

Makine öğrenimi alanında kümeleme algoritmaları, veri noktalarını benzerliklerine göre gruplara (kümeler) ayırmak için kullanılır. Bu gruplar, benzer özelliklere veya davranışlara sahip veri noktalarını içerir. Kümeleme algoritmaları, denetimsiz öğrenme kategorisine girer, çünkü etiketlenmiş veri kullanmadan veri setindeki doğal yapıları keşfederler.

Bazı popüler kümeleme algoritmaları şunlardır:

- 1- K-Means Kümeleme: K-means, belirli bir sayıda küme (K) oluşturmak için kullanılan en yaygın kümeleme algoritmalarından biridir. Başlangıçta, rastgele K merkez (küme merkezleri) seçilir ve her veri noktası en yakın merkeze atanır. Daha sonra, her veri noktası için atanmış merkezlerin ortalama konumu hesaplanır ve bu adımlar tekrarlanarak kümeleme süreci devam eder.
- 2- Hiyerarşik Kümeleme: Hiyerarşik kümeleme, aglomeratif (birleştirici) ve bölümleyici (bölücü) olmak üzere iki ana yaklaşıma sahiptir. Aglomeratif yöntemde, her veri noktası bir küme olarak başlar ve benzer kümeler birleştirilir. Bölümleyici yöntemde ise, tüm veri noktaları bir küme olarak başlar ve benzerliklere göre alt kümelere bölünürler.
- 3- DBSCAN (Density-Based Spatial Clustering of Applications with Noise): DBSCAN, veri setindeki yoğun alanları bulmak ve bu alanları kümelere dönüştürmek için kullanılan bir kümeleme algoritmasıdır. Bu algoritma, veri noktalarının yoğunluğuna dayanarak kümeleme yapar ve ayrık olmayan kümeleme problemlerinde iyi performans gösterir.

4-Gaussian Mixture Models (GMM): GMM, veri setinin olası bileşenlerini ve her bir bileşenin ağırlıklarını modelleyen bir olasılık tabanlı bir kümeleme algoritmasıdır. Veri, birbirinden farklı Gauss dağılımları ile temsil edilen bileşenlere ayrılır.

Bu algoritmaların her biri farklı veri yapıları ve kullanım senaryoları için uygun olabilir. Hangi kümeleme algoritmasının seçileceği, veri setinin özelliklerine, küme sayısına ve sonuçların beklenen formuna bağlı olarak değişir. Ayrıca, kümeleme sonuçlarının değerlendirilmesi ve yorumlanması da önemlidir, çünkü bu sonuçlar genellikle insanlar tarafından anlamlandırılmalı ve kullanılmalıdır.

Kümeleme kod parçası :

```
from sklearn.cluster import KMeans
import numpy as np
# Örnek veri oluşturma
X = np.array([[1, 2], [5, 8], [1.5, 1.8], [8, 8], [1, 0.6], [9, 11]])
# K-Means modeli oluşturma
kmeans = KMeans(n_clusters=2) # Küme sayısı belirlenir
# Veriyi modelimize uydurma
kmeans.fit(X)
# Küme merkezlerini ve etiketleri alma
centroids = kmeans.cluster_centers_
labels = kmeans.labels_
print("Küme merkezleri:")
print(centroids)
print("\nKüme etiketleri:")
print(labels)
# Sonuçları görselleştirme
import matplotlib.pyplot as plt
colors = ["g.", "r.", "c.", "y."]
for i in range(len(X)):
    plt.plot(X[i][0], X[i][1], colors[labels[i]], markersize=10)
plt.scatter(centroids[:, 0], centroids[:, 1], marker="x", s=150,
linewidths=5, zorder=10) , plt.show()
```