

Talha Khan (2303.009.KHI.DEG)

Muhammad Moiz Khan (2303.022.KHI.DEG)

ASSIGNMNET 3.3

```
In [50]: import numpy as np
import pandas as pd
from sklearn import datasets
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
import warnings
warnings.filterwarnings('ignore')
```

```
In [51]: # load dataset
data = datasets.load_iris()
```

```
In [52]: x = data.data
```

```
In [53]: # load dataset into a dataframe
df = pd.DataFrame(data['data'], columns=data['feature_names'])
df
```

Out[53]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
...
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

150 rows × 4 columns

Without Reducing Dimensions

```
In [54]: # model implementation
model = KMeans(n_clusters=3, n_init=1, max_iter=100)
model.fit(x)
```

```
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at 0x7fcc2d562e60>
Traceback (most recent call last):
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at 0x7fcc28bd7e20>
Traceback (most recent call last):
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
```

Out[54]:

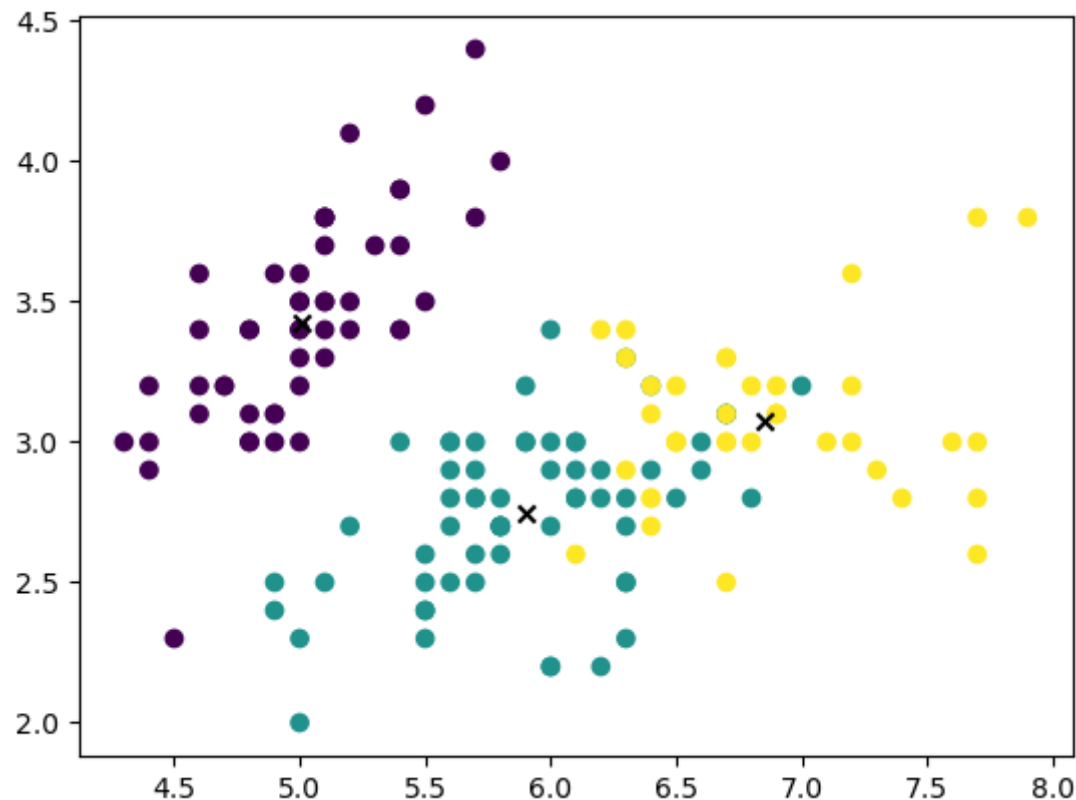
```
▼ KMeans
KMeans(max_iter=100, n_clusters=3, n_init=1)
```

