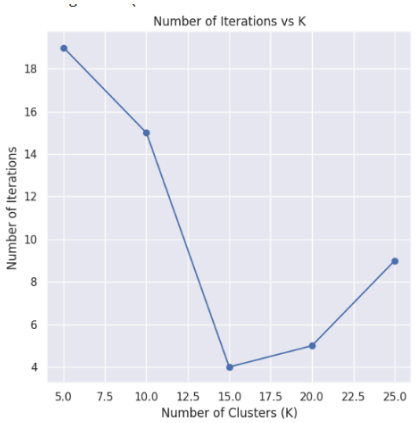
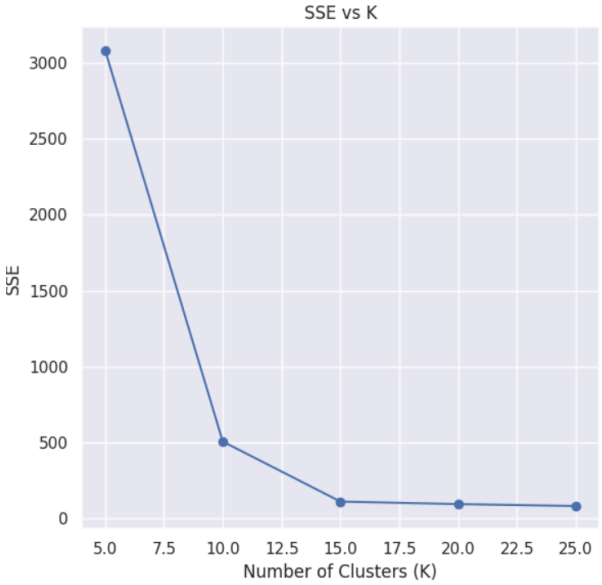
HW8-1: Clustering Download the “SynData1.txt” dataset from the Canvas. a) Study sklearn.cluster (https://scikit-learn.org/stable/modules/classes.html#module-sklearn.cluster) b) Find the optimal number of clusters. c) Using sklearn.cluster, use k-means to cluster the dataset into k = 5,10,15,20, and 25 clusters. For each k: (i) how many iterations until convergence? (ii) what is the within cluster sum of squared error SSE? Is there any correlation between k and SSE? (iii) plot the results.

1b. K = 10
1c.



15 Clusterrs is then it has the least amount of iterations.
Also around 15 SSE adding more clusters doesn’t improve the performance.



HW8-2: Clustering Use the distance matrix in the following table to perform (a) single link, (b) complete link, and (c) Group Average hierarchical clustering. Show your results by drawing a Dendrogram. The Dendrogram should clearly show the order in which the points are merged

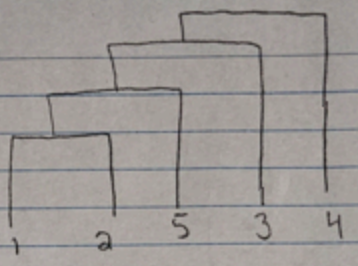
Single link

a lowest first P1-P2

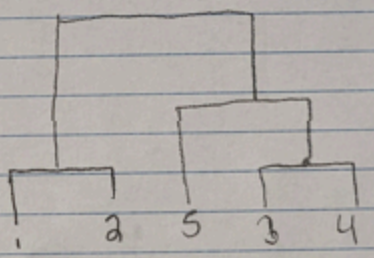
	P1	P2	P3	P4	P5
P1	000				
P2	0.10	0			
P3	0.11	.64	0		
P4	0.55	.47	.44	0	
P5	0.35	.98	.85	0.76	0

