

TD2 Analyse en composante principale

PCA : pas à pas

Ex1 :

Partie 1 : Calcul manuel

On donne le jeu de donnée A suivant :

X	Y
3	21
3	32
4	43
6	64

1. Calculer la moyenne et l'écart type de chacune des deux variables. Que remarquer vous ?
2. Standardiser les variables X et Y en appliquant la transformation suivante :

$$X_{std} = \frac{X - moyenne}{\text{écart_type}}$$

3. Calculer la matrice de covariance donnée par la formule suivante :

$$V = \frac{1}{n-1} A_{std}^T A_{std}$$

4. Calculer les valeurs propres de V et ses vecteurs propres
5. Quelles sont les taux de variances expliqués par chaque vecteur propre ?
6. Normaliser les vecteurs propres comme suit :

$$U_{norm} = \frac{U}{\|U\|}$$

7. Projeter le jeu de donnée selon les vecteurs propres comme suit :

$$C = A_{std} \cdot U_{norm}$$

Partie 2 : Sur Python

1. Faites ces calculs sur Python sans utiliser les bibliothèques pandas et sklearn et comparer les résultats
2. Appliquer la méthode PCA de sklearn et comparer les résultats

Ex 2 :

On donne le jeu de donnée suivant :

```
Import numpy as np
```

```
rng = np.random.RandomState(1)
```

```
data = (rng.rand(2, 2)@rng.randn(2, 200)).T
```

3. Visualisez le nuage de point
4. Standardiser les données en utilisant le standardScaler
5. Appliquer la PCA de sklearn avec 2 composantes principales
6. Visualiser sur un même graphique :

- a. Le nuage de point d'origine
 - b. Le nouveau nuage de points projeté sur les axes principaux dans le repère d'origine en utilisant la fonction [pca.inverse_transform\(\)](#)
 - c. Les axes des composantes principales en utilisant la fonction [plt.quiver](#)(x_pos, y_pos, x_dir, y_dir, couleur)
7. Calculer les variances expliquées par les axes et les variances expliquées cumulées
8. Visualiser le diagramme en bar des variances expliquées
9. En déduire d'appliquer la PCA avec 1 composantes principales
10. Visualiser sur un même graphique le nuage standardisé et le nouveau nuage projeté avec l'axe de composante principale