In [43]:

```
all_predictions = model.predict(x)
centroids = model.cluster_centers_
```

```
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<
locals>.match_module_callback at 0x7fcc28bb6200>
Traceback (most recent call last):
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
```

```
In [44]: # plot clusters
plt.scatter(x[:,0], x[:,1], c=all_predictions)
plt.scatter(centroids[:,0], centroids[:,1], marker='x', color="black")
plt.show()
```



PCA

```
In [45]: # dimensionality reduction using PCA
pca = PCA(n_components=2)
x_reduced = pca.fit_transform(x)
```

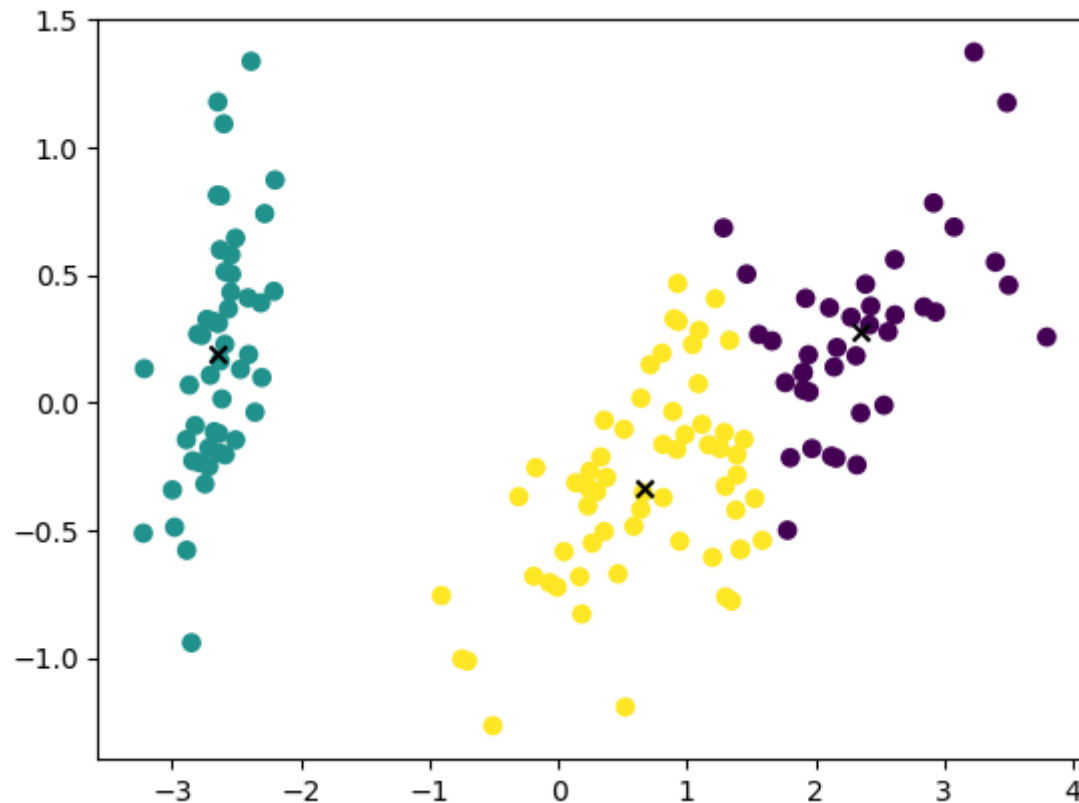
```
In [47]: x_reduced.shape
```

```
Out[47]: (150, 2)
```

```
In [48]: # model implementation on reduced dimensioned data
model = KMeans(n_clusters=3, n_init=1, max_iter=100)
model.fit(x_reduced)
all_predictions = model.predict(x_reduced)
centroids_PCA = model.cluster_centers_
```

```
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<
locals>.match_module_callback at 0x7fcc28bd6440>
Traceback (most recent call last):
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<
locals>.match_module_callback at 0x7fcc28bd6440>
Traceback (most recent call last):
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<
locals>.match_module_callback at 0x7fcc28bd40d0>
Traceback (most recent call last):
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/talhakhan/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
```

```
In [49]: # plot clusters
plt.scatter(x_reduced[:,0], x_reduced[:,1], c=all_predictions)
plt.scatter(centroids_PCA[:,0], centroids_PCA[:,1], marker='x', color="black")
plt.show()
```

