Due Date: See Webcampus
How to submit: Webcampus

For programming homework, please submit your ipynb code and also put your results into the pdf file you submit.

HW4-1: Clustering

Download the “SynData1.txt” dataset from the Webcampus.
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.

HW4-2: 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.

	P1	P2	P3	P4	P5
P1	0.00	0.10	0.41	0.55	0.35
P2	0.10	0.00	0.64	0.47	0.98
P3	0.41	0.64	0.00	0.44	0.85
P4	0.55	0.47	0.44	0.00	0.76
P5	0.35	0.98	0.85	0.76	0.00

HW4-3: 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.