



AFYON KOCATEPE ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

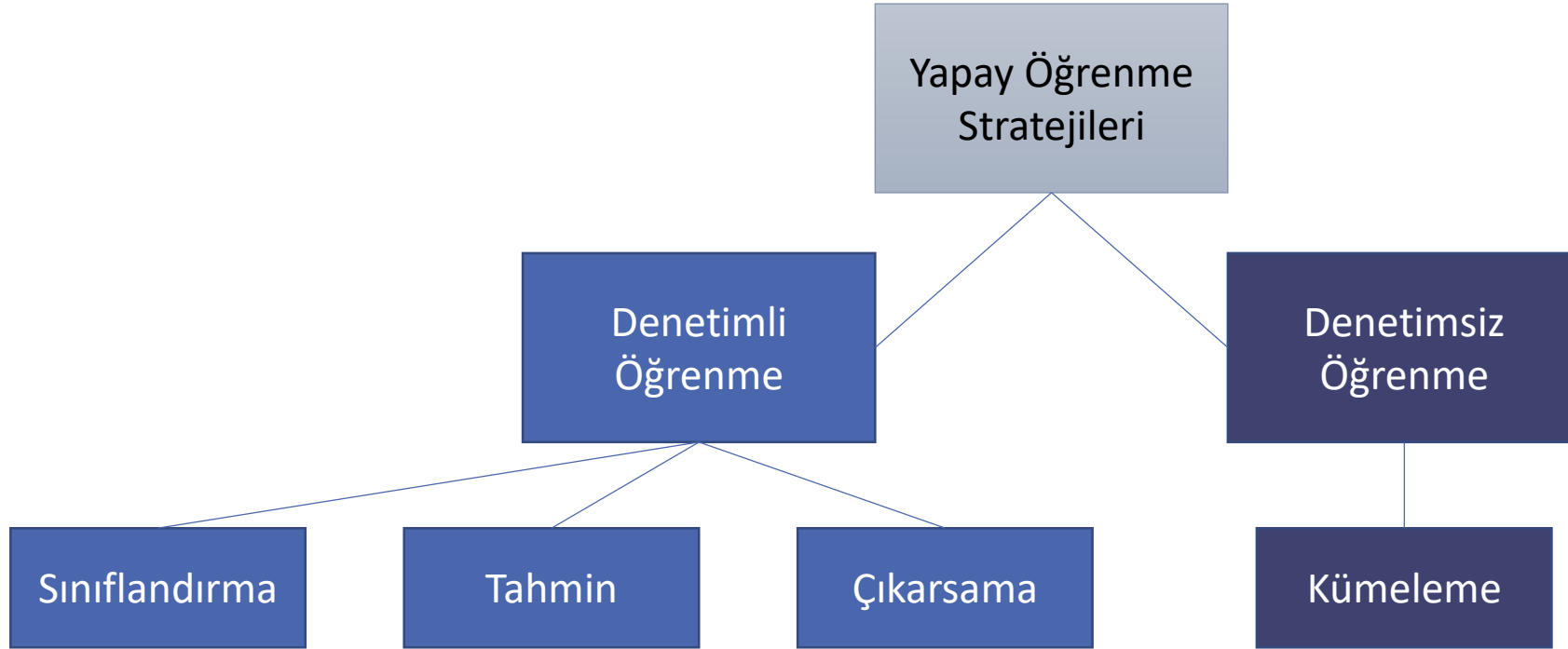
BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

- KÜMELEME TEMEL PRENSİPLER



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ





BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

- Kümeleme bir denetimsiz öğrenme metodudur.
- Kümeleme, sınıf bilgisi içermeyen veri içerisinde benzerliği yüksek verileri grupta işlemidir.
- Bir veri seti içinde verilerin öznitelikleri çıkartılarak birbirine en çok benzeyen veriler aynı gruba atanacak şekilde bir dağılım yapılır.
- Farklı gruptaki verilerin benzerliklerinin an az olması sağlanmaya çalışılır.

