

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

# Reading the data from csv file
df = pd.read_csv("2020_em_clustering.csv", sep=',', header=None)
df = df.transpose()

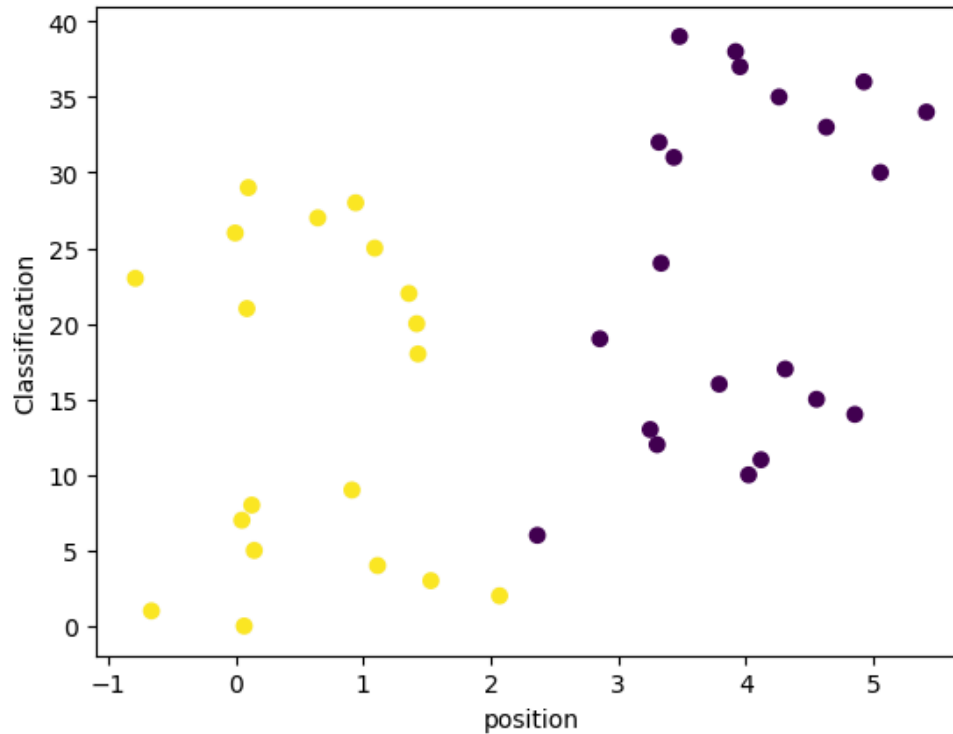
# Defining the number of clusters
kmeans = KMeans(n_clusters=2)

# Fitting the model
kmeans.fit(df)

# Predicting the clusters
predictions = kmeans.predict(df)

# Plotting the clusters
plt.scatter(df.iloc[:,0], [i for i in range(df.shape[0])], c=predictions)
plt.xlabel("position")
plt.ylabel("Classification")
plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 100 in the future. Please set `n_init` to the new default value of 100 for this warning to be silenced.
warnings.warn(
```



```
# Reading the data from csv file
df = pd.read_csv("2020_em_clustering.csv", sep=',', header=None)
df = df.transpose()

# Defining the number of clusters
em = GaussianMixture(n_components=2)

# Fitting the model,
"""
Fitting the model: The fit method is used to fit the EM algorithm to the data.
This step involves iteratively estimating the parameters of the Gaussian distribution for each cluster until convergence.
"""
em.fit(df)

"""
Predicting the clusters: The predict method is used to make predictions on which cluster each data point belongs to.
"""
```

```
predictions = em.predict(df)

# Plotting the clusters
plt.scatter(df.iloc[:,0], [i for i in range(df.shape[0])], c=predictions)
plt.xlabel("position")
plt.ylabel("Classification")
plt.show()
```

