

## MAKİNE ÖĞRENMESİNDE KÜMELEME ALGORİTMALARI:

### 1. K-Means Kümeleme

Tanım:

K-Means kümeleme, verileri K sayıda küme olacak şekilde böler. Her küme, verilerin ortalamasına (merkeze) dayalı olarak tanımlanır ve her veri noktası en yakın merkez noktasına atanır.

Adımlar:

Küme sayısını K olarak belirle.

Rastgele K küme merkezi (centroid) seç.

Her veri noktasını, en yakın küme merkezine ata.

Her küme için yeni merkez noktalarını hesapla.

Adım 3 ve 4'ü merkezler değişmeyene kadar veya belirli bir iterasyon sayısına ulaşana kadar tekrarla.

Avantajlar: Basit ve hızlı, büyük veri kümeleriyle çalışabilir.

Dezavantajlar: Küme sayısını önceden belirlemek gerekir, başlangıç merkezlerine duyarlıdır, yalnızca küresel (spherical) kümeleri iyi işler.

Kullanım Alanı: Müşteri segmentasyonu, görüntü sıkıştırma, biyomedikal veri analizi.

### 2. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Tanım:

DBSCAN, yoğunluk tabanlı bir kümeleme algoritmasıdır. Yoğun bölgeler küme olarak kabul edilirken, yoğun olmayan bölgeler gürültü olarak değerlendirilir.

Adımlar:

Her veri noktası için belirli bir yarıçap içinde kaç komşusu olduğunu say (parametre: epsilon).

Komşu sayısı belirli bir eşik değerin (minPts) üzerinde olan noktaları çekirdek noktası (core point) olarak belirle.

Çekirdek noktaları birbirine bağlı noktaları içerir.

Gürültü noktalarını belirle ve bu noktaları kümelere dahil etme.

Özellikler:

Avantajlar: Gürültü ve sapmaları iyi yönetir, küme sayısını önceden belirtmeye gerek yoktur, kümeler herhangi bir şekle sahip olabilir.

Dezavantajlar: Yoğunluk parametreleri ( $\epsilon$  ve minPts) iyi seçilmelidir, büyük veri setlerinde hesaplama maliyeti yüksek olabilir.

Kullanım Alanı: Coğrafi veri analizi, anomalilerin tespiti, astronomi verisi kümeleme.

### 3. Mean Shift Kümeleme

Tanım:

Mean Shift, yoğunluk tabanlı bir kümeleme algoritmasıdır. Veri noktalarını yoğunluk modlarının (yoğunluk tepelerinin) etrafında iteratif olarak kaydırarak kümeleri belirler.

Adımlar:

Veri noktalarının yoğunluk bölgelerini hesapla.

Her veri noktasını, yoğunluk gradyanı boyunca en yakın yoğunluk tepesine doğru kaydır.

Veri noktaları yoğunluk tepelerinde toplanana kadar bu işlemi tekrarla.

Yoğunluk tepesinde toplanan veri noktaları kümeleri oluşturur.

Özellikler:

Avantajlar: Küme sayısı önceden belirlenmez, yoğunluk tepelerini bulur.

Dezavantajlar: Hesaplama maliyeti yüksektir, bant genişliği (bandwidth) parametresi iyi seçilmelidir.

Kullanım Alanı: Görüntü segmentasyonu, nesne izleme, radyo astronomi.

#### 4. Hiyerarşik Kümeleme (Hierarchical Clustering)

Tanım:

Hiyerarşik kümeleme, verileri hiyerarşik bir yapıda kümeleştirir. İki ana yaklaşımı vardır:

**Aglomeratif (Birleştirici) Yöntem:** Her veri noktası kendi kümesi olarak başlar ve en yakın kümeler iteratif olarak birleştirilir.

**Divizif (Bölücü) Yöntem:** Tüm veri noktaları tek bir kümede başlar ve iteratif olarak bölünür.

Adımlar (Agglomeratif Yöntem):

Her veri noktası kendi kümesi olarak başlar.

En yakın iki kümeyi birleştir.

Adım 2'yi tek bir küme kalana kadar tekrarla.

Kümeler arasındaki mesafeyi ölçmek için çeşitli ölçütler (örneğin, tek bağlantı, tam bağlantı, ortalama bağlantı) kullan.

Özellikler:

**Avantajlar:** Küme sayısını önceden belirlemeye gerek yoktur, dendrogram (ağaç diyagramı) kullanılarak görselleştirilebilir.

**Dezavantajlar:** Büyük veri setlerinde hesaplama maliyeti yüksektir, küme sayısını belirlemek zor olabilir.

**Kullanım Alanı:** Genetik veri analizi, belge sınıflandırma, müşteri segmentasyonu.