

Hierarchical

Hierarchical Clustering (Hiyerarşik Kümeleme), bir veri kümesindeki benzer veri noktalarını gruplara ayırmak için kullanılan bir kümeleme algoritmasıdır. Bu algoritma, hiyerarşik bir ağaç yapısı oluşturur, bu nedenle "hiyerarşik" terimi kullanılır. Her bir veri noktası, bir küme veya kümeleme içinde tek bir eleman olarak başlar ve ardından birleştirme (bir araya getirme) adımlarıyla daha büyük kümeler oluşturulur. Bu birleştirme işlemleri, benzer veri noktalarını içeren kümeleme yapılarının oluşturulmasını sağlar.

Uzaklık Ölçüsü: Hiyerarşik kümeleme algoritmalarında, veri noktaları arasındaki benzerlik veya uzaklık ölçüsü belirlenmelidir. Bu ölçüm, genellikle Euclidean mesafe veya Pearson korelasyonu gibi bir metrik kullanılarak hesaplanır.

Birleştirme Stratejisi: Hiyerarşik kümeleme algoritmaları, birleştirme stratejisine bağlı olarak farklı türde kümeleme yapıları oluşturabilir. En yaygın birleştirme stratejileri şunlardır:

- **Single Linkage (En Yakın Komşu):** Her kümenin en yakın iki veri noktası arasındaki mesafeyi ölçer. İki küme arasındaki mesafe, bu iki veri noktası arasındaki mesafe olarak belirlenir.
- **Complete Linkage (En Uzak Komşu):** Her kümenin en uzak iki veri noktası arasındaki mesafeyi ölçer. İki küme arasındaki mesafe, bu iki veri noktası arasındaki mesafe olarak belirlenir.
- **Average Linkage (Ortalama Bağlantı):** Her iki küme arasındaki tüm veri noktaları arasındaki ortalama mesafeyi ölçer. İki küme arasındaki mesafe, bu ortalama mesafe olarak belirlenir.
- **Ward's Method:** Kümeleme işlemlerinde küme içindeki veri noktalarının kovaryans matrisine dayanarak, kümeleme işlemlerinde en az kare hatayı (varyansı) minimize eden iki kümenin birleştirilmesini sağlar.

Başlangıç Aşaması: Her bir veri noktası, tek elemanlı kümelere atanır.

Benzerlik veya Uzaklık Matrisi Oluşturma: Tüm veri noktaları arasındaki benzerlik veya uzaklık ölçüleri hesaplanır. Bu, veri noktaları arasındaki ilişkileri ifade eden bir matris oluşturur.

Birleştirme (Agglomerative) Aşaması: Başlangıçta, her veri noktası bir küme olarak kabul edilir. Daha sonra, her bir adımda, birleştirme stratejisine göre, birleştirilecek iki en yakın küme veya veri noktası bulunur. Bu işlem, bir hiyerarşik ağaç yapısı oluşturur.

Dendrogram Oluşturma: Son olarak, bu ağaç yapısı, dendrogram adı verilen ve kümeleme işlemlerini görselleştiren bir yapıya dönüştürülür.

Uygulama

Rastgele verinin birbirlerine yakınlığına göre hiyerarşik kümeleme

