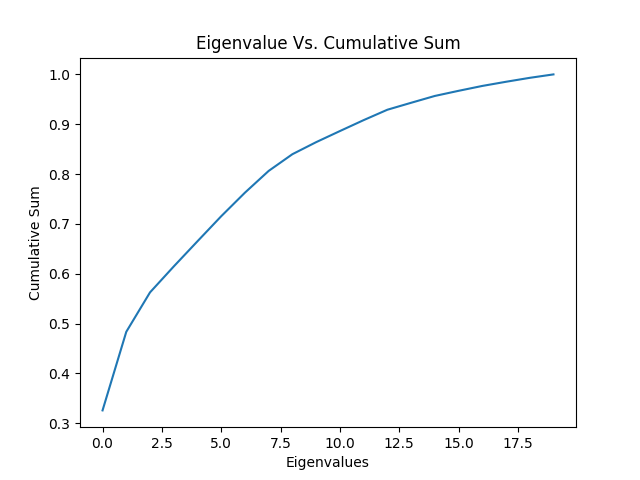
Prof. Kinsman

CSCI 420

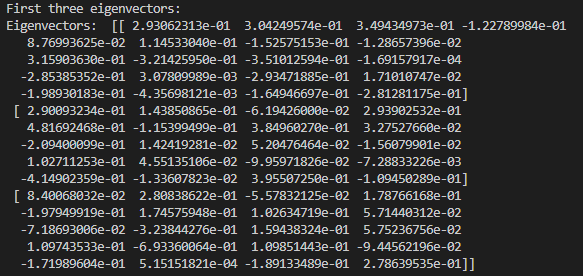
November 15, 2022

1. We use 10 visits because having more than this would likely overflow our system and keeping this amount of visits also allows the program to run efficiently.

***4)***

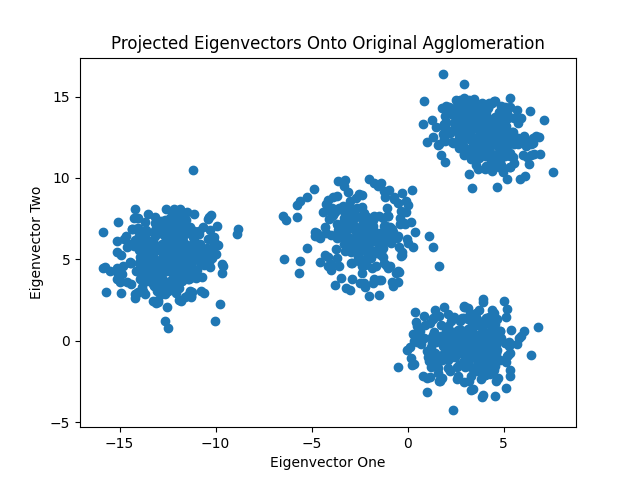


***5)*** The first three eigenvectors are:

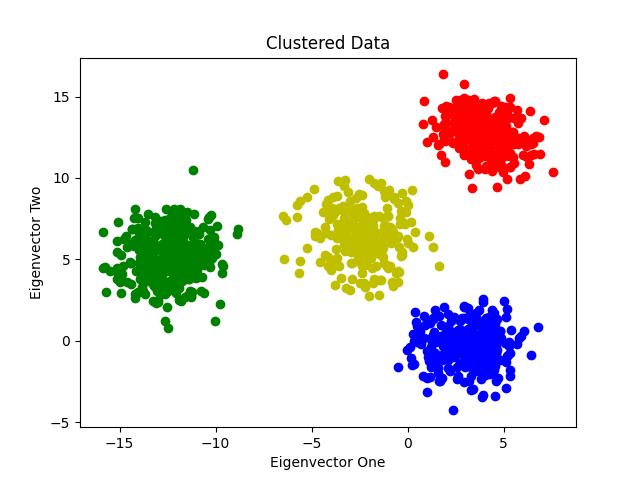


***6)*** Showing all of the original Agglomeration data onto the first two eigenvectors

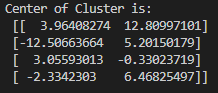
We first had to compute the dot product of the two eigenvectors, however we had to first transpose that data into a *n x m* matrix. Upon displaying the data, we get four visible clusters.



***7)*** For finding the Kmeans we used the sklearn.cluster library. This has a built-in function where it can calculate the center of the clusters. Upon using this function, we were able to get the four unique clusters.



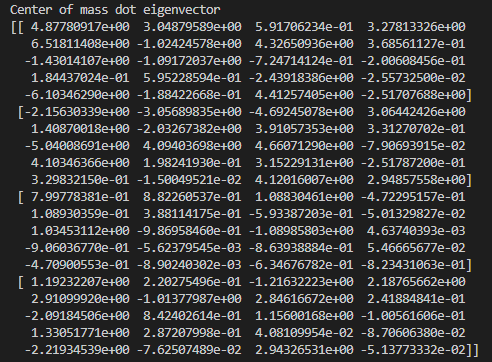
***8)*** In order for us to calculate the center of the mass of each cluster, we used sklearn cluster\_center. Upon using this function, we got four results due to having four clusters. The order for the clusters goes Red, Green, Blue, Yellow.



***9)*** When multiplying the center of the mass with the first two eigenvectors, we needed to make sure that the dimensions were the same. At first we did the same method when calculating the agglomeration data with the first two eigenvectors, but it had dimension problems. To fix that

problem we need to transpose the two eigenvectors when doing dot product with the center of the mass. The only things odd that we noticed about this were that we had some negative values.

When multiplying the center of mass with the first two eigenvectors, we get



***10)*** If you project all of the data onto all of the eigenvectors, it would not help understanding the data because you are not reducing the dimensions. Also the dimension itself will just be a massive number, where it would probably make it harder to understand the data.

***Conclusion***:

This homework felt slightly easier than the previous one, and we were able to understand how to find the covariance, center of the mass, and finding the eigenvectors. NumPy allowed us to pretty easily find the eigenvalues and eigenvectors while also having a built-in function to find dot products. These allowed us to pretty easily create the four clusters shown in the graphs. Also, using sklearn’s KMeans class allowed us to find the different clusters and their center masses.

Most of the setbacks that occurred was when we didn't realize that we had to change the dimensions of the matrices since it didn’t really occur to us at the time (for example when we were calculating the agglomeration data for the first two eigenvectors). Also when we had to multiply the center of mass with two eigenvectors, we had to transpose each in order to make the dimensions match correctly. Overall we got to learn about eigenvectors, eigenvalues, covariance calculations, kmeans, and plot creation. Also we were able to learn about utilizing more useful libraries, like sklearn, which we will probably use in the future.