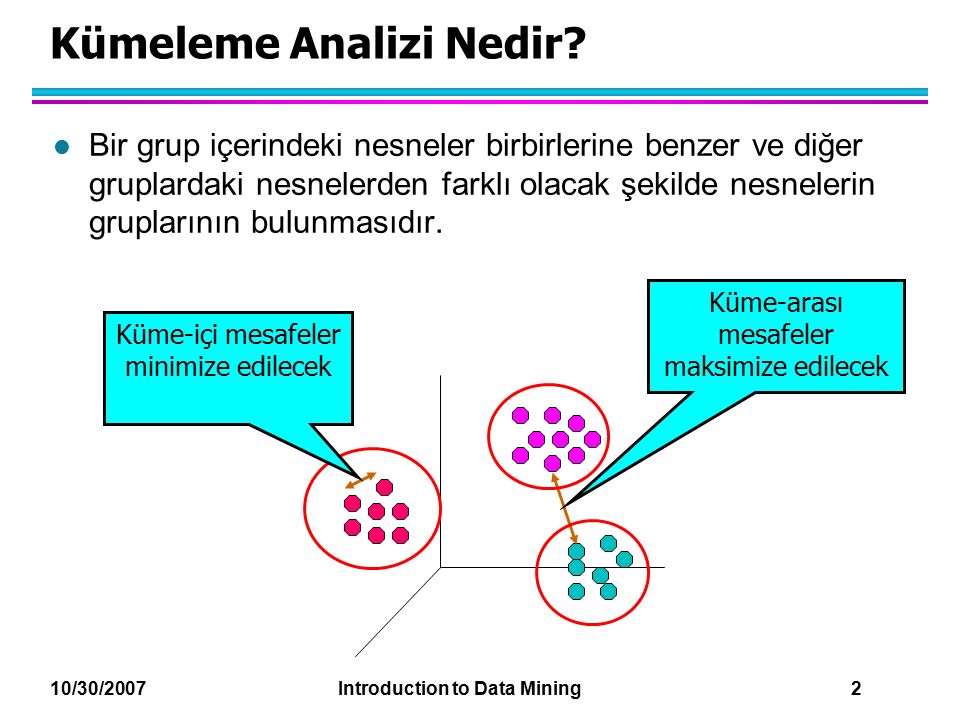
VERİ YAPILARI VE ALGORİTMALAR DERSİ 2. KISA SINAV 2. BÖLÜM ÖDEVİ

\* KÜMELEME ALGORİTMALARI\*

KÜMELEME VE KÜMELEME ANALİZİ NEDİR ?

Kümeleme en basit tanımıyla benzer özellik gösteren veri elemanlarının kendi aralarında gruplara ayrılmasıdır. Aynı küme içindeki elemanların benzerliği fazla, kümeler arası benzerlik ise az olmalıdır.

Kümeleme analizi, bir veri kümesindeki bilgileri belirli yakınlık kriterlerine göre gruplara ayırma işlemidir. Bu grupların her birine ‘küme’ denir. Kümeleme analizine işlemine ise kümeleme adı verilir



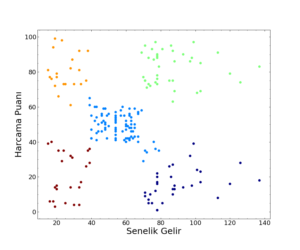
Kümeleme Yöntemleri Nelerdir?

Kümeleme analizi, veri noktalarını benzerliklerine göre gruplandırmak için kullanılan farklı yöntemleri içerir bunlardan bazıları;