Yani:

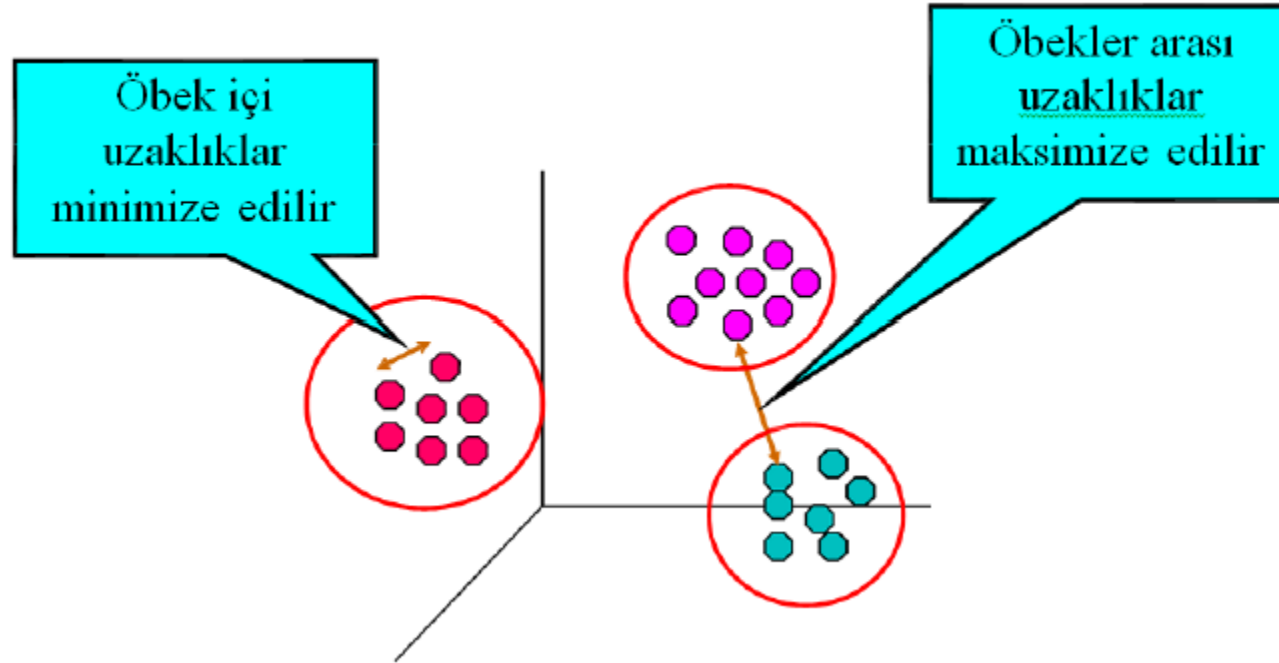
Aynı kümedeki örnekler birbirine daha çok benzer, Farklı kümedeki örnekler birbirine daha az benzer



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

Kümelemenin amacı; aşağıdaki özellikleri sağlayan kümeleri bulmaktır.





BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ÖRNEKLERİ

1. Müşteri Segmentasyonu

Örnek: Bir e-ticaret sitesi, müşteri verilerini kullanarak farklı müşteri gruplarını belirlemek istiyor. Bu gruplar, benzer alışveriş alışkanlıklarına, harcama miktarlarına veya ziyaret sıklıklarına sahip müşterilerden oluşabilir. Kümeleme algoritmaları sayesinde, benzer özelliklere sahip müşteriler gruplandırılabilir ve her bir grup için özel pazarlama stratejileri geliştirilebilir.

2. Belge Kümeleme

Örnek: Bir haber ajansı, güncel haber makalelerini konularına göre gruplandırmak istiyor. Kümeleme algoritmaları, benzer içerik ve anahtar kelimelere sahip haber makalelerini bir araya getirerek, kullanıcılara ilgili haberleri kolayca sunabilir.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ÖRNEKLERİ

3. Sosyal Ağ Analizi

Örnek: Bir sosyal medya platformu, kullanıcılarını ilgi alanlarına göre gruplandırmak istiyor. Kümeleme algoritmaları, kullanıcıların beğenileri, paylaşımları ve arkadaşlık bağlantılarını analiz ederek, benzer ilgi alanlarına sahip kullanıcıları belirleyebilir ve bu sayede daha kişiselleştirilmiş içerik sunabilir.

4. Biyoinformatik

Örnek: Bir araştırma laboratuvarı, gen ekspresyon verilerini analiz ederek, benzer genetik profillere sahip bireyleri gruplandırmak istiyor. Kümeleme algoritmaları, benzer genetik özelliklere sahip bireyleri belirleyerek, hastalıkların genetik temellerini anlamada yardımcı olabilir.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ÖRNEKLERİ

5. Makine Parçalarının Aşınma Analizi

Örnek Durum:

Bir üretim tesisi, çeşitli makinelerde kullanılan parçaların aşınma ve yıpranma durumlarını izlemek ve analiz etmek istiyor. Makine mühendisleri, farklı parçaların aşınma profillerini anlamak ve bakım programlarını optimize etmek için bu verileri kullanabilir. Kümeleme algoritmaları, benzer aşınma profillerine sahip makine parçalarını gruplandırmak için kullanılabilir.

