Exercice 6

6.1 Sale temps sur Seattle (suite)

- À partir de l'histogramme représentant la distribution des hauteurs de précipitations à Seattle (*cf.* précédent TD), préciser les axes de la figure ainsi que le titre de l'histogramme
- Changer l'axe y pour passer en échelle logarithmique
- Annoter la figure pour faire apparaître la valeur moyenne et l'écart-type des précipitations
- Représenter sur une seconde figure, la variation temporelle des températures minimale et maximale. On s'aidera du code suivant pour convertir les jours extraits du fichier `csv` en une grandeur exploitable par `numpy`

```
import pandas as pd
day = pd.to_datetime(day, format="%Y%m%d")
```

- Sur une troisième figure, représenter la distribution des valeurs minimale et maximale des températures. Faire en sorte que les distributions puissent être visible y compris au niveau des zones de recouvrement

6.2 Fonctions discontinues[a]

- Représenter la fonction d'Heaviside $\Theta(x)$ définie par

$$\begin{cases} \Theta(x) & = 1 \quad \text{si } x \geq 0 \\ & = 0 \quad \text{sinon} \end{cases}$$

- Le noyau radioactif de ^{11}C est un émetteur β^+ utilisé lors de tomographie par émission de positrons. La réaction permettant la production de cet élement est la suivante

$$p +_7^{14} \mathrm{N} \rightarrow_6^{11} \mathrm{C} + \alpha$$

En tenant compte du taux de production de ^{11}C par irradiation et du nombre de noyaux se désintégrant, on peut montrer que le nombre de noyaux de ^{11}C au cours du temps s'exprime de la façon suivante

$$\begin{cases} n(t) & = \frac{n_i}{\lambda}\left(1 - e^{-\lambda t}\right) \quad \text{si } t \leq t_0 \\ & = n(t_0)\, e^{-\lambda(t-t_0)} \quad \text{si } t > t_0 \end{cases}$$

où $\lambda = \frac{\ln 2}{T_{1/2}}$ et $T_{1/2} = 20.36$ minutes. n_i correspond au taux d'irradition et est égal à 3 10^8 noyaux/s. Représenter $n(t)$ pour $t_0 = 3$ heures.

6.3 *Iris setosa, Iris virginica et Iris versicolor*

Le jeu de données *Iris* contient les propriétés morphologiques de 3 espèces de fleur d'iris collectées par Edgar Anderson. Ce jeu de données est surtout reputé par l'utilisation faite en 1936 par Ronald Fisher pour démontrer la puissance de son algorithme d'analyse discriminante linéaire à même de séparer les 3 espèces de fleur d'iris. Ces données sont devenues depuis un cas typique pour de nombreuses techniques de classification automatique en *machine learning*.

[a] on pourra s'aider ou pas de la fonction `piecewise` de `numpy`

- Télécharger le fichier `iris.csv` qui contient la longueur et la largeur des sépales en cm (colonne 1 et 2), la longueur et la largeur des pétales en cm (colonne 3 et 4) ainsi qu'une dernière colonne dont la valeur, 0, 1 ou 2, est relative à l'espèce de la fleur d'iris (0 = *iris setosa*, 1 = *iris versicolor*, 2 = *iris virginica*). Charger ce fichier dans un tableau `numpy`
- Représenter les distributions normalisées de longueur et de largeur des sépales et des pétales pour les 3 espèces
- Représenter dans un diagramme (largeur des sépales *vs.* longueur des sépales), la largeur des pétales ainsi que l'espèce de fleur d'iris considérée
- Représenter l'ensemble des combinaisons possibles de données (largeur des sépales *vs.* longueur des sépales, largeur des sépales *vs.* largeur des pétales…), les figures situées dans la diagonale devant correspondre aux distributions normalisées des différentes grandeurs.