GIMBERT Vincent
GUERIN Clément

L3 Informatique

Rapport TP1
Probabilité et statistique

Table des matières :

Table des figures :

Table des tableaux :

Objectif du TP

L'objectif de ce TP est de se familiariser avec les différents outils mathématiques permettant de déduire, à partir d'un nuage de points, une approximation et ainsi pouvoir prédire avec le plus de précision possible des cas de figure qui ne sont pas présents dans les tableaux d'observations.

Présentation des résultats

Dans cette partie, nous allons présenter les résultats de nos implémentations, ainsi qu'en faire une comparaison avec les fonctions proposées par la bibliothèque externe de python, sklearn.

Il s'agira alors de déterminer quelle est l'implémentation la plus précise.

Résultats pour l'été

Premièrement, nous allons présenter les résultats pour le tableau d'été :

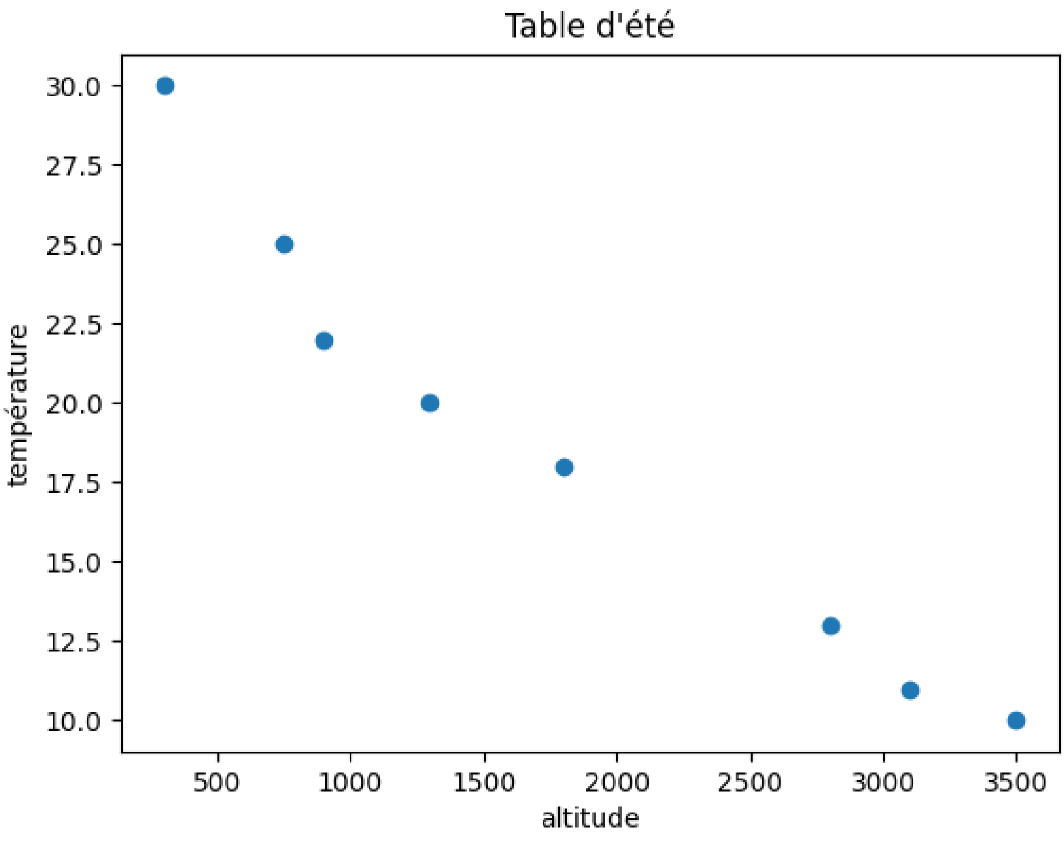

Figure 1 : Tableau montrant la température en fonction de l'altitude en saison d'été.

Nous pouvons voir que les points semblent suivre une fonction affine. Nous allons donc déterminer le coefficient de cette fonction grâce à la méthode des moindres carrés.

