Machine Learning III
Introduction à scikit-learn

Nicolas Bourgeois

Télécharger

Data and Cheatsheets :

`ouralou.fr/Resources/epita/C3.zip`

Exercice

Exercice

Importez les données de data1.csv et testez une régression linéaire entre la longueur et l'épaisseur des pétales.

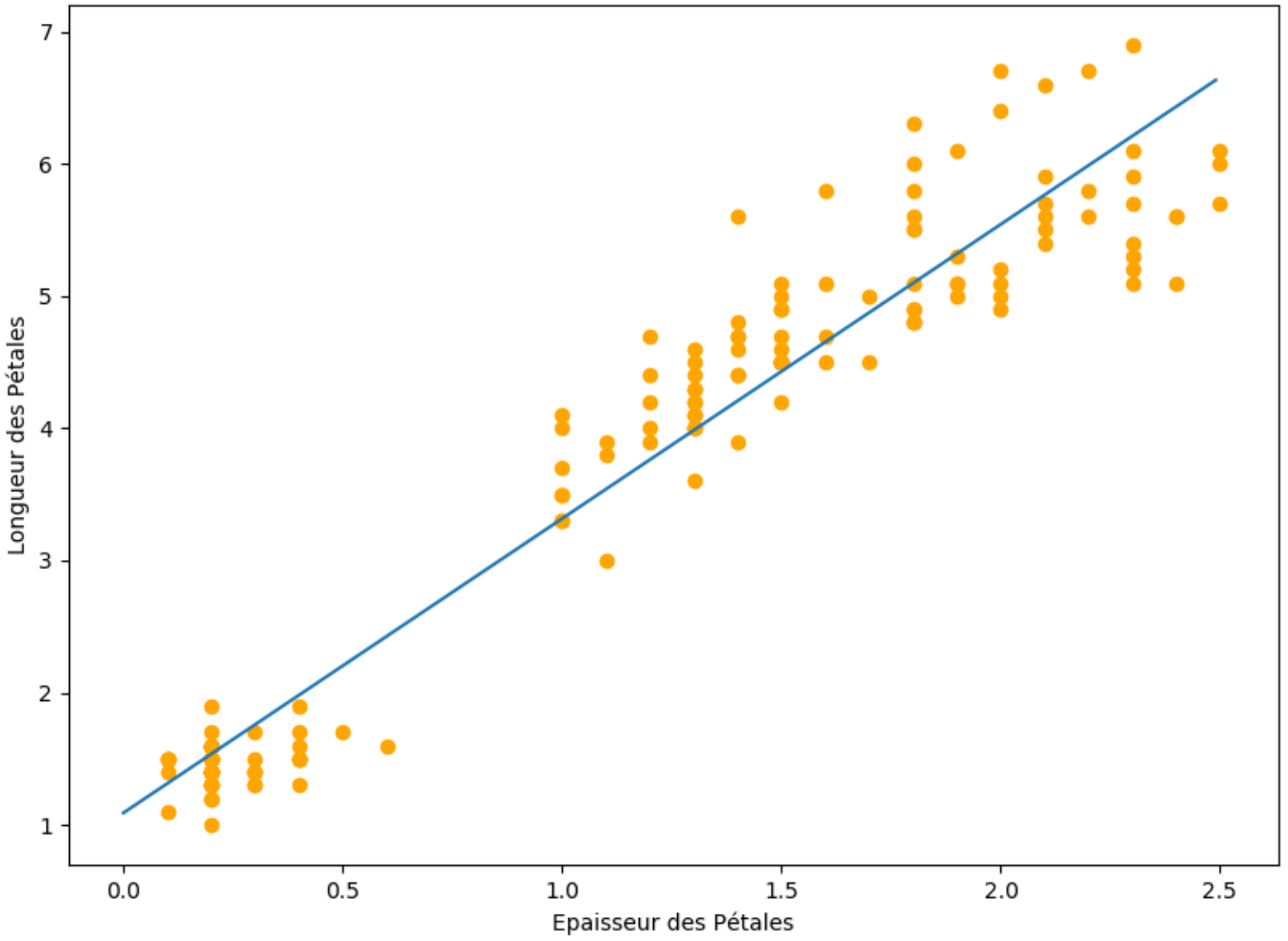
Exercice

Même question, cette fois entre la longueur des cépales et la largeur des pétales. Comparez les scores des deux régressions.

