

# Exploration visuelle des données

## TP 2

### Exercice 1 :

Ouvrez le module « PCA » sur Knime et étudiez toutes les options proposées;

### Exercice 2 :

Chargez la base Iris et visualisez les données initiales. Visualisez la base en utilisant les variables par paires. Que constatez-vous ? Quelle sera la meilleure combinaison ?

### Exercice 3 :

a. Créez un scénario par rapport à la méthode ACP et visualiser le résultat obtenu;

b. Testez la méthode sur la base *Iris*. Que constatez-vous, par rapport à l'exercice 2 ?

c. Testez cette méthode sur les bases de données proposées (*Churn*, *pays*) et analysez les résultats obtenus ;

d. Sauvegardez les résultats obtenus en .png ;

e. Changez les paramètres du modèle pour chaque jeu de données et testez le à chaque fois ;

### Exercice 4 :

a. Chargez le jeu de données 'vagues de Breiman' (Waveform) et visualisez ce jeu de données. Analysez le résultat. Testez plusieurs combinaisons de variables. Que constatez-vous ?

b. Appliquez le scénario de l'exercice 3 (l'ACP) sur cette base et comparez le résultat de l'ACP avec la visualisation de la base initiale.

**Note :** La base *waveform* est composée de 5000 individus divisés en 3 classes. La base originale comportait 21 variables et par la suite 19 variables additionnelles distribuées selon une loi normale ont été rajoutées sous forme de bruit. Chaque individu a été généré comme une combinaison de 2 sur 3 vagues.

Pour analyser l'impact des variables pertinentes sur le résultat de l'ACP, divisez la base *waveform* en deux parties : la base composée de variables pertinentes et la base de données avec l'ensemble de variables non pertinentes (bruités).

Testez le module ACP sur les deux ensembles et visualisez le résultat. Que constatez-vous ?

### **Exercice 5 :**

**a.** Utilisez la connaissance des classes (labels) qui est disponible pour chaque base et visualisez le résultat de l'ACP avec des couleurs différentes pour chaque classe. Testez ce scénario pour la base *waveform*.

**b.** Créez un scénario de clustering (avec *k-means*) et visualiser le résultat obtenu. Utilisez ce scénario pour la base *waveform*.

**c.** Appliquez l'ACP sur la même base et visualisez le résultat en utilisant les classes obtenues avec *k-means*.

**d.** Comparez le résultat de l'ACP sur la base initiale avec les classes réelles et le résultat de l'ACP sur la même base, mais avec les classes obtenues grâce à la méthode *k-means*. Comparez les résultats.

**e.** Déterminez les centres des clusters résultants. Projetez les barycentres de chaque cluster sur les axes de l'ACP. Testez ce scénario sur les jeux de données (*waveform*, *pays*, *Churn*).

**Attention :** Pour un bon fonctionnement du scénario configurez bien tous les nœuds (modules) du modèle.