En utilisant la méthode que nous avons implémenté, nous trouvons une équation ax+b avec :

a = -0.005772634870992963

b = 29.05182173573104

Ci-dessous une représentation graphique de cette fonction affine :

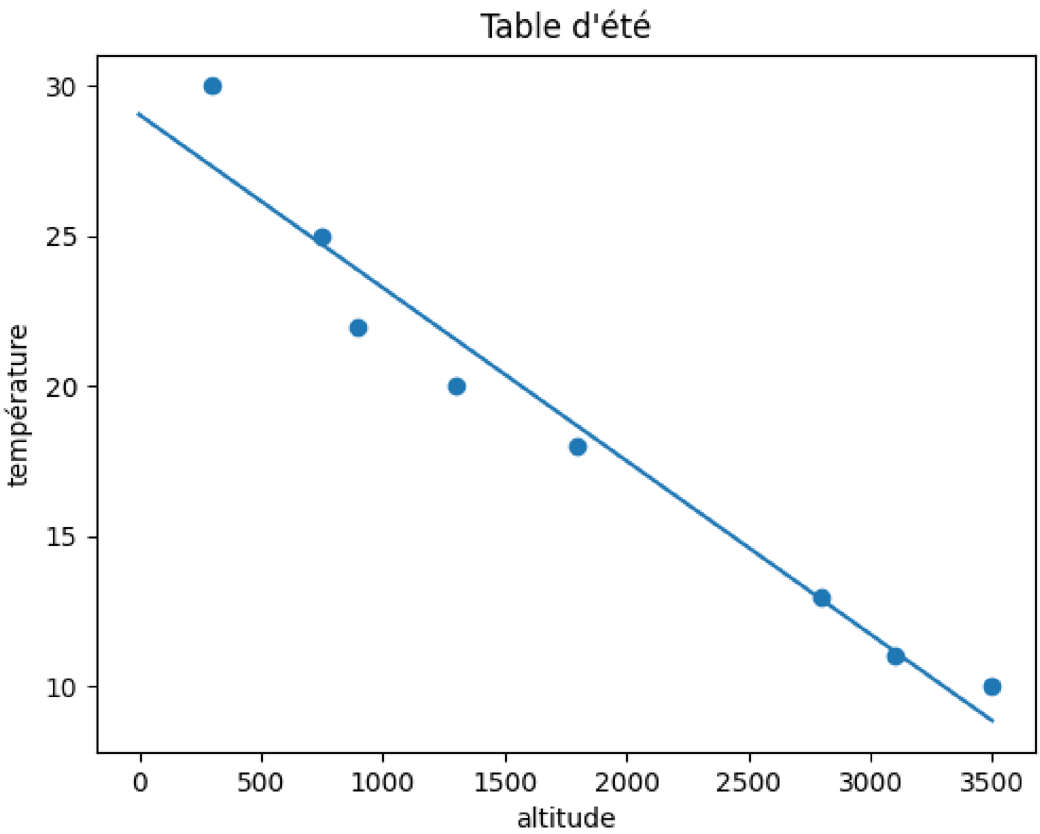

Figure 2: Tableau montrant la température en fonction de l'altitude en saison d'été avec fonction affine approximative.

Ci-dessous, le tableau comparatif qui compare nos méthodes avec celles de sklearn :

Tableau 1 : Tableau comparatif entre nos implémentations et ceux de sklearn pour le cas de l'été.

	Nos Implémentations	**Bibliothèque sklearn**
A	-0.005772634870992963	-0.005772634870992964
B	29.05182173573104	29.05182173573104
Maximum de vraisemblance	1.8627756059421445	1.8627756059421419
Méthode Optimisation	6303652135746414.0	6303652135746414.0
RMSE	1.36483537686497	1.862775605942142
Coefficient de détermination	0.9571621851310466	0.9571621851310467

Nous pouvons remarquer que les différences sont très faibles :

- Le coefficient A de notre implémentation des moindres carrés diffère d'une unité de l'ordre de 10^{-18}.

- Le coefficient B est exactement le même
- Le maximum de vraisemblance diffère de l'ordre de 10^{-15}.
- Le coefficient de détermination lui aussi diffère l'ordre de 10^{-15}.
- La méthode d'optimisation est identique pour les deux implémentations.
- Le RMSE en revanche diffère de l'ordre 10^{-1} en notre faveur, car plus le RMSE est petit, plus notre modèle est précis. Ce qui est étrange ici, c'est le fait que le RMSE de sklearn est très proche du maximum de vraisemblance, a 10^{-15} près.

