USL_KNN_IPYNB

March 2, 2025

[ ]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.datasets import load_iris
from mpl_toolkits.mplot3d import Axes3D
```

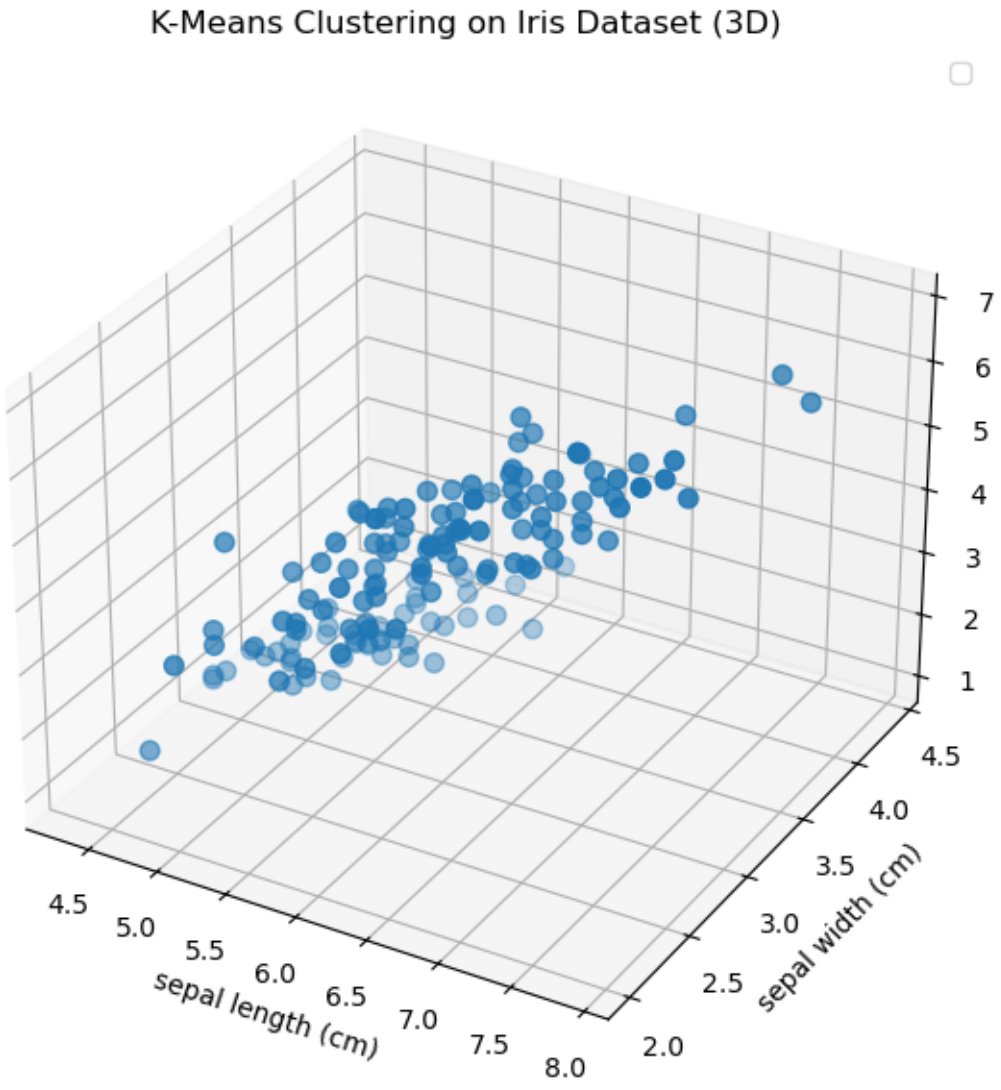
[ ]:
```
# Load the iris dataset
iris = load_iris()
df = pd.DataFrame(iris.data, columns=iris.feature_names)
```

[ ]:
```
# Select three features for 3D visualization
features = ['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)']
data = df[features]
```

[ ]:
```
fig = plt.figure(figsize=(10, 7))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(data.iloc[:, 0], data.iloc[:, 1], data.iloc[:, 2], cmap='viridis',␣
 ↪marker='o', s=50)
ax.set_xlabel(features[0])
ax.set_ylabel(features[1])
ax.set_zlabel(features[2])
ax.set_title('K-Means Clustering on Iris Dataset (3D)')
ax.legend()
```

