

BI Data Mining - TP2 - Visualisation

Ludovic Denoyer et Laure Soulier

1 Iris

Nous allons travailler sur le jeu de données Iris disponible sous RapidMiner.

- Réaliser une ACP sur les données sans estimer la matrice de variance co-variance. Comparer les résultats avec l'étape de variance co-variance.
- Identifier l'ébouillissement des valeurs propres pour prendre une décision sur le nombre de composantes principales
- Visualiser la projection des individus pour identifier des classes d'individus. Utiliser les fonctionnalités "Advanced Charts" pour mettre en forme les données.
- Pour connaître la contribution des variables dans la classification des individus, appliquer après l'ACP la boîte "Weight by Component Model". Interpréter. Tester également en appliquant la boîte "Weight by PCA" directement sur les données. Que remarquez-vous ?

2 La suite....

- Sur le jeu de données Sonnar disponible sous RapidMiner, réaliser une ACP sur ces données, identifier les classes et les poids des variables dans la classification.
- Faire de même avec le dataset "weighting". Jouer avec les dimensions pour identifier quels axes permettent de visualiser la classification (au delà des croisements 1-2).
- Faire une ACP sur le datasets "voitures". Attention, un traitement des données est nécessaire avant de réaliser l'ACP. Interpréter.