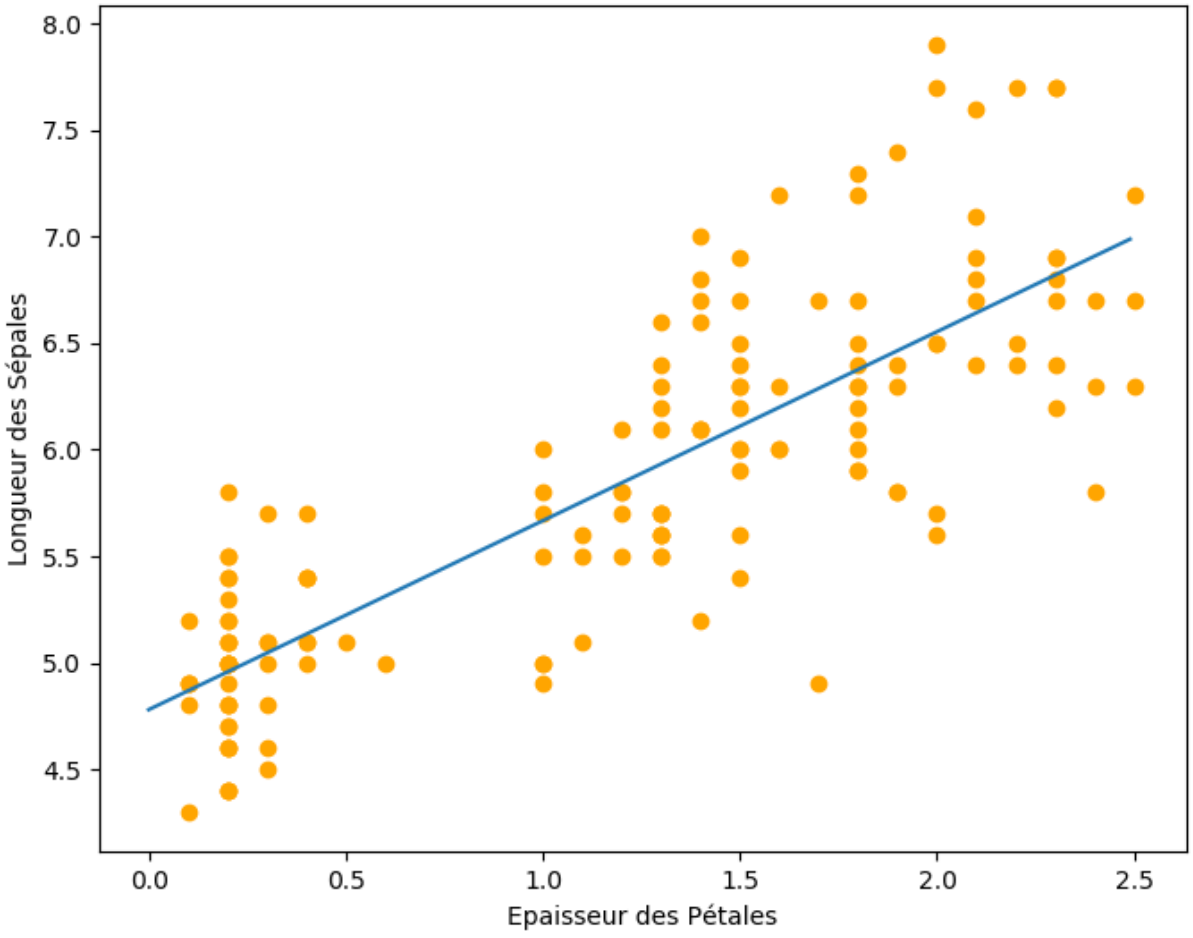
Exercice

Dans les deux cas, représentez les données et la droite de régression.

Résultat attendu (1)



Résultat attendu (2)



Solution (parties 1 et 2)

```
import pandas as pd
from sklearn.linear_model import LinearRegression
iris = pd.read_csv('./C3/data1.csv')
X = iris.PetalWidth.values.reshape(-1,1)
Y = iris.PetalLength
lr = LinearRegression()
lr.fit(X,Y)
print(lr.coef_)

##question 2
print(lr.score(X,Y))
Y = iris.SepalLength
lr.fit(X,Y)
print(lr.score(X,Y))
#attention ce score est trompeur
```

Solution (partie 3)

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
from matplotlib import pyplot as plt
iris = pd.read_csv('./C3/data1.csv')
X = iris.PetalWidth.values.reshape(-1,1)
Y = iris.PetalLength
lr = LinearRegression()
lr.fit(X,Y)
plt.scatter(X,Y,c="orange")
plt.xlabel("Epaisseur des Petales")
plt.ylabel("Longueur des Petales")
test = np.arange(0,2.5,0.01).reshape(-1,1)
plt.plot(test,lr.predict(test))
plt.show()
```

Exercice

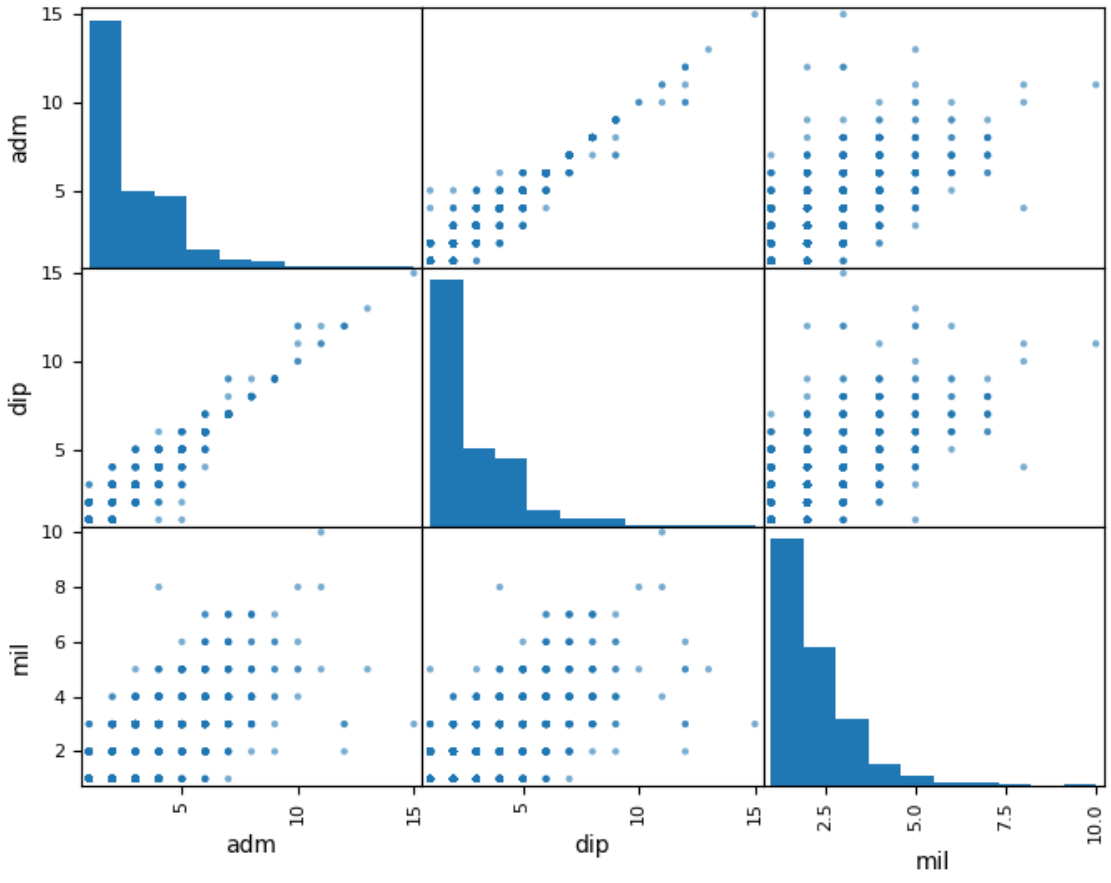
Exercice

Importez les données de data2.csv, gardez uniquement les champs adm, dip et mil et entraînez une ACP dessus.

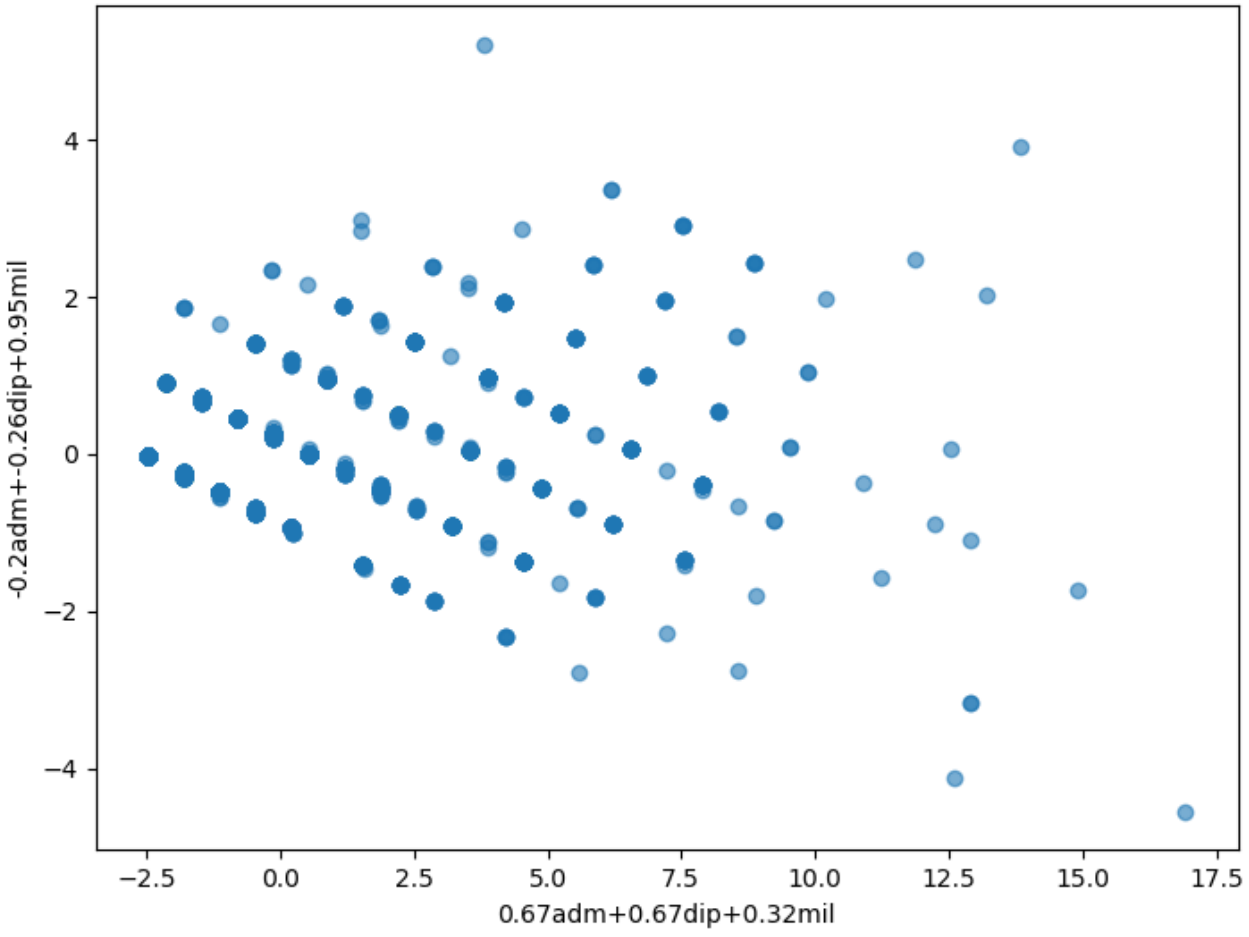
Exercice

Comparez graphiquement les représentations des données utilisant deux axes standards (via une scatter matrix) et celles utilisant les axes de l'ACP.

Résultat attendu (1)



Résultat attendu (2)



Solution (partie 1)