Örnek Sonuç:

Grup 1: Yüksek sıcaklık ve basınca maruz kalan parçalar, belirli bir süre sonunda ciddi aşınma gösteriyor.

Bu grup için sık bakım gereklidir.

Grup 2: Düşük sıcaklık ve titreşim koşullarında çalışan parçalar, daha uzun süre dayanıyor. Bu grup için daha seyrek bakım yeterlidir.

Grup 3: Anormal aşınma özelliği gösteren parçalar, belki de üretim hatası veya malzeme kalitesi sorunlarına işaret edebilir. Bu parçalar daha detaylı incelenmelidir.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ÖRNEKLERİ

6. Elektrik Şebekesi Anomali Tespiti ve Yük Profili Analizi

Örnek Durum:

Bir elektrik dağıtım şirketi, farklı bölgelerdeki elektrik tüketim verilerini analiz ederek anormal tüketim davranışlarını tespit etmek ve yük profillerini sınıflandırmak istiyor. Bu analiz, elektrik dağıtımının daha verimli yönetilmesi ve olası sorunların önceden tespit edilmesi için önemlidir. Kümeleme algoritmaları, benzer elektrik tüketim profillerine sahip bölgeleri veya tüketicileri gruplandırmak için kullanılabilir.

Örnek Sonuç:

Grup 1: Konut bölgeleri, genellikle sabah ve akşam saatlerinde yüksek tüketim gösteriyor. Bu grup için pik saatlerde yeterli enerji sağlanmalıdır.

Grup 2: Ticari bölgeler, gün içinde sabit ve yüksek tüketim gösteriyor. Bu grup için kesintisiz enerji sağlanması önemli.

Grup 3: Endüstriyel tesisler, gece ve hafta sonu dahil olmak üzere sürekli yüksek tüketim gösteriyor. Bu grup için sürekli enerji akışı ve yedekleme sistemleri kritik.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

Diğer örnekler:

- Twitter ya da LinkedIn profillerini kullanarak aynı meslek grubunda olan hesapları veya aynı hobilere sahip kişileri gruplayarak arkadaş önerilerinde bulunmak amacıyla kullanılabilir.
- Binlerce öğrencinin kişisel ve demografik verilerini kullanarak burs almaya hak kazanacak öğrenci kümesinin tespit edilmesi,
- Marketlerde farklı müşteri gruplarının keşfedilmesi ve bu grupların alışveriş örüntülerinin ortaya konması,
- İşlevlerine göre benzer genlerin sınıflandırılması,
- Dokümanların içeriklerine göre kümelenmesi



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

Oluşturulan kümelerin kalitesi seçilen **benzerlik ölçütüne** ve bu ölçütün ne kadar gerçekleştirilebildiğine bağlıdır. Aynı kümedeki veriler arası benzerliği enbüyüten, farklı kümelerdeki veriler arası benzerliği enküçülten model aranmalıdır.

Veriyi gruplama için uygun kümeleme ölçütleri bulunmalıdır.

Benzerlik ölçütü olarak çoğunlukla veriler arasındaki uzamsal uzaklık kullanılır. Uzaklığın da çeşitli hesaplama yöntemleri vardır.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

Uzaklık için en çok kullanılan ölçüt Öklit (Euclidean distance) uzaklığıdır. İki nokta (n boyutlu) arasındaki uzaklık şöyle hesaplanır:

$$Uzaklık_{a,b} = \|a - b\| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

Benzerlik ise genellikle uzaklığın tersidir.

$$Benzerlik_{a,b} = \frac{1}{1 + \|a - b\|}$$



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

Fakat uzaklık hesabı için Öklit dışında önerilmiş Mahalanobis, Manhattan ve Chebyshev gibi farklı birçok ölçüt bulunmaktadır.

Manhattan ya da City block uzaklığı:

$$Uzaklık(i,j) = \left(|x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}| \right)$$



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

Minkowski Uzaklığı:

Öklid ve Manhattan uzaklığının genelleştirilmiş bir halidir.

