

kmean

May 10, 2022

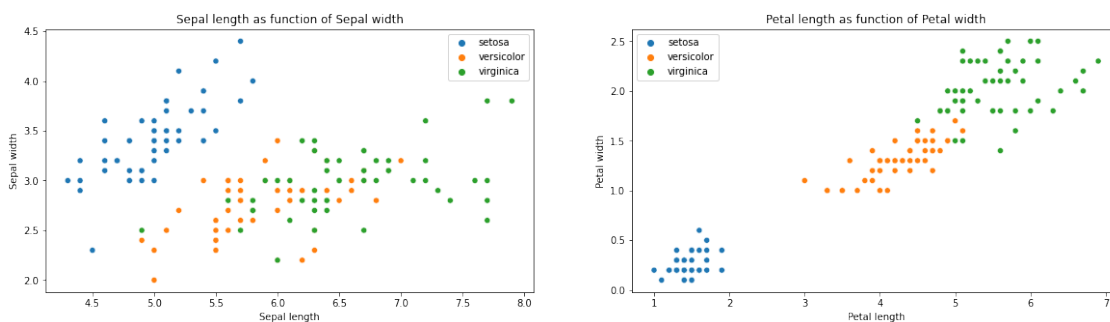
```
[ ]: import matplotlib.pyplot as plt
import seaborn as sbn
import numpy as np
from sklearn.datasets import load_iris
from sklearn.cluster import KMeans
```

```
[ ]: data = load_iris()
X = data["data"]
y = data["target"]
labels = data["target_names"]
y_labels = [labels[i] for i in y]
X_labels = ["Sepal length", "Sepal width", "Petal length", "Petal width"]
```

```
[ ]: plt.figure(figsize=(20,5))

j=1
for i in range(0, len(X_labels), 2):
    plt.subplot(1, 2, j)
    sbn.scatterplot(x=X[:,i], y=X[:,i+1], hue=y_labels)
    plt.xlabel(X_labels[i])
    plt.ylabel(X_labels[i+1])
    plt.title(f"{X_labels[i]} as function of {X_labels[i+1]}")
    j += 1

plt.show()
```



```
[ ]: kmeans = KMeans(n_clusters=3)
kmeans.fit(X)
predictions = kmeans.predict(X)
clusters = [f"Cluster {i+1}" for i in range(3)]
predictions_labels = [clusters[i] for i in predictions]
```

```
[ ]: fig = plt.figure(figsize=(20,5))

j=1
for i in range(0, len(X_labels), 2):
    plt.subplot(1, 2, j)
    sbn.scatterplot(x=X[:,i], y=X[:,i+1], hue=y_labels)
    plt.xlabel(X_labels[i])
    plt.ylabel(X_labels[i+1])
    plt.title(f"{X_labels[i]} as function of {X_labels[i+1]}")
    fig.suptitle("Before clustering")
    j += 1

plt.show()

fig = plt.figure(figsize=(20,5))

j=1
for i in range(0, len(X_labels), 2):
    plt.subplot(1, 2, j)
    sbn.scatterplot(x=X[:,i], y=X[:,i+1], hue=predictions_labels)
    plt.scatter(x=kmeans.cluster_centers_[i], y=kmeans.cluster_centers_[i+1], c='r', label="Center of clusters")
    plt.legend()
    plt.xlabel(X_labels[i])
    plt.ylabel(X_labels[i+1])
    plt.title(f"{X_labels[i]} as function of {X_labels[i+1]}")
    fig.suptitle("After clustering")
    j += 1

plt.show()
```