```
C:\Users\Neil\AppData\Local\Temp\ipykernel_38364\4258469181.py:3: UserWarning:
No data for colormapping provided via 'c'. Parameters 'cmap' will be ignored
  ax.scatter(data.iloc[:, 0], data.iloc[:, 1], data.iloc[:, 2], cmap='viridis',
marker='o', s=50)
C:\Users\Neil\AppData\Local\Temp\ipykernel_38364\4258469181.py:8: UserWarning:
No artists with labels found to put in legend.  Note that artists whose label
start with an underscore are ignored when legend() is called with no argument.
  ax.legend()
```

[ ]:
```
<matplotlib.legend.Legend at 0x29ad4a8b4a0>
```



[ ]:
```
# Perform K-Means clustering
kmeans = KMeans(n_clusters=3, random_state=42)
kmeans.fit(data)
labels = kmeans.labels_
centroids = kmeans.cluster_centers_
```

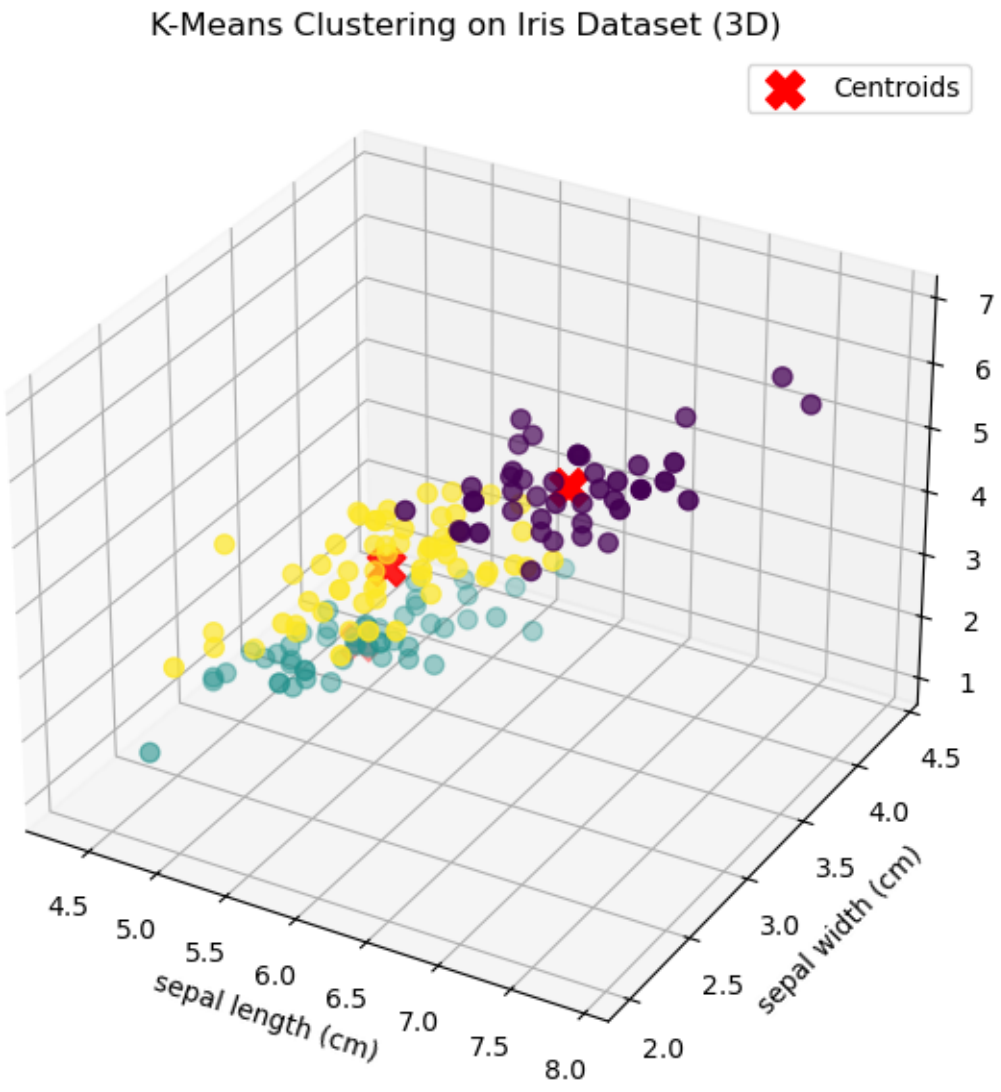
```
c:\Users\Neil\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1429:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=1.
  warnings.warn(
```

[ ]:
```
# Plot the clusters in 3D
fig = plt.figure(figsize=(10, 7))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(data.iloc[:, 0], data.iloc[:, 1], data.iloc[:, 2], c=labels,␣
 ↪cmap='viridis', marker='o', s=50)

# Plot centroids
ax.scatter(centroids[:, 0], centroids[:, 1], centroids[:, 2], c='red',␣
 ↪marker='X', s=200, label='Centroids')

# Labels and title
ax.set_xlabel(features[0])
ax.set_ylabel(features[1])
ax.set_zlabel(features[2])
ax.set_title('K-Means Clustering on Iris Dataset (3D)')
ax.legend()

plt.show()
```


K-Means Clustering on Iris Dataset (3D)
Centroids
sepal length (cm)
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
sepal width (cm)
2.0
2.5
3.0
3.5
4.0
4.5
1
2
3
4
5
6
7