CHAPITRE 4

Réponses aux exercices du Chapitre 4

MTH3302 - Méthodes probabilistes et statistiques pour l'I.A.

Jonathan Jalbert – Automne 2021

1. a) Les valeurs propres sont $\lambda_1 = 6$ et $\lambda_2 = 1$. Les vecteurs propres correspondants sont $\boldsymbol{x}_1 = [2\ 1]^\top$ et $\boldsymbol{x}_2 = [1\ \ -2]^\top$.
   b) Le premier propre correspondant à la plus grande valeur propre, donc $\boldsymbol{x}_1 = [2\ 1]^\top$.
   c) $\frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{6}{7}$.

2. Lorsque la variance échantillonnale des valeurs propres est grande. Dans ce cas, plusieurs valeurs propres sont très différentes. Donc une d'entre elles est probablement beaucoup plus petite qu'une autre.

3. Si les variables ne sont pas standardisées, alors il se peut que leurs échelles ne soit pas comparables. Les premières composantes principales pourraient alors provenir de la variable avec l'échelle la plus grande car elle correspond à la plus grande part de la variabilité absolue. L'analyse en composantes principales sert plutôt à extraire les combinaisons linéaires non triviales des variables qui expliquent le plus de variabilité relative. On est plutôt intéressé à la relation entre les variables qu'à leurs échelles.

4. Si une valeur singulière est très grande par rapport aux autres, cela signifie qu'une combinaison linéaire de variables permet de compresser le jeu de données en limitant

la perte d'information. En effet, les composantes principales correspondants aux petites valeurs singulières peuvent être négligées sans trop d'effet.

De plus, une grande disparité entre les valeurs singulières suggère un problème de multicolinéarité. En effet, le déterminant de la matrice, qui est égal aux produit des valeurs propres, sera petit. On divisera donc par un petit nombre la matrice des cofacteurs pour calculer la matrice inverse, ce qui engendre des problèmes d'instabilité numérique.

5. a) Pour trouver les valeurs singulières de A, les valeurs propres de la matrice $AA^\top$ peuvent d'abord être calculées. On obtient le système d'équation suivant :

$$\begin{vmatrix} 2-\lambda & 2 \\ 2 & 5-\lambda \end{vmatrix} = 0$$

qui donne les valeurs propres $\lambda_1 = 6$ et $\lambda_2 = 1$. Les valeurs singulières de A sont donc $\gamma_1 = \sqrt{6}$ et $\gamma_2 = 1$.

b) La matrice V s'obtient avec les valeurs propres de $A^\top A$:

$$V = \frac{1}{\sqrt{30}} \begin{bmatrix} -5 & 0 & -\sqrt{5} \\ 2 & -\sqrt{6} & 2\sqrt{5} \\ 1 & 2\sqrt{6} & \sqrt{5} \end{bmatrix};$$

et la matrice U par les vecteurs propres de $AA^\top$:

$$U = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}.$$

La meilleure approximation de rang 1 de la matrice A est donnée par

$$\begin{aligned} A_1 &= \gamma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^\top \\ &= \sqrt{6} \times \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \times \frac{1}{\sqrt{30}} \begin{bmatrix} -5 & 2 & 1 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} -5 & 2 & 1 \\ -10 & 4 & 2 \end{bmatrix}. \end{aligned}$$

6. a) Avec une composante principale, 0.68882 de la variance est expliquée. Avec deux composantes principales, 0.99976 de la variance est expliquée. Toute la variance est expliquée en utilisant les trois composantes principales.

b) On obtient que $\hat{\eta} = \begin{bmatrix} 0.575 & -0.323 & 5.363 \end{bmatrix}^\top$.

c) On obtient que $\hat{\beta} = \begin{bmatrix} 4.264 & -2.929 & -1.561 \end{bmatrix}^\top$.

d) On obtient 31.3. Attention à la normalisation des variables !