$q=1$ için Manhattan uzaklığını, $q=2$ için Öklid uzaklığını verir. q sayısı büyüdükçe hassasiyet artacaktır.

$$Uzaklık(i,j) = \left(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots + |x_{ip} - x_{jp}|^q \right)^{1/q}$$

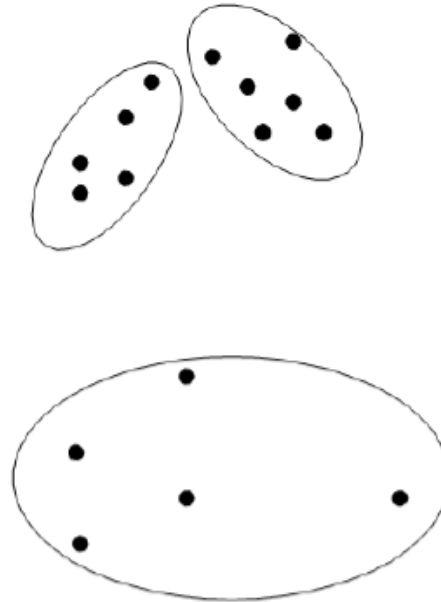


BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

Kümeleme Tipleri:

- **Bölümlenmeli kümeleme** :Veri nesnelerinin, birbirini kapsamayan alt kümelere ayrılmasıdır. Her bir veri nesnesi alt kümelerden sadece birinde yer alır.



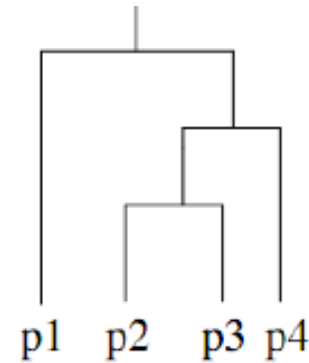
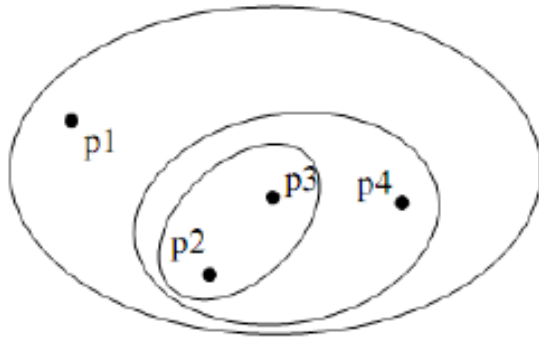


BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

Hiyerarşik kümeleme: veri noktalarını hiyerarşik bir yapı içinde gruplandırılan bir kümeleme yöntemidir. İki temel türü vardır:

- Birleştirici (Agglomerative) Hiyerarşik Kümeleme:
- Bölücü (Divisive) Hiyerarşik Kümeleme:





BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

- **Yoğunluk Tabanlı Kümeleme:**

Yoğunluk tabanlı yöntemler, nesnelerin doğal dağılımını bir yoğunluk fonksiyonu aracılığı ile tespit ederek bir eşik yoğunluğunu aşan bölgeleri küme olarak adlandırır. Düzgün şekilli olmayan kümeleri bulma başarısı, gürültü ve istisnalardan etkilenmeme ve tek tarama ile sonuca ulaşma avantajları ile en başarılı kümeleme yöntemleri arasındadır [5].

Örnek olarak DBSCAN (Density-Based Spatial Clustering of Applications with Noise) yöntemi verilebilir.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

- **Model Tabanlı Yöntemler**

Eldeki verileri bir matematiksel model ile ifade etmeye çalışırlar. Model tabanlı yöntemler iki temel yaklaşımı kullanırlar; istatistiksel yaklaşım ve yapay zekâ yaklaşımı [5].



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

VERİ TIPLERİ

Veri tipleri basitçe 3'e ayrılabilir.

1. Kategorik veri
2. Ordinal/sıralı veri
3. Sayısal/Numerik/Continuous veri



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

VERİ TIPLERİ

1. Kategorik veri

	A	B	C	D	E	F	G
1		Kategori 1	Kategori 2	Kategori 3	Kategori 4	Kategori X	
2	Cinsiyet	Erkek	Kadın				
3	ASA skoru	ASA 1	ASA 2	ASA 3	ASA 4		
4	Tm evresi	Evre 1	Evre 2	Evre 3	Evre 4		
5	Elma	Kırmızı elma	Yeşil elma	Beyaz elma			
6	Elma 2	Ekşi elma	Tatlı elma				
7	Kan grubu	A	B	AB	O		
8	DM öyküsü	Var	Yok				
9	Cerrahi öyküsü	Yes	No				
10							

Kategorik veri için Nominal veri ismi de kullanılabilir ve bazı alt isimlerle anılabilir.