Résultats pour l'hiver

Ensuite, voici nos résultats pour le tableau d'hiver :

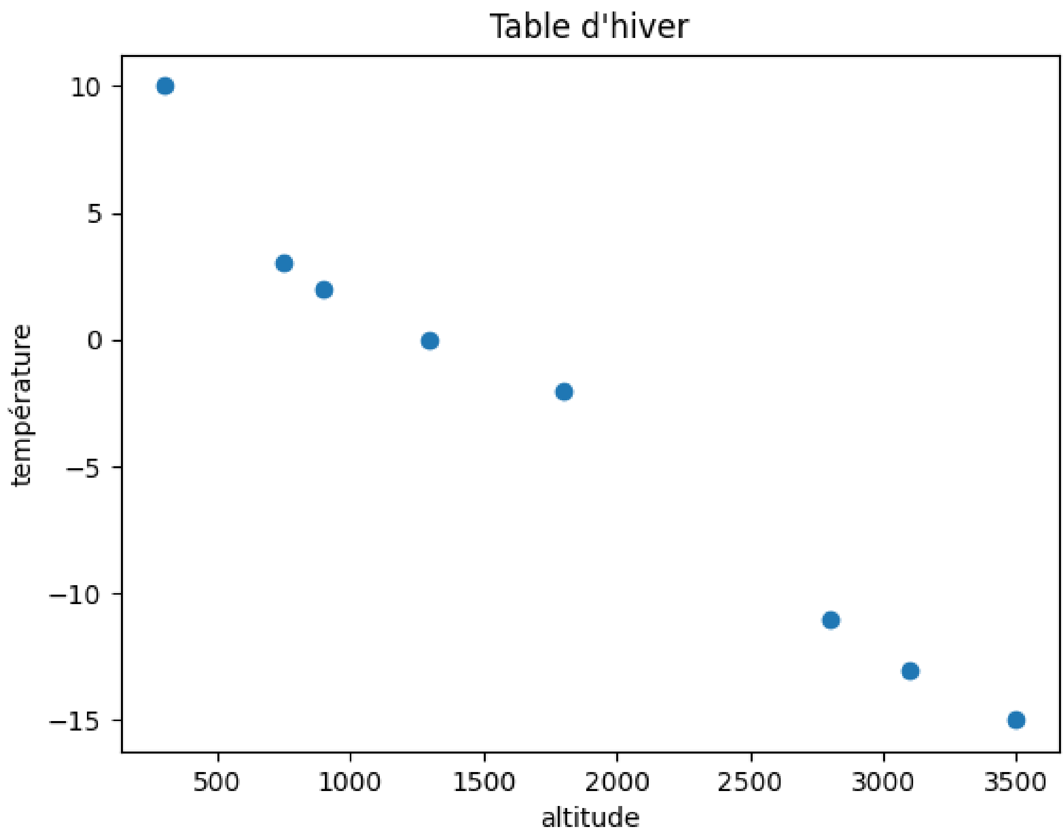

Figure 3 : Tableau montrant la température en fonction de l'altitude en saison d'hiver

Grâce à la méthode des moindres carrés, nous trouvons une équation ax+b avec :

a = -0.007324472243940579

b = 9.97982799061767

Ci-dessous une représentation graphique de cette fonction affine :