1. K-Means Kümeleme: K-means, küme merkezlerini belirleyerek ve veri noktalarını bu merkezlere en yakın olanlara atayarak çalışan bir kümeleme yöntemidir. Kullanıcı tarafından belirlenen bir K değeri ile başlar ve iteratif olarak merkezleri optimize eder.
2. Hiyerarşik Kümeleme: Hiyerarşik kümeleme, veri noktalarını bir ağaç yapısı şeklinde gruplandırır. İki ana yaklaşım vardır: aglomeratif (birleştirici) ve bölücü (ayırıcı). Aglomeratif yöntemde, her veri noktası ayrı bir küme olarak başlar ve benzer kümeleme yöntemleri kullanılarak birleştirilir. Bölücü yöntemde ise tüm veri noktaları tek bir kümeye aitken, iteratif olarak ayrıştırılır.
3. DBSCAN (Yoğunluk Tabanlı Kümeleme): DBSCAN, veri noktalarını yoğunluk tabanlı olarak gruplandıran bir kümeleme yöntemidir. Bu yöntemde, yoğun veri bölgeleri bir küme oluştururken, düşük yoğunluklu bölgeler arasında ayrım yapar. DBSCAN, küme sayısını önceden belirtme ihtiyacı olmadan esnek bir şekilde çalışır.
4. Gaussian Mixture Models (GMM): GMM, veri noktalarını olasılık dağılımlarına dayalı olarak gruplandıran bir kümeleme yöntemidir. Bu yöntemde, her bir kümenin bir olasılık dağılımını temsil eden Gauss eğrileri kullanılır. Veri noktaları, bu Gauss eğrilerine göre en iyi şekilde tahmin edilir.
5. Fuzzy C-Means (Bulanık C-Means): Fuzzy C-means, bulanık kümeleme analizinde kullanılan bir yöntemdir. Her bir veri noktası, birden fazla kümeyle ilişkilendirilir ve üyelik derecesi kullanılarak bu ilişki ifade edilir. Veri noktaları, küme merkezlerine olan uzaklıkları ve üyelik dereceleri göz önünde bulundurularak gruplandırılır.(Bizim konu başlığıdır.)
6. SOM (Self-Organizing Map): SOM, bir yapay sinir ağı tabanlı bir kümeleme yöntemidir. Veri noktaları, ızgaraya yerleştirilmiş düğümler veya nöronlar tarafından temsil edilir. SOM, benzer veri noktalarını bir araya getirirken, topolojik bir yapıyı korur ve veriye özgü örüntüleri ortaya çıkarmak için kullanılır.

Kümeleme Analizinin Uygulama Alanları Nelerdir?

1. Pazar Segmentasyonu: Pazarlama alanında, kümeleme analizi müşteri segmentlerini belirlemek ve pazarı daha iyi anlamak için kullanılır. Benzer özelliklere sahip müşteriler kümelendirilerek, pazarlama stratejileri ve hedeflemeleri optimize edilebilir.
2. Müşteri Segmentasyonu: Perakende, e-ticaret ve hizmet sektörlerinde müşterileri farklı segmentlere ayırmak için kümeleme analizi kullanılır. Bu, müşteri davranışlarını ve tercihlerini anlamak, kişiselleştirilmiş pazarlama yapmak ve müşteri memnuniyetini artırmak için kullanılan bir tekniktir.
3. Görüntü İşleme: Görüntü işleme alanında, kümeleme analizi benzer pikselleri veya görüntü bölgelerini bir araya getirerek nesneleri veya desenleri tanımak için kullanılır. Örneğin, görüntü segmentasyonunda kümeleme analizi kullanılarak, bir görüntüdeki farklı nesneler veya bölümler ayrıştırılabilir.
4. Biyomedikal Veri Analizi: Biyomedikal araştırmalarda, hastalıkların tanısı, ilaç etkinliği veya genetik analiz gibi konular için kümeleme analizi kullanılır. Benzer özelliklere sahip hastalar veya genetik profillere sahip bireyler gruplandırılarak, daha iyi teşhisler yapılabilir veya tedavi stratejileri geliştirilebilir.
5. Hava Kalitesi Analizi: Hava kalitesi izleme istasyonlarından elde edilen verilerin analizi için kümeleme analizi kullanılır. Benzer hava kalitesi ölçümlerine sahip bölgeler gruplandırılarak, hava kirliliği kaynakları ve etkileri daha iyi anlaşılabilir.
6. Sosyal Ağ Analizi: Sosyal medya veya iletişim ağlarındaki kullanıcıları veya grupları kümelemek için kümeleme analizi kullanılır. Bu, benzer ilgi alanlarına sahip kullanıcıları bir araya getirmek ve sosyal ağlardaki etkileşimleri veya davranışları anlamak için kullanılan bir tekniktir.
7. Veri Keşfi ve Desen Tanıma: Veri madenciliği alanında kümeleme analizi, büyük veri kümelerinde gizli yapıları veya desenleri ortaya çıkarmak için kullanılır. Benzerliklere dayalı olarak veri noktalarını gruplandırmak, veri keşfi sürecinde önemli bir adımdır.



