Name: Osama Abdul Razzak (2303.KHI.DEG.029)
Peer Name: Rahima Siddiqui (2303.KHI.DEG.030)

Assignment 3.3

Import the required libraries

[51]:
```
import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.metrics import adjusted_rand_score
```

Load the Iris datasets

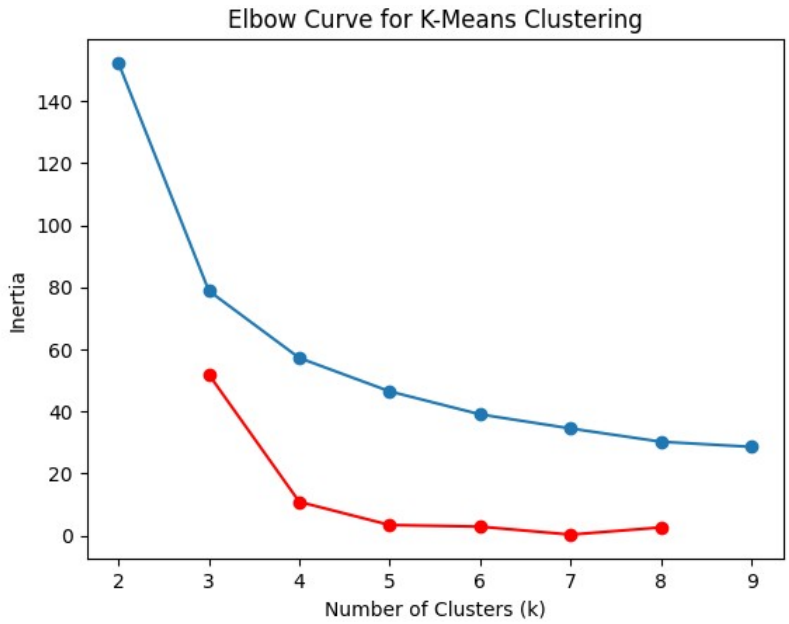
[10]:
```
# Load the Iris dataset
iris_data = datasets.load_iris()
x = iris_data.data
y = iris_data.target
```

Use Elbow method to find optimal number of clusters

[33]:
```
# Using Elbow method to determine the optimal number of clusters
k_values = []
inertia_scores = []

for k in range(2, 10):
    model = KMeans(n_clusters=k)
    model.fit(x)
    inertia_scores.append(model.inertia_)
    k_values.append(k)

module_of_second_derivative = np.abs(np.diff(np.diff(inertia_scores)))

# Plotting the Elbow curve
plt.plot(k_values, inertia_scores)
plt.scatter(k_values, inertia_scores)
plt.plot(k_values[1:-1], module_of_second_derivative, color='red')
plt.scatter(k_values[1:-1], module_of_second_derivative, color='red')
plt.xlabel("Number of Clusters (k)")
plt.ylabel("Inertia")
plt.title("Elbow Curve for K-Means Clustering")
plt.show()
```

```
C:\Users\LENOVO\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\LENOVO\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\LENOVO\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\LENOVO\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\LENOVO\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\LENOVO\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\LENOVO\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\LENOVO\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
```

Name: Osama Abdul Razzak (2303.KHI.DEG.029)
Peer Name: Rahima Siddiqui (2303.KHI.DEG.030)

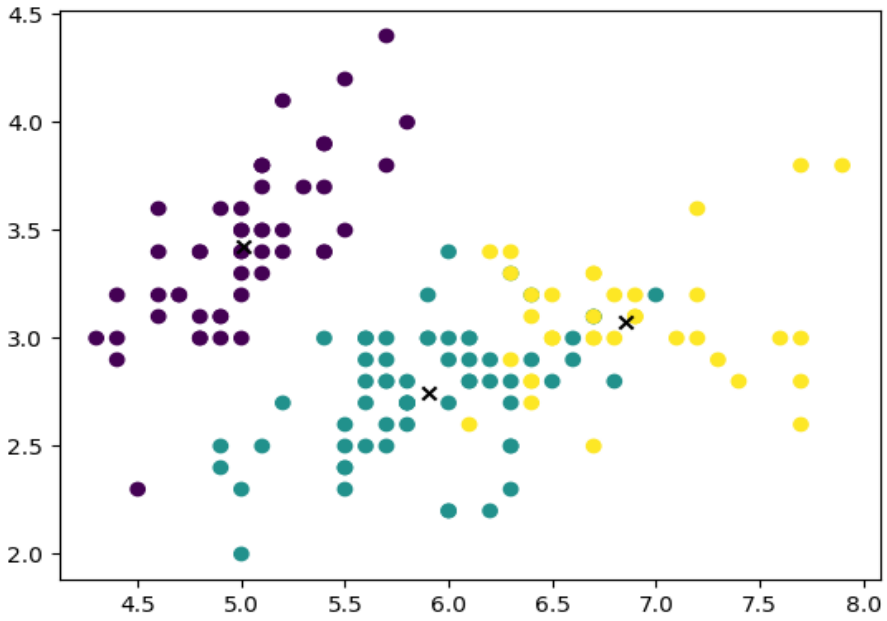


And from plot it is observed the sharp curves appear at k=3

Now, Applying KMean on original dataset

[72]:
```
#performing K-mean cluster on original dataset
model = KMeans(n_clusters=3 ,n_init=1,max_iter=100)
model.fit(x)
y_pred= model.predict(x)
centroids = model.cluster_centers_
plt.scatter(x[:,0], x[:,1], c= y_pred)
plt.scatter(centroids[:,0], centroids[:,1], marker = 'x', color = 'black')
plt.show
```