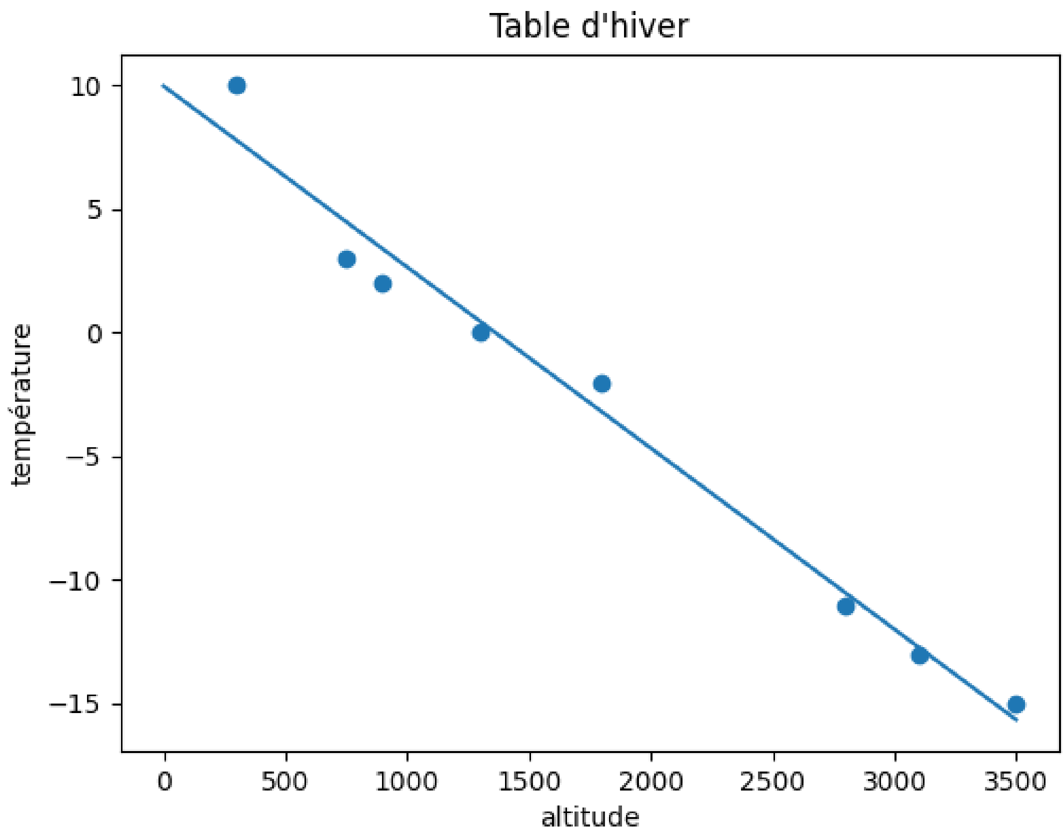

Figure 4 : Tableau montrant la température en fonction de l'altitude en saison d'hiver avec fonction affine approximative

Ci-dessous, le tableau comparatif qui compare nos méthodes avec celles de sklearn :

Tableau 2 : Tableau comparatif entre nos implémentations et ceux de sklearn pour le cas de l'hiver.

	Nos Implémentations	**Bibliothèque sklearn**
A	-0.007324472243940579	-0.007324472243940579
B	9.97982799061767	9.97982799061767
Maximum de vraisemblance	1.430023455824863	1.430023455824863
Méthode Optimisation	2035280353161662.2	2035280353161662.2
RMSE	1.1958358816429886	1.430023455824864
Coefficient de détermination	0.9791046800975364	0.9791046800975363

Nous pouvons remarquer, ici encore, que les différences sont très faibles :

- Le coefficient A et B sont exactement les mêmes quel que soit le modèle
- Par conséquent, le maximum de vraisemblance ne diffère pas non plus.
- Le coefficient de détermination lui en revanche diffère l'ordre de 10^{-15}.
- La méthode d'optimisation est identique pour les deux implémentations.

- Le RMSE diffère de l’ordre 10^{-1} toujours en notre faveur. Le RMSE de sklearn est encore une fois très proche du maximum de vraisemblance, a 10^{-15} près.

Réponses aux questions :

Pour cette partie nous allons répondre aux questions posées par le TP.

Quel est le modèle le plus précis ?

D’après nos données, il semblerait que le modèle le plus précis soit notre implémentation plutôt que celle de la bibliothèque sklearn. En effet, comme vu plus haut, le RMSE est plus petit pour nos implémentations, ce qui signifie qu’elles sont précises, et le coefficient de détermination est, pour nos implémentations, plus proche de 1.

Il est à noter que ces différences sont très faibles, presque insignifiante.

Quelle est la température à 1000 mètres en été et en hiver ?

Pour répondre à cette question, il suffit de résoudre ax+b, x étant le paramètre 1000, a et b étant calculé par la méthode des moindres carrés. Nous trouvons donc les résultats suivants :

Tableau 3 : Tableau comparatif sur la prédiction de température en degré à 1000 mètres selon la saison

Altitude\Saison	**Eté**	**Hiver**
1000 m	23.279186864738076	2.655355746677091

Supposons qu’il fasse 15 degrés à 300m, combien devrait-il faire à 1000m ?

Cette question est un peu plus difficile à répondre, car nous n’avons pas d’informations concernant la saison. Pour ce faire, nous allons donc faire une moyenne des coefficients directeurs trouvées.

Puis résoudre l’équation b = A’ * 300 – 15, A’ étant le coefficient directeur calculé précédemment, afin de trouver la constante.
On obtient ainsi une nouvelle droite, et on effectue la fonction affine.
On obtient ainsi le résultat suivant : 10.41601250977326 degrés.

Conclusion

Bien que les différences de résultats entre nos implémentations et celles de sklearn soient très faible, il semblerait que les nôtres soient légèrement plus précise, pour des raisons inconnues.