

# day-1-handin

October 12, 2023

## 1 Basic Data Science in Python - Handin 1

### 1.1 Leihui Li, 12/10/2023, leihui@mpe.au.dk

This handin is individual and mandatory to pass the course.

Hand in this .ipynb file and the compiled pdf, no later than 13/10 kl 09:30.

```
[1]: import os
os.environ["OMP_NUM_THREADS"] = '1'

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import sklearn
from sklearn import datasets
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.mixture import GaussianMixture as GM
```

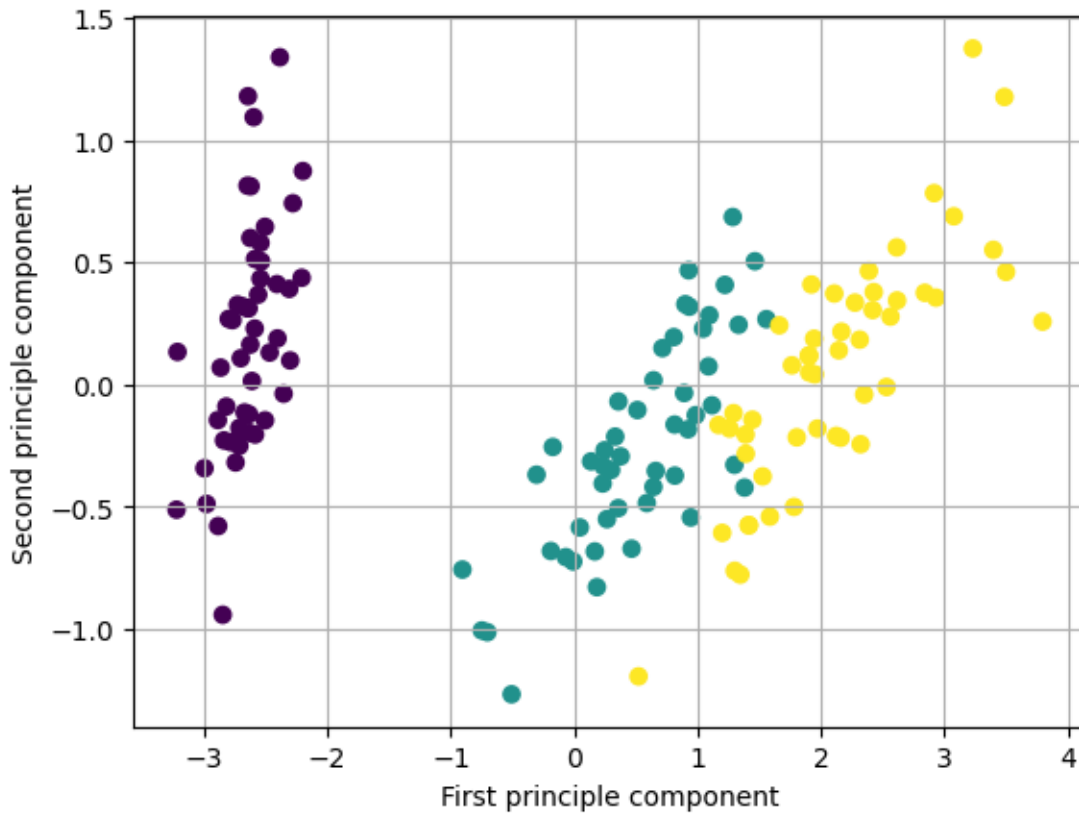
#### 1.1.1 PCA and EM

Use Principal Component Analysis to reduce the dimensionality of the Iris Dataset to 2D, and then use Gaussian Mixtures to assign the points to three clusters. Plot the result

```
[2]: iris = datasets.load_iris()
X = iris.data
### YOUR CODE HERE
pca = PCA(n_components=2)
pca.fit(X)
x_pca = pca.transform(X)
print (X.shape, x_pca.shape, pca.explained_variance_ratio_)

# PLOT
plt.scatter(x_pca[:, 0], x_pca[:, 1], c= iris.target)
plt.xlabel('First principle component')
plt.ylabel('Second principle component')
plt.grid()
plt.show()
```

(150, 4) (150, 2) [0.92461872 0.05306648]



```
[4]: X_train = np.vstack(X)
      clustering = GM(n_components=3, covariance_type='full')
      clustering.fit(X_train)
      labels = clustering.predict(X_train)
      plt.scatter(X_train[:, 0], X_train[:, 1], c=labels)
      plt.title('Gaussian Mixtures')
      plt.grid()
      plt.show()
```

