

DS - INFO - Info

Q1.

```
Select idpatient from MEDICAL where etat = "hemiparésie"
```

Q2.

```
Select nom, prenom from PATIENT  
Join MEDICAL on MEDICAL.idpatient = PATIENT.id  
Where MEDICAL.etat = "Spondylolisthésis";
```

Q3.

```
Select etat, count(*) as nbdepersonnes  
from MEDICAL group by MEDICAL.etat;
```

Q4.

Numpy est utile pour les grandes données car les données sont plus facilement et rapidement manipulées dans des vecteurs.

Q5.

• Pour le tableau :

On a 100 000 lignes et 6 colonnes. Chaque réel est codé sur 32 bit. En octet on a :

$$\frac{1}{8} \times 100\,000 \times 6 \times 32 = \underline{2,4\text{ Mo}}$$

• Pour le vecteur :

On en a 1 000 000 codés sur un octet.
Cela représente 0,7 Mo.

Donc au total la quantité de données nécessaires est de 2,5 Mo.

Q6.

```
def separation Par Groupe (data, etat) :  
    nc = [[]; [[]; [[]]  
    for i in range(len(data)):  
        nc = [  
  
    return nc
```


Q 8.

L'utilité de diagrammes de la diagonales et de présenter le résultat de façon visuelle.

Ces deux diagrammes servent à compter les relations, de corrélation entre les variables.

Q 9.

$$x_{\text{norm } j} = \frac{x_j - \min(x)}{\max(x) - \min(x)}$$

Q 10.

```
def min_max(x):  
    min = x[0]  
    max = x[0]  
    for i in x:  
        if i < min:  
            min = i  
        elif i > max:  
            max = i  
    return min, max
```


Q 11.

$d=0$, $L=[]$

```
def distance (z, data):  
    For i in range(len(data)):  
        For j in range(len(z)):  
             $d = d + (z[j] - data[i][j])^2$   
         $L = L.append(\sqrt{d})$   
    return L.
```

Q 12.

De la ligne 3 à 7, on crée une liste de liste, ie un vecteur. Chaque coordonnée du vecteur correspond la distance d'un n -uplet et le nombre correspondant à ce n -uplets.

A la ligne 7 on crée cette liste.

dist correspond à la distance euclidienne entre le n -uplet z et un n -uplet x .

Q 13.

La diagonale correspond au nombre de personnes correctement skinné pour chaque état, un état correspondant à une ligne.

Cette matrice sert donc au rendu compte du nombre de personnes justement prédit et aux erreurs.

Q14.

Il y a taux d'état prédit correctement compris entre
5.9% et 74%.

L'algorithme donne donc une valeur approchée avec une
erreur d'au moins 20%, ce qui paraît assez empâté.

Q15.

def moyenne (x):

 t = 0

 for i in range (len(x)):

 t += x[i]

 return t / (len(x))

def variance (x):

 v = 0, n = 0

 for i in x:

 n += i²

 v = moyenne(x) ** 2 - moyenne (v / len(x))

 return v