Hiyerarşik kümeleme örneği.

Bulanık Kümeleme Analizi Nedir?

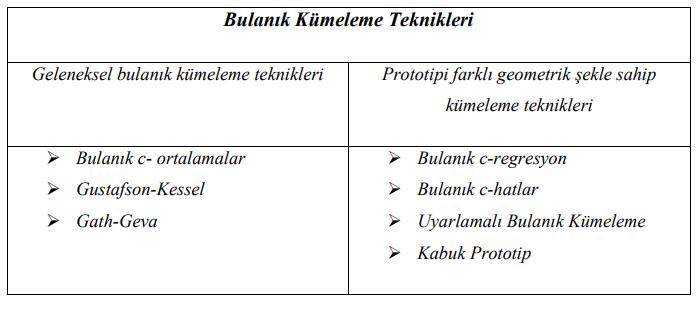
Kümeleme analizi kümelere olan üyelik kayıtlarına göre katı, bulanık ve olasılıklı olmak üzere 3 farklı açıdan incelenebilinir. Katı kümelemede kümeye üyelik kaydı booleca (boolean)’dır .Yani gözlemler ya bir kümeye üyedirler ya da değildirler. Bulanık kümelemede, 0 ile 1 arasında değişen bir üyelik kaydı ve verilerin aynı anda birden fazla kümeye üyeliği söz konusu olmaktadır. Olasılıklı kümeleme analizinde ise üyelik yine boolecadır ama bu yöntemde kümelere atanmada bir olasılık dağılımı vardır .

Bulanık kümeleme analizi (fuzzy clustering analysis), veri kümesindeki benzerlikleri ve farklılıkları belirlemek için kullanılan bir veri analizi yöntemidir. Geleneksel kümeleme analizi yöntemlerinden farklı olarak, verileri sadece kesin kümeler halinde değil, bulanık (fuzzy) kümeler halinde gruplandırır.

Bulanık kümeleme analizi, her veri noktasını birden fazla kümeyle ilişkilendirebilir ve bu ilişkiyi bir belirsizlik derecesi veya olasılık değeriyle ifade eder. Bu, her veri noktasının farklı kümelerdeki üyelik derecesini belirlemek için kullanılır. Örneğin, bir veri noktası %80 olasılıkla bir küme içinde yer alırken, aynı zamanda %20 olasılıkla başka bir kümeye de ait olabilir.

Bulanık kümeleme analizi genellikle, veri kümesindeki içerik, özellik veya örüntülerin belirsiz olduğu durumlarda kullanılır. Bu analiz yöntemi, verilerdeki gürültüyü, eksik bilgiyi veya belirsizlikleri dikkate alarak daha esnek bir gruplandırma sağlar.

Bulanık Kümeleme Teknikleri:



Bulanık Kümeleme Analizi Nerede Kullanılır?

Bulanık kümeleme analizi, makine öğrenimi, veri madenciliği, desen tanıma ve karar destek sistemleri gibi birçok alanda kullanılır. Özellikle, pazar segmentasyonu, görüntü işleme, biyomedikal veri analizi ve hava kalitesi analizi gibi uygulamalarda sıklıkla kullanılır.

Nasıl Hesaplanabilir?

Bulanık kümeleme analizi, küme merkezlerini belirlemek ve veri noktalarının her bir kümeye olan üyelik derecelerini hesaplamak için bazı formüller kullanır. İşte bulanık kümeleme analizinde yaygın olarak kullanılan bazı formüller:

1. Küme Merkezi (Centroid) Hesaplama:

* Veri noktaları xᵢ, i = 1, 2, ..., n
* Kümelerin merkezi vᵢ, i = 1, 2, ..., c
* Üyelik derecesi uᵢj, i = 1, 2, ..., n; j = 1, 2, ..., c (veri noktasının i. kümedeki üyelik derecesi)

Küme merkezini hesaplamak için aşağıdaki formül kullanılabilir:

vᵢ = (Σ(uᵢj^m \* xᵢ) / Σ(uᵢj^m)), for i = 1, 2, ..., c

NOT:

Burada, m bir bulanıklık parametresidir (genellikle 2'den büyük bir değer alır) ve Σ sembolü toplam anlamına gelir.