[72]: `<function matplotlib.pyplot.show(close=None, block=None)>`

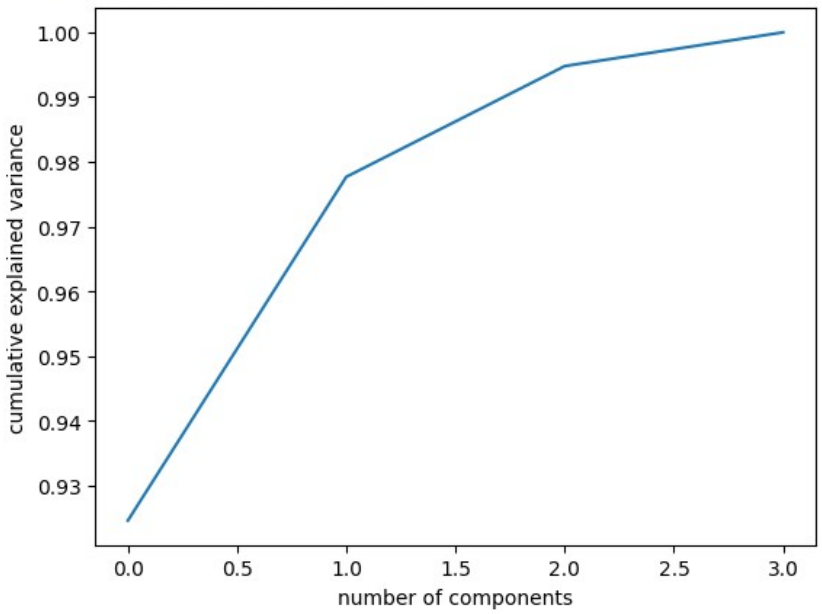


Name: Osama Abdul Razzak (2303.KHI.DEG.029)
Peer Name: Rahima Siddiqui (2303.KHI.DEG.030)

For performing PCA on original dataset, we have to choose the optimal number of components

```
[44]: pca = PCA().fit(x)
      plt.plot(np.cumsum(pca.explained_variance_ratio_))
      plt.xlabel('number of components')
      plt.ylabel('cumulative explained variance')
```

```
[44]: Text(0, 0.5, 'cumulative explained variance')
```



We can see that more than 99% of the variance is contained in the first 2 components

Applying PCA on Original dataset with two number of components

```
[48]: pca = PCA(n_components=2)
      x_reduced = pca.fit_transform(x)
      x_reduced.shape
```

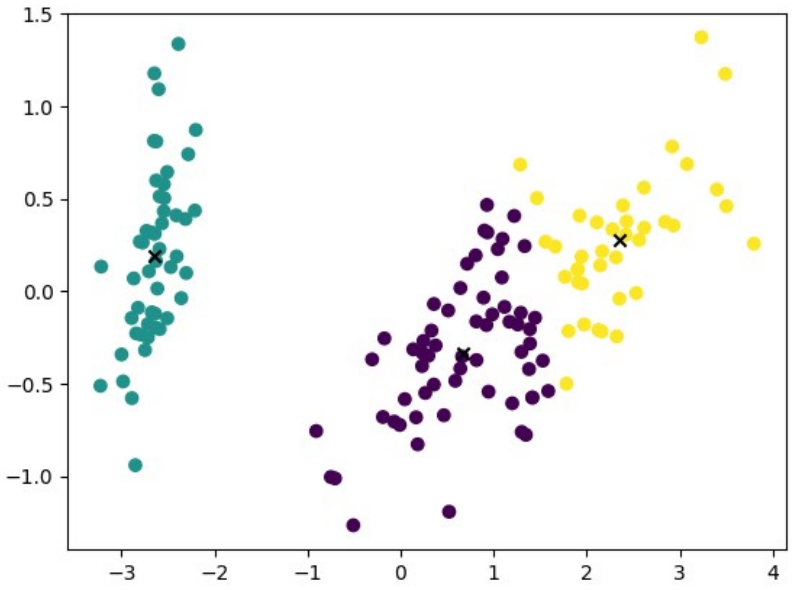
```
[48]: (150, 2)
```

Applying KMean on reduced dataset

Name: Osama Abdul Razzak (2303.KHI.DEG.029)
Peer Name: Rahima Siddiqui (2303.KHI.DEG.030)

Apply KMean on reduced dataset

[64]:
```
model_pca = KMeans(n_clusters=3, n_init =1, max_iter=100)
model_pca.fit(x_reduced)
y_pred_pca = model_pca.predict(x_reduced)
centroid_pca = model_pca.cluster_centers_
plt.scatter(x_red[:,0], x_red[:,1], c= y_pred_pca)
plt.scatter(centroid_pca[:,0],centroid_pca[:,1], marker='x',color = 'black')
plt.show()
```



Now comparing both plot before and after PCA using Adjusted_rand_score

[74]:
```
adj_rand_score = adjusted_rand_score(y_pred, y_pred_pca)
print(f"Adjusted Rand Score: {adj_rand_score:.4f}")
```

```
Adjusted Rand Score: 0.9803
```

The adjusted Rand score of 0.9803 indicates a high level of agreement between the cluster assignments obtained from K-means clustering on the original dataset and the PCA-reduced dataset