Örneğin; kategori sayısı İKİ olduğunda (Cinsiyet, Elma 2, DM öyküsü, cerrahi öyküsü örneği gibi); BINARY (ikili) veri ismi alır. Çoğunlukla kullanılan ikili veri YES/NO cevabı olan değişkenlerdir.

Yani; iki kategori varsa ikili (binary), daha fazla varsa nominal, ya da hepsi için kategorik veri ismi kullanılabilir.

[https://www.aliguner.com/veri-tipi/#:~:text=1.-,Kategorik%20veri,ismi%20al%C4%B1r%20\(dichotomous%20variable\).](https://www.aliguner.com/veri-tipi/#:~:text=1.-,Kategorik%20veri,ismi%20al%C4%B1r%20(dichotomous%20variable).)



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

VERİ TIPLERİ

2. Ordinal (Sıralı) Veri

Bazı kaynaklarda ayrı bir sınıf olarak tanımlansa da aslında kategorik verinin bir alt tipi olarak düşünülebilir.

Farkı şudur; Ordinal veride kategoriler arasında bir sıralama vardır. Bu tip veriler de Nominal değişkenlerde olduğu gibi sonlu sayıda farklı durum içerirler fakat Ordinal veriler anlamlı bir sıralama takip ederler. Sıralamada daha üstte olan değişken bir altta olan değişkenden daha değerlidir. Örn: eğitim durumu verileri

ASA skoru ya da Tm evresi örneklerindeki gibi;

ASA1->ASA2->ASA3->ASA4 ya da Evre1->Evre2->Evre3->Evre4 gibi sıralı kategorilerden oluşur.

Cinsiyet ya da Kan grubu örneklerinde ise bir sıra yoktur. Rastgeledirler.

[https://www.aliguner.com/veri-tipi/#:~:text=1.-,Kategorik%20veri,ismi%20al%C4%B1r%20\(dichotomous%20variable\).](https://www.aliguner.com/veri-tipi/#:~:text=1.-,Kategorik%20veri,ismi%20al%C4%B1r%20(dichotomous%20variable).)



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

VERİ TIPLERİ

3. Sayısal/Numerik/Continuous veri

Continuous veri olarak bilinen bir ailedir ancak SAYISAL veri olarak da ifade edilir.

Yaş: 34,

Ağırlık: 78 kg

Boy: 182 cm

BMI: 26.8

Ameliyat süresi: 128 dakika

gibi örnekler verilebilir.

[https://www.aliguner.com/veri-tipi/#:~:text=1.-,Kategorik%20veri,ismi%20al%C4%B1r%20\(dichotomous%20variable\).](https://www.aliguner.com/veri-tipi/#:~:text=1.-,Kategorik%20veri,ismi%20al%C4%B1r%20(dichotomous%20variable).)



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

VERİ TIPLERİ

Veri tipleri arası dönüşüm

Bir sayısal veri kategorik veriye çevrilebilir. Örneğin;

Yaş verisi sayısal olarak kullanılabilir. 23,34,35,46,56,87,76,54,75,29,..... gibi.

Ya da kategorik veriye dönüştürülebilir: 50 yaş altı ve üstü gibi.

Bu durumda YES/NO içeren bir binary-nominal veri olacaktır.

Ya da 45 yaş altı/45-60 arası/60 üzeri gibi üçlü kategori de yapılabilir.

[https://www.aliguner.com/veri-tipi/#:~:text=1.-,Kategorik%20veri,ismi%20al%C4%B1r%20\(dichotomous%20variable\).](https://www.aliguner.com/veri-tipi/#:~:text=1.-,Kategorik%20veri,ismi%20al%C4%B1r%20(dichotomous%20variable).)



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

Bu ders içeriğinde Bölümlenmeli Kümeleme Analizi yapılacaktır.

Yöntem olarak da

- Bir adet matematiksel modele dayalı yöntem (K-Means),
- Bir adet de yapay öğrenmeye dayalı yöntem (Fuzzy C-Means)

Yöntemleri kullanılacaktır.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

K-MEANS YÖNTEMİ

k -means algoritmasının genel mantığı «n» adet veri nesnesinden oluşan bir veri setini, giriş parametresi