A2DI - PCA

Antonin Carette - Quentin Baert

15 décembre 2015

Question 1

Les visualisation des différentes réductions sont représentées par les Figure 1, 2 et 3. Les éléments de la première classe sont représentés en vert, ceux de la seconde classe sont représentés en rouge et ceux de la troisième classe en jaune.

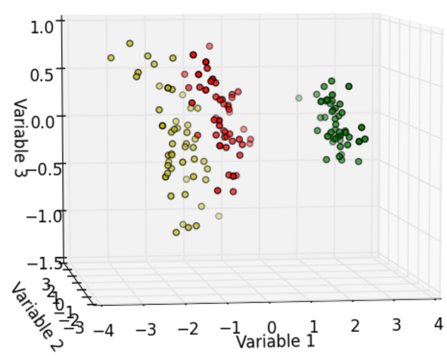

Figure 1: Réduction en 3 dimensions

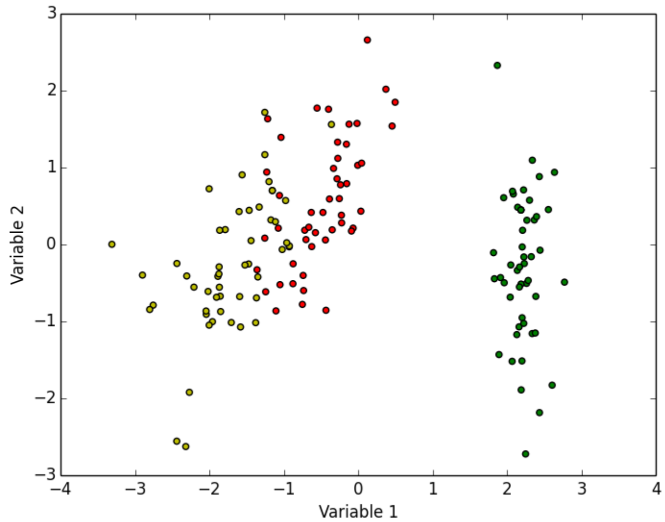

Figure 2: Réduction en 2 dimensions

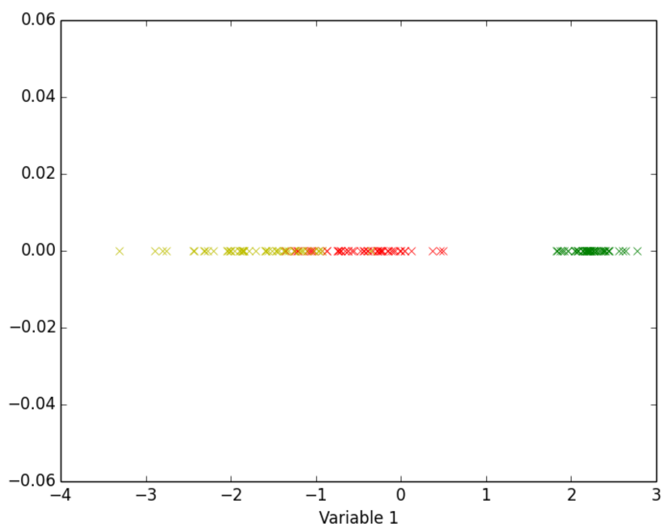

Figure 3: Réduction en 1 dimension

Question 2

Les éléments de la première classe semble être bien différents des éléments des deux autres classes car quelque soit la représentation, ces derniers sont facilement identifiables.

- La réduction en 3 dimensions permet de distinguer les 3 classes facilement.
- La réduction en 2 dimensions permet de distinguer la première classe. Les éléments de la seconde et troisième classes sont peu identifiables car très proches les uns des autres.
- La réduction en 1 dimension permet de distinguer la première classe. Comme pour la réduction en 2 dimensions, les éléments de la seconde et troisième classe sont beaucoup moins différenciés.

Question 3

Pour une matrice X^* initiale de taille $n \times p$ réduite en k dimensions et de matrice de covarience R, on calcul le pourcentage de variance totale de la manière suivante :

$$\frac{\sum_{i}^{k} \lambda_i}{p} \text{ où } \lambda_i \text{ est la valeur propre du ième vecteur propre de } R.$$

Voici les différents pourcentages de variance totale pour chaque projection.

- Réduction en 3 dimensions : $99,48$ %
- Réduction en 2 dimensions : $95,80$ %
- Réduction en 1 dimension : $72,77$ %

Question 4

On note Y les données réduite en k dimensions. Pour reconstruire une approximation X' des données initiales on calcul $X' = Y.U_k^T$ où U_k est la matrice composée des k premiers vecteurs propres en colonnes.

Question 5

Le taux d'erreur de reconstruction est obtenu en calculant la moyenne des différences entre la matrice initiale X^* et la matrice reconstruite X'.

Les taux d'erreur de reconstruction sont les suivants :

- Reconstruction depuis 3 dimensions : $0,05$
- Reconstruction depuis 2 dimensions : $0,15$

- Reconstruction depuis 1 dimension : $0,34$

Question 6

La reconstruction la plus efficace est celle obtenue à partir des données en 3 dimensions. Cela parait logique car c'est la réduction qui réduit le moins les données, il est donc plus facil d'en tirer les données initiales.