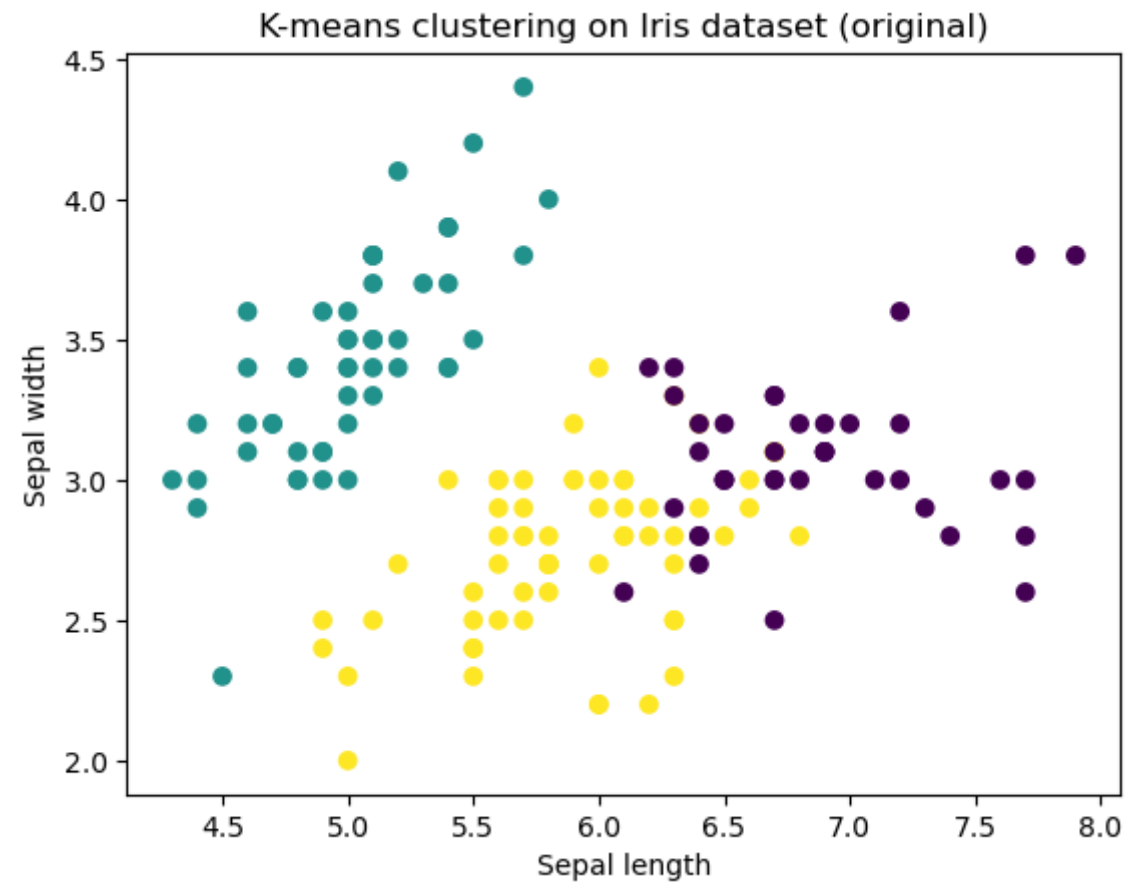


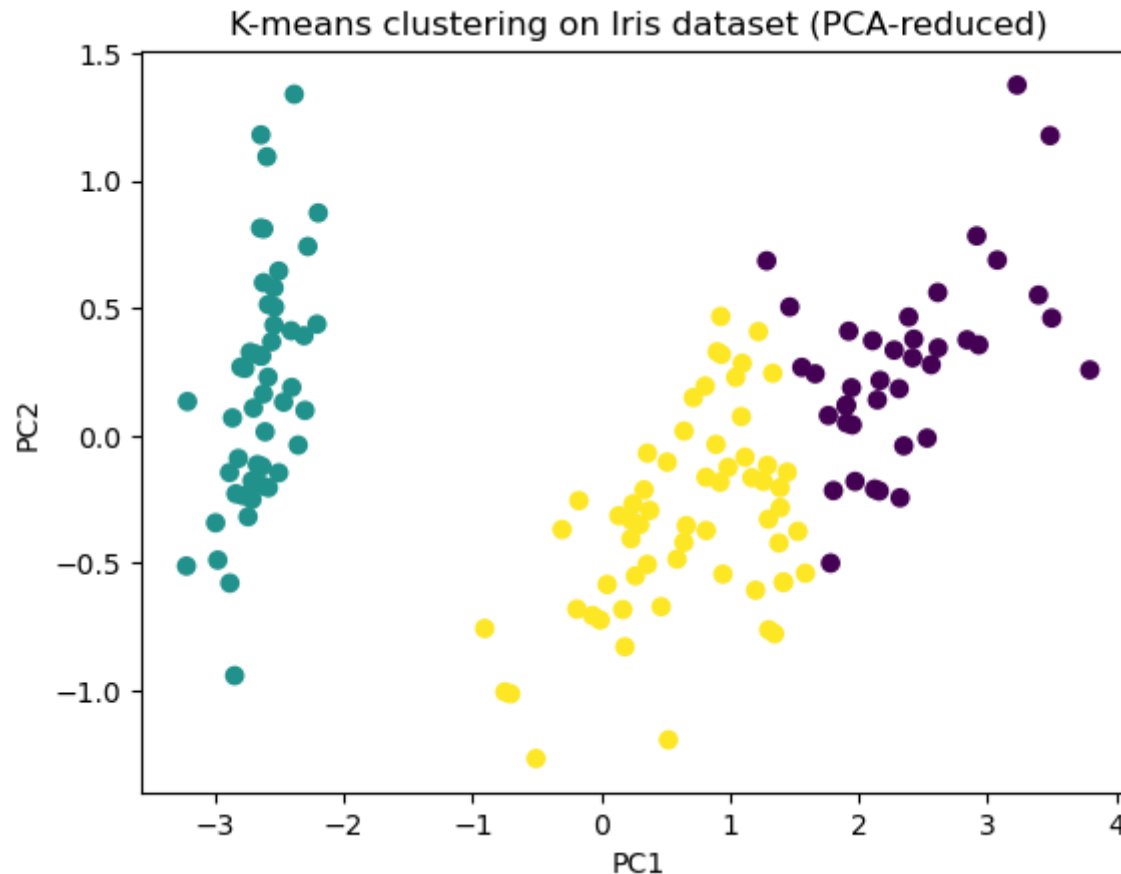
Comparison

```
In [59]: # Plot original data
plt.scatter(x[:, 0], x[:, 1], c=all_predictions)
plt.title("K-means clustering on Iris dataset (original)")
plt.xlabel("Sepal length")
plt.ylabel("Sepal width")
plt.show()

# Plot reduced data
plt.scatter(x_reduced[:, 0], x_reduced[:, 1], c=all_predictions)
plt.title("K-means clustering on Iris dataset (PCA-reduced)")
plt.xlabel("PC1")
```

```
plt.ylabel("PC2")  
plt.show()
```





It is observed that the clustering is similar in both cases, but the clusters are more compact in the reduced dataset. This is because PCA reduces the dimensionality of the data and removes the less informative dimensions. As a result, the remaining dimensions contain more relevant information and are more suitable for clustering.

In []: