

Data Analytics Foundations

Online Workshop 8
Clustering (2)

Questions from Module 5?

k-means clustering in KNIME

- Start up KNIME and import the iris dataset
- Set up a **k-means** node with 3 clusters
 - What are the centroids of the clusters?
- Add a **Color Manager** and set up the colours to reflect the clusters coming from the clustering
- Add a **Shape Manager** and set a different shape for each species
- Add a **Scatter Plot** to visualise the results
 - How do the clusters change when you increase the number of clusters? How would you choose the best number of clusters?

k-medoid clustering in KNIME

- Do the same thing for k-medoids.
 - k-medoids requires a distance matrix to be supplied – use [Distance Matrix Calculator](#)
 - What are the medoids of the clusters?
- Now repeat the clustering with [iris-outlier.csv](#).
 - How do k-means and k-medoids compare for both datasets?

Wine dataset

- Download [wine.xls](#) from Canvas
- Use this dataset to practice doing some data exploration using the methods you'll be using in the assignment