

K-MEANS:

K-Means, veri noktalarını kümelemek için denetimsiz bir öğrenme yöntemidir. Algoritma, her kümedeki varyansı en aza indirerek veri noktalarını yinelemeli olarak K kümelerine böler.

-Dirsek yöntemini kullanarak K için en iyi değeri nasıl tahmin edeceğinizi göstereceğiz

-Ardından veri noktalarını kümeler halinde gruplandırmak için K-ortalama kümelemeyi kullanacağız.

İlk olarak, her veri noktası rastgele K kümelerinden birine atanır. Ardından, her kümenin merkezini (işlevsel olarak merkezi) hesaplıyoruz ve her veri noktasını en yakın merkeze sahip kümeye yeniden atıyoruz. Her veri noktası için küme atamaları artık değişmeyene kadar bu işlemi tekrarlarız. K-ortalama kümeleme, verileri gruplandırmak istediğimiz küme sayısı olan K'yi seçmemizi gerektirir. Dirsek yöntemi, ataletin grafiğini (mesafeye dayalı bir ölçüm) çizmemize ve doğrusal olarak azalmaya başladığı noktayı görselleştirmemize olanak tanır. Bu nokta "elbow" olarak adlandırılır ve verilerimize dayanarak K için en iyi değer için iyi bir tahmindir.

CODE:

```
import sys

import matplotlib
matplotlib.use('Agg')

import matplotlib.pyplot as plt

x = [4, 5, 10, 4, 3, 11, 14, 6, 10, 12]

y = [21, 19, 24, 17, 16, 25, 24, 22, 21, 21]

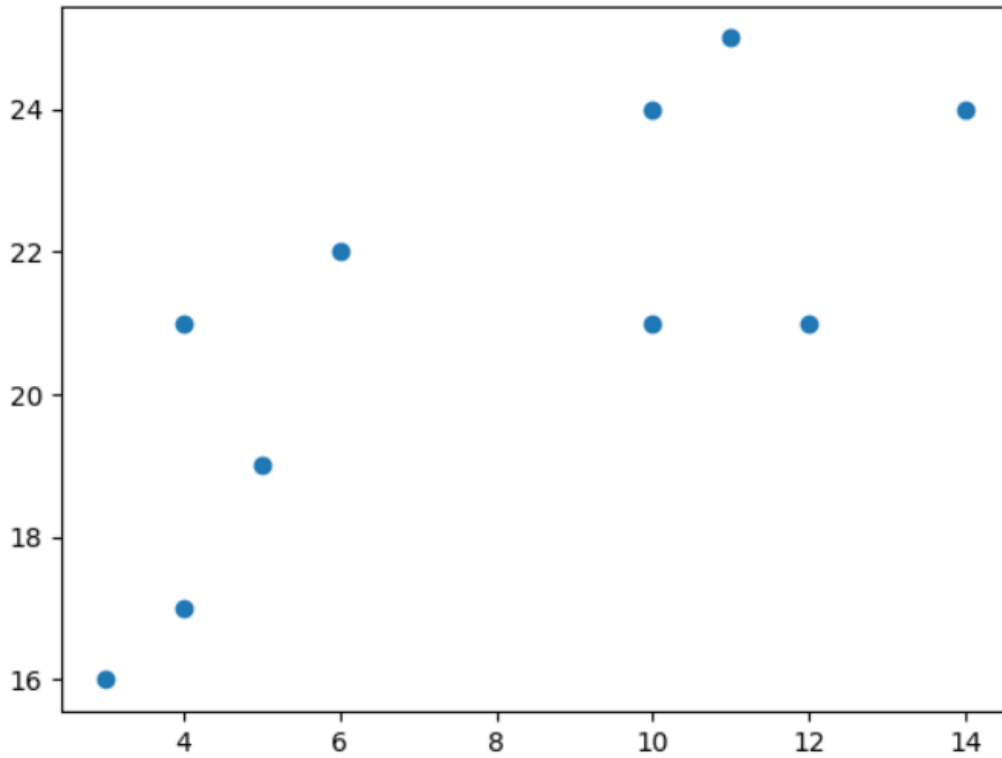
plt.scatter(x, y)

plt.show()

plt.savefig(sys.stdout.buffer)

sys.stdout.flush()
```

OUTPUT:



Şimdi farklı K değerlerinde görselleştirmek için dirsek yöntemini kullanıyoruz:

CODE:

```
import sys

import matplotlib
matplotlib.use('Agg')

import matplotlib.pyplot as plt from sklearn.cluster

import KMeans

x = [4, 5, 10, 4, 3, 11, 14, 6, 10, 12]

y = [21, 19, 24, 17, 16, 25, 24, 22, 21, 21]

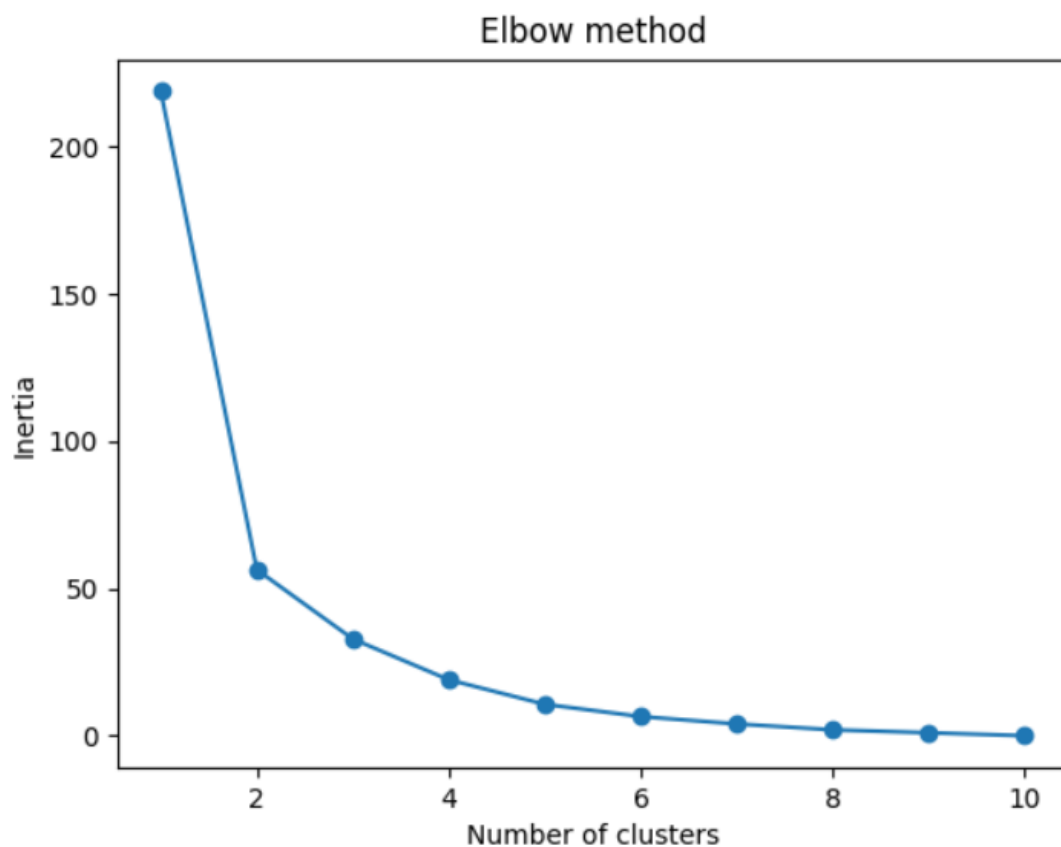
data = list(zip(x, y))

inertias = [ ]

for i in range(1,11):
```

```
kmeans = KMeans(n_clusters=i)
kmeans.fit(data)
inertias.append(kmeans.inertia_)
plt.plot(range(1,11), inertias, marker='o')
plt.title('Elbow method')
plt.xlabel('Number of clusters')
plt.ylabel('Inertia')
plt.show()
plt.savefig(sys.stdout.buffer)
sys.stdout.flush()
```

OUTPUT:



Dirsek yöntemi, 2'nin K için iyi bir değer olduğunu gösterir, bu nedenle sonucu yeniden eğitir ve görselleştiririz:

CODE:

```
import sys

import matplotlib
matplotlib.use('Agg')

import matplotlib.pyplot as plt from sklearn.cluster

import KMeans

x = [4, 5, 10, 4, 3, 11, 14, 6, 10, 12]

y = [21, 19, 24, 17, 16, 25, 24, 22, 21, 21]

data = list(zip(x, y))

kmeans = KMeans(n_clusters=2)

kmeans.fit(data)

plt.scatter(x, y, c=kmeans.labels_)

plt.show()

plt.savefig(sys.stdout.buffer)

sys.stdout.flush()
```

OUTPUT:

