

DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise), yoğunluk-tabanlı bir kümeleme algoritmasıdır. Gürültüyü (noise) tolere edebilir ve küme sayısını önceden belirlemeye gerek yoktur. DBSCAN algoritması, yoğun bölgeleri bulur ve bu yoğun bölgeleri küme olarak kabul ederken, düşük yoğunluklu bölgeleri gürültü olarak tanımlar.

Epsilon (ϵ): Epsilon, bir veri noktasının komşularını belirlemek için kullanılan maksimum mesafedir. Belirli bir noktadan ϵ mesafede olan diğer noktalar, bu noktanın komşuları olarak kabul edilir.

MinPts (Minimum Points): Bir veri noktasının merkezi (core) olarak kabul edilmesi için gereken minimum komşu sayısıdır. Bir veri noktası, en az MinPts sayısına sahip bir komşuluk içinde bulunmuyorsa, bu nokta yoğun olmayan bir bölgede (outlier) olarak kabul edilir.

Noktaların Komşularının Belirlenmesi:

Her veri noktası için, ϵ mesafesindeki komşuları belirlenir. Bu, veri noktalarının birbirine ne kadar yakın olduğunu belirler.

Core Noktaların Bulunması:

Bir veri noktasının, kendisi dahil olmak üzere en az MinPts sayısına sahip komşuları varsa, bu nokta bir core (merkez) nokta olarak kabul edilir. Core noktalar, bir kümenin çekirdeğini oluşturur.

Kümelerin Oluşturulması:

Core noktaları kullanarak, yoğun bölgelerin (clusters) kümelerini oluştururuz. Bu, bir core noktasından başlayarak, bu noktanın yoğunlukla bağlı (density-reachable) olduğu tüm noktaları bir küme olarak kabul etmek anlamına gelir.

Gürültü (Noise) Noktalarının Belirlenmesi:

Yoğun bölgelerin dışında kalan ve herhangi bir kümeye atanmayan noktalar, gürültü (noise) olarak kabul edilir.

Uygulama

Rastgele oluřturulan veri iin dbscan uygulaması

