

Laboratory 3 : Visualisation with KNIME

Install the KNIME software with all the extension on your computer.

Exercise 1:

Open the "PCA" module on Knime and study all the options offered;

Exercise 2:

Load the Iris base and view the initial data. Visualise the data using the variables pairs by pairs. What do you see? What will be the best combination?

Exercise 3:

- Create a scenario by using the PCA method that will transform the correlated variables of a database and visualize the result obtained;
- Test the method on the Iris database. What do you see in relation to exercise 2?
- Save the results obtained in .png;
- Change the model parameters (the number of PCA eigenvectors) for each base and test it each time;

Exercise 4:

- Load the waveform dataset and scatter the data. Analyze the result. Test several combinations of variables. What do you see?
- Apply the scenario of Exercise 3 (PCA) on this data and compare the result of the PCA with the visualisation of the initial base.

Note: The waveform base is made up of 5000 individuals divided into 3 classes. The original base included 21 variables, but 19 additional variables distributed according to a normal distribution were added in the form of noise. Each object was generated as a combination of 2 out of 3 waves.

To analyse the impact of relevant variables on the PCA result, divide the waveform base into two parts: the base made up of relevant variables and the database with the set of irrelevant (noisy) variables.

Test the PCA module on both bases and visualise the result. What do you see?

Exercise 5:

- Use the knowledge of classes that is available for each base and visualise the PCA result with different colours for each base. Test this scenario on the waveform base.
- Create a clustering scenario (with k-means) that will allow you to classify a database and visualise the result obtained. Use this scenario for the basic waveform.

vs. Apply PCA on the same basis and visualise the result using the classes obtained with k-means.

d. Compare the result of PCA on the base with the real classes and the result of PCA on the same base but with the classes obtained by the k-means method. Compare the results.

e. Determine the centers of the resulting clusters. Project the barycenters of each cluster onto the axes of the PCA. Test this scenario on an iris and waveform basis.

Warning: For the scenario to function correctly, configure all the nodes (modules) of the model.