

Calcul tensoriel

January 23, 2017

INB : l sera apprécié que les matrices dans votre rapport soient bien présentées : sous forme d'image (cf. les codes fournis) et/ou sous forme de sortie numpy bien arrondies.

1 Calcul tensoriel

Considérez les vecteurs `a=np.array(range(4))` et `b=np.array(range(3))`. Calculez sans faire de boucle les tenseurs suivants.

- $c_{ij} = a_i b_j$ (Aide : il faut glisser un indice en position 1 chez a et un indice en position 0 chez b)
- $d_{ij} = c_{ij} - a_i - a_j$
- $e_{ijk} = c_{ij} - d_{ik}$
- $\sum_{ik} e_{ijk}$

Pour la dernière, vous utiliserez que `sum` peut réduire plusieurs dimensions à la fois, mais attention, il faut mettre un t-uplet (et pas une liste), par exemple `np.sum(e, [0,1])`

Construisez des matrices carrées a et b puis calculez

- $c_{ik} = \sum_j a_{ij} b_{jk}$
- $c_{jk} = \sum_i a_{ij} b_{ik}$

On fera cela de deux manières différente : en utilisant `np.matmul` et `np.transpose` et en utilisant `expand_dim` et le produit `*` (ou bien `np.multiply`).

2 Centrer réduire, corrélation

Considérons une matrice de shape (n, p) . Chaque colonne correspondant à des caractéristiques d'un individu, chaque ligne correspondant à un individu. Par exemple les caractéristiques sont :

- Colonne $p = 0$, le poids.
- Colonne $p = 1$, la taille.
- Colonne $p = 1$, l'indice corporel.
- Colonne $p = 2$, la quantité de sang.
- Colonne $p = 3$, le nombre d'enfant.
- Colonne $p = 4$, le numéro de sécurité social.

- Colonne $p = 5$, le sexe (0=homme), (1=femme)
- Colonne $p = 6$, le tour de taille.
- Colonne $p = 6$, le tour de crane.

Ensuite, la ligne $n = 0$ correspond aux caractéristiques de l'individu numéro 0, la ligne $n = 1$ correspond aux caractéristiques de l'individu numéro 1 ...

Un tel tableau s'appelle une "dataframe" et la présentation ligne/colonne est toujours la même. Ainsi, si une dataframe représente des individus tirés aléatoirement dans la population les lignes sont indépendantes mais pas forcément les colonnes.

Question : dans l'exemple ci-dessus, quelles sont les colonnes les plus corrélées entre elles ? Essayer de faire un dessin pour présenter ces corrélations, toutes les présentations sont permise (graphique, flèches, arbres).

Construisons avec numpy une data-frame x toute simple.

```
sample_size=10
x0=np.random.random([sample_size])
x1=3*np.random.random([sample_size])
epsilon=0.1
x2=x0 +x1 +epsilon*np.random.random([sample_size])
x=np.transpose(np.array([x0,x1,x2]))
```

Les lignes/colonnes sont-elles indépendantes ?

On va dire que la première colonne représente l'age, la seconde la taille, et la troisième le poids.

- Calculez les caractéristiques moyennes de cette dataframe (age moyen, taille moyenne, poids moyen). Utiliser `np.mean` ou bien `np.sum`
- Calculez les écart-types de cette dataframe. Utilisez `np.std`, ou bien essayez sans, c'est un bon exercice.
- Calculez la dataframe centrée-réduite (ex: la première colonne sera l'age centré réduit des individus). Utilisez ce que l'on a appris sur le calcul tensoriel pour éviter les boucles.
- Calculer la matrice de corrélation de la dataframe de deux manière différente : 1/ avec `np.corrcoef` et 2/ en utilisant le travail précédent. Interprétez ces coefficient.

N'hésitez pas à faire ce même travail avec une sconde dataframe issue de votre imagination.