1. Üyelik Derecesi Hesaplama:

* Veri noktası xᵢ
* Kümelerin merkezi vᵢ, i = 1, 2, ..., c
* Üyelik derecesi uᵢj, i = 1, 2, ..., n; j = 1, 2, ..., c (veri noktasının i. kümedeki üyelik derecesi)

Üyelik derecesini hesaplamak için aşağıdaki formül kullanılabilir:

uᵢj = (1 / Σ((||xᵢ - vᵢ|| / ||xᵢ - vⱼ||)^(2 / (m-1)))), for i = 1, 2, ..., n; j = 1, 2, ..., c

NOT:

Burada, ||xᵢ - vᵢ||, Euclidean mesafesini ifade eder ve ^ sembolü üssü gösterir.

Bu formüller, bulanık kümeleme analizinde kullanılan temel hesaplamaları temsil etmektedir. Gerçek uygulamalarda, bu formüller çeşitli algoritmaların içinde kullanılarak iteratif bir şekilde küme merkezleri ve üyelik dereceleri güncellenir.

BULANIK KÜMELEME ANALİZİ İLE TÜRKİYE’DEKİ İLLERİN SOSYOEKONOMİK AÇIDAN SINIFLANDIRILMASI ÖRNEĞİmetin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu :

Bulanık Analiz Kümeleme Örnek Kodu :

metin, ekran görüntüsü, yazılım, bilgisayar içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, yazılım, bilgisayar, web sayfası içeren bir resim

Açıklama otomatik olarak oluşturuldu

Yukarıda verilen kod, su markalarını belirli özelliklerine göre bulanık kümeleme ile sınıflandıran bir C programıdır.

Programın çalışma mantığı şu şekildedir :

1. ‘’SuMarkasi’’ struct yapısı, su markalarının özelliklerini (isim, kalite, tat, fiyat) tutmak için kullanılır.
2. ‘’siniflandirma’’ fonksiyonu, su markalarını bulanık kümeleme ile sınıflandırır. İlk olarak, ‘’bulaniklikDegerleri’’ adında bir dizi oluşturulur ve her bir marka için kümeler arasındaki bulanıklık değerleri hesaplanır.
3. Bulanıklık değerleri hesaplanırken, su markasının kalite, tat ve fiyat uyelik değerleri kullanılır. Örneğin, üçgen bulanıklaştırma yöntemi kullanılarak, her bir kombinasyon için bulanıklık değerleri hesaplanır.
4. Sınıflandırma işlemi, her bir su markası için en yüksek bulanıklık değerine sahip olan küme tarafından yapılır.
5. Program, her bir su markasının hangi kümeye ait olduğunu ekrana yazdırır.
6. Kullanıcıdan su markası sayısı, su markalarının özellikleri ve küme sayısı alınır. Bu bilgiler kullanılarak tekrar ‘’siniflandirma’’ fonksiyonu çağrılır.
7. Bu C programı, su markalarını bulanık kümeleme kullanarak sınıflandırmaktadır.

KAYNAKÇA:

**https://prezi.com/ml-nnxafkeky/kumeleme-nedir/**

**http://www.papatyabilim.com.tr/PDF/Biyoinformatik-Veri-Analizinde-R-ile-Hiyerarsik-Kumeleme-PDF-icindekiler.pdf**

**https://dergipark.org.tr/tr/download/article-file/194560**

**https://earsiv.anadolu.edu.tr/xmlui/bitstream/handle/11421/5572/469334.pdf?sequence=1&isAllowed=y**

**https://dergipark.org.tr/tr/download/article-file/194568**

HAZIRLAYANLAR:

AD/SOYAD: ÖĞRENCİ NO:

REYYA TACEMEN 5210505034

KEREM YALIN TAŞKIN 5210505028

ARDA IŞIK 5210505058