```
import pandas as pd
from sklearn.decomposition import PCA
df = pd.read_csv('./C3/data2.csv',sep=";")
data = df[['adm','dip','mil']].dropna()
acp = PCA()
acp.fit_transform(data)
print(acp.explained_variance_)
print(acp.components_)
```

Solution (partie 2)

```
import pandas as pd
from pandas.plotting import scatter_matrix
from matplotlib import pyplot as plt
df = pd.read_csv('./C3/data2.csv',sep=";")
data = df[['adm','dip','mil']]
scatter_matrix(data,alpha=0.6,diagonal='hist')
plt.show()
```

Solution (partie 3)

```
import pandas as pd
from sklearn.decomposition import PCA
from matplotlib import pyplot as plt
df = pd.read_csv('./C3/data2.csv',sep=";")
data = df[['adm','dip','mil']].dropna()
acp = PCA()
rot = acp.fit_transform(data)
plt.scatter(rot[:,0],rot[:,1],alpha=0.6)
cf = acp.components_
plt.xlabel("{0:.2}adm+{1:.2}dip+{2:.2}mil".format(
    cf[0][0],cf[0][1],cf[0][2]))
plt.ylabel("{0:.2}adm+{1:.2}dip+{2:.2}mil".format(
    cf[1][0],cf[1][1],cf[1][2]))
plt.show()
```