ASSIGNMENT 3.3

on

Unsupervised Machine Learning

Submitted by:

Haseebullah Shaikh (2303.KHI.DEG.015)

and

Faiza Gulzar Ahmed (2303.khi.deg.001)

Dated : 20th Apr 2023

Task 01: Perform k-means clusterization on the Iris dataset. Repeat the procedure on the dataset reduced with PCA, and then compare the results.

In [1]:
```
import matplotlib.pyplot as plt
from sklearn import datasets
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
```

In [2]:
```
iris = datasets.load_iris()
```

In [3]:
```
x = iris.data
```

In [4]:
```
x.shape
```

Out[4]:
```
(150, 4)
```

Training the model on 4 features

In [5]:
```
model = KMeans(n_clusters=3, n_init=1, max_iter=100)
model.fit(x)

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

Out[5]:
```
array([[6.85384615, 3.07692308, 5.71538462, 2.05384615],
       [5.006     , 3.428     , 1.462     , 0.246     ],
       [5.88360656, 2.74098361, 4.38852459, 1.43442623]])
```

Reducing dataset with 2 features

In [6]:
```
pca = PCA(n_components=2)
x_reduced = pca.fit_transform(x)

x_reduced.shape
```

Out[6]:
```
(150, 2)
```

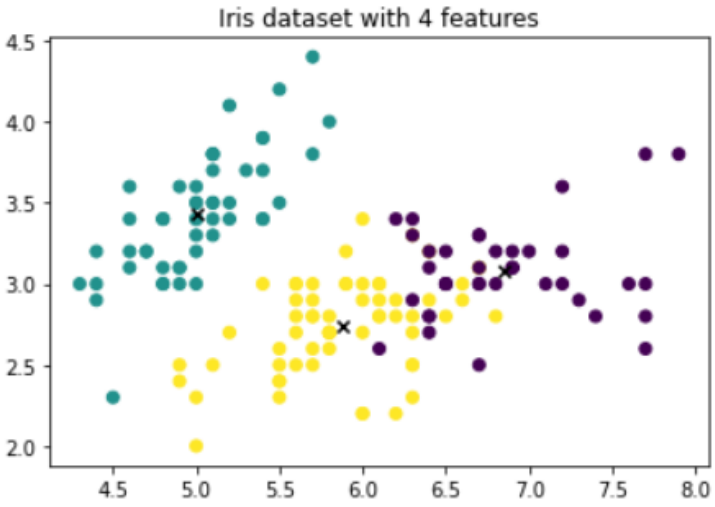
Training model on reduced dataset having 2 features

In [7]:
```
model_2 = KMeans(n_clusters=3, n_init=1, max_iter=100)
model_2.fit(x_reduced)

all_predictions_2 = model_2.predict(x_reduced)
centroids_2 = model_2.cluster_centers_
centroids_2
```

Out[7]:
```
array([[ 2.34652659,  0.27393856],
       [-2.64241546,  0.19088505],
       [ 0.66567601, -0.3316042 ]])
```

In [8]:
```
plt.scatter(x[:,0], x[:,1], c=all_predictions)
plt.scatter(centroids[:,0], centroids[:,1], marker='x', color="black")
plt.title("Iris dataset with 4 features")
plt.show()
```



In [9]:
```
plt.scatter(x_reduced[:,0], x_reduced[:,1], c=all_predictions_2)
plt.scatter(centroids_2[:,0], centroids_2[:,1], marker='x', color="black")
plt.title("Iris dataset with 2 features")
plt.show()
```

Results can be seen in graphs, we have got pretty good results with reduced dataset.
