Data Analytics Foundations

Online Workshop 7
Clustering (1)

Questions from Module 5?

Hierarchical clustering in KNIME

- Start up KNIME and import the iris dataset
- Set up a Hierarchical Clustering node with 3 clusters
  - Why does the configuration not allow you to include Species?
- Add a Color Manager and set up the colours to reflect the clusters coming from the hierarchical clustering
- Add a Shape Manager and set a different shape for each species
- Add a Scatter Plot to visualise the results

Hierarchical clustering in KNIME

- How well does the hierarchical clustering recover the known classes?

- Play around with the linkage type and see how it affects the clusters found

- Have a look at the dendrogram (View: Dendrogram/Distance View). Does it give you any clues about what a 'good' number of clusters might be? What about the distance plot?

Choosing the number of clusters

- Good clustering gives clusters **with high intra-cluster similarity** and **low inter-cluster similarity**

- Silhouette method
  - Maximise the 'silhouette' coefficient: normalised difference between intra-cluster and inter-cluster distances
  - Try out the Silhouette Coefficient node in KNIME