

Kümeleme (Clustering): Denetimsiz Öğrenmenin Temel Yöntemi

Giriş

Kümeleme, denetimsiz öğrenmenin (unsupervised learning) önemli bir yöntemidir ve veri analizinde yaygın olarak kullanılır. Denetimsiz öğrenme, etiketlenmemiş verilerden anlamlı bilgiler çıkarma sürecidir. Kümeleme, veriler arasındaki benzerlikleri ve farklılıkları belirleyerek verileri gruplar halinde organize eder. Bu yazıda, kümelemenin temellerini, kullanılan algoritmaları ve uygulama alanlarını inceleyeceğiz.

Denetimsiz Öğrenme ve Kümeleme

Denetimsiz öğrenme, veri kümesi hakkında önceden bilgi verilmeden, verilerin kendi içindeki ilişkiler ve yapıların keşfedilmesine dayanır. Kümeleme, bu sürecin bir parçası olarak, benzer özelliklere sahip veri noktalarını gruplara ayırarak daha anlamlı veriler oluşturur. Örneğin, bir alışveriş sitesinde kullanıcıların satın aldığı ürünlere göre diğer ürünlerin tavsiye edilmesi, kümeleme ile mümkündür.

Kümeleme Algoritmaları

Kümeleme algoritmaları, verileri belirli kriterlere göre gruplara ayırmak için çeşitli yaklaşımlar kullanır. İşte en yaygın kullanılan kümeleme algoritmalarından bazıları:

- K-Means Kümeleme:** Verileri k adet kümeye bölen bu algoritma, her küme için bir merkez (centroid) belirler ve veri noktalarını bu merkezlere olan uzaklıklarına göre kümelere ayırır. Algoritma, merkezlerin konumunu optimize ederek kümeleri günceller ve bu süreç, hata oranı minimize edilene kadar devam eder.
- Hiyerarşik Kümeleme:** Bu yöntem, veri noktalarını hiyerarşik bir yapı içinde organize eder. Aşağıdan yukarıya (agglomerative) veya yukarıdan aşağıya (divisive) yaklaşımlar kullanılabilir. Aşağıdan yukarıya yaklaşımda, her veri noktası başlangıçta tek bir küme olarak kabul edilir ve benzer kümeler birleşerek daha büyük kümeler oluşturur.
- DBSCAN (Gürültülü Uygulamaların Yoğunluğa Dayalı Konumsal Kümelenmesi):** Bu algoritma, yoğun bölgeleri küme olarak tanımlar ve düşük yoğunluklu bölgeleri gürültü olarak değerlendirir. DBSCAN, küme sayısını önceden belirlemeye gerek duymaz ve şekil ve boyut bakımından farklı kümeleri keşfetmede etkilidir.
- GMM (Gauss Karışım Modelleri) ile Beklenti-Maksimizasyon (EM) Kümeleme:** GMM, veri noktalarının Gauss dağılımına sahip olduğunu varsayar ve her küme için bir Gauss dağılımı belirler. EM algoritması, bu dağılımların parametrelerini optimize ederek kümeleri günceller.

Uygulama Alanları

Kümeleme, çeşitli alanlarda geniş uygulama yelpazesiyle kullanılır:

- **Tıp:** Görüntü analizinde farklı niteliklerin belirlenmesi.
- **Suç Yerlerinin Belirlenmesi:** Şehirlerdeki suç yoğunluklarının haritalanması.
- **Spor:** Oyuncu istatistiklerinin analizi ve performans değerlendirilmesi.
- **Telekom:** Çağrı kaydı detaylarının analizi ve müşteri segmentasyonu.
- **Pazarlama:** Müşteri segmentasyonu ve davranışsal analizler.

Sonuç

Kümeleme, denetimsiz öğrenmenin güçlü bir aracıdır ve çeşitli veri analizi uygulamalarında önemli rol oynar. K-Means, Hiyerarşik Kümeleme, DBSCAN ve GMM gibi algoritmalar, verilerin daha iyi anlaşılmasını ve daha verimli kullanılmasını sağlar. Kümeleme teknikleri, tıptan pazarlamaya kadar geniş bir yelpazede uygulanabilir ve gelecekte daha da önemli hale gelecektir.

KAYNAK

- [Unsupervised Learning: Clustering](#) (Kaggle)
- [Unsupervised Learning: 3-6 Clusters | K-Means | EDA](#) (Kaggle)
- [Unsupervised ML - Clustering \(Iris DataSet\)](#) (Kaggle)