	P12	P3	P4	P5
P12	0			
P3	.41	0		
P4	.55	0.44	0	
P5	0 35	.95	.76	0

	P1235	P4
P1235	0	
P4	.44	0



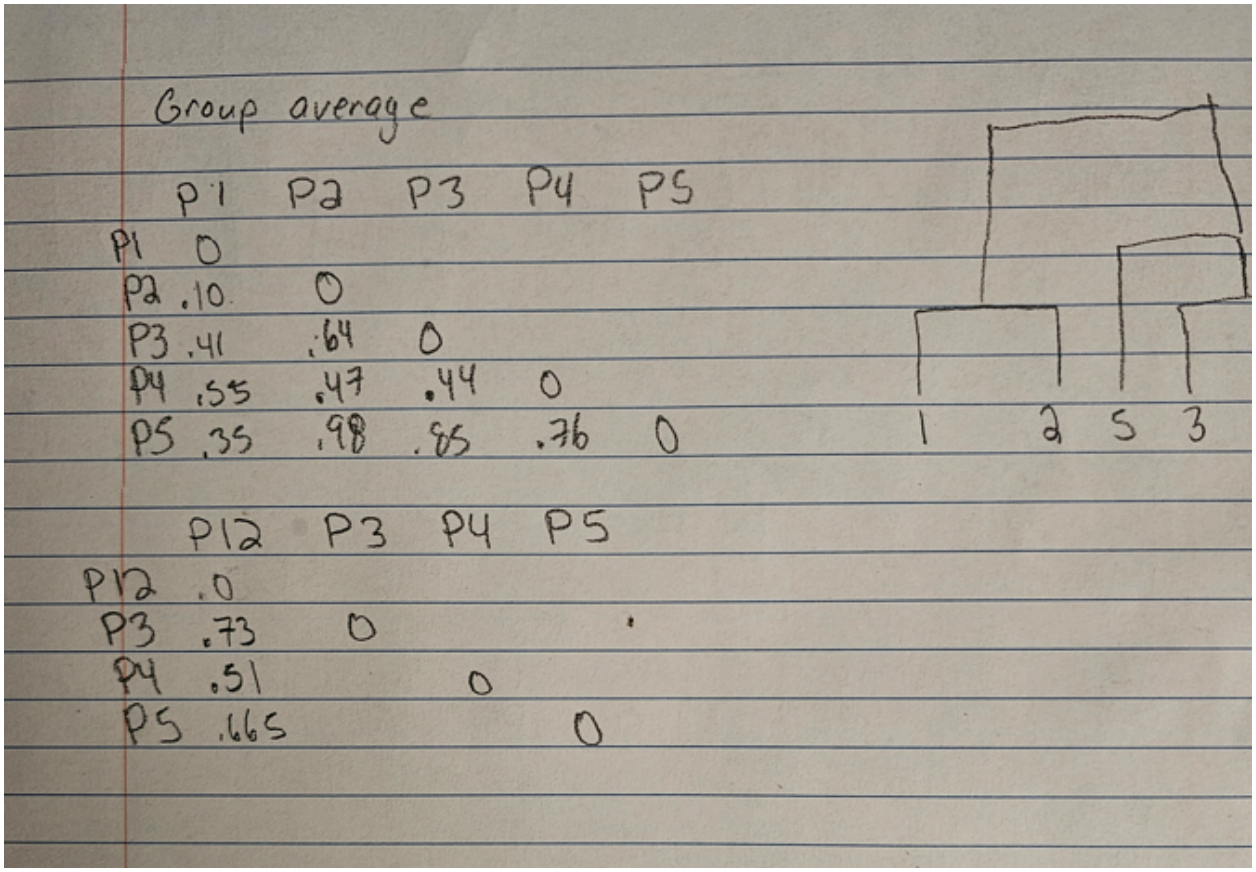
	P1235	P3	P4
P1235	0		
P3	.41	0	
P4	.55	.44	0



Complete Max clostest first P1 -> P2

	P12	P3	P4	P5
P12	0			
P3	.64	0		
P4	.98	.44	0	
P5	.98	.85	.76	0

	P12	P34	P5
P12	0		
P34	.98	0	
P5	.98	.85	0



HW8-3: Clustering Download the “Shape Sets” datasets from http://cs.joensuu.fi/sipu/datasets/ Using sklearn.cluster, run DBScan clusterer on the dataset and find the parameter values that find the optimal number of clusters (if such parameter values exist) - the optimal number of clusters is provided in the site above.

```
File: spiral.txt
Optimal EPS: 4.2
Optimal Min Samples: 10
Number of Clusters: 2

File: pathbased.txt
Optimal EPS: 1.6
Optimal Min Samples: 4
Number of Clusters: 9

File: flame.txt
Optimal EPS: 1.0
Optimal Min Samples: 7
Number of Clusters: 4
```

```
File: Compound.txt
Optimal EPS: 3.0000000000000004
Optimal Min Samples: 2
Number of Clusters: 2

File: jain.txt
Optimal EPS: 2.7
Optimal Min Samples: 2
Number of Clusters: 2

File: D31.txt
Optimal EPS: 1.5000000000000002
Optimal Min Samples: 5
Number of Clusters: 31

File: Aggregation.txt
Optimal EPS: 1.5000000000000002
Optimal Min Samples: 2
Number of Clusters: 7

File: R15.txt
Optimal EPS: 0.7000000000000001
Optimal Min Samples: 2
Number of Clusters: 15
```