

Jan Nyberg 665380
Assignment 1
26.9.2020

Using jupyter notebook to solve the exercises

1.A

Done in the notebook.

Best K for:

Silhouette Coefficient, K=2

Calinski Harabasz, K=2

Davies-Bouldin, K=2

.B

Using the sum of squared distances to determine an optimal value for k. The optimal value for K=3 (taken from the graph in jupyter notebook). If we choose from 2,5 and 10 then 2 is the best choice.

2.A

Done in the notebook.

.B

By watching the dendograms in the notebook we can conclude that the complete and average methods are best at creating evenly distributed clusters. The single method was by far the worst.

3

Plotted graphs of the shuffled data in the jupyter notebook and couldn't find any difference in the dendograms.

4.A

Done in the notebook. Used heatmap to graph similarities.

.B

Wrote the goodall algorithm in the notebook. Graphed it with a heatmap.

.C

Done in the notebook.

.D

Done in the notebook.

5.A

Calculated in jupyter notebook by using numpy library.

Eigenvalues: [-5.55111512e-17 5.00000000e-01 0.00000000e+00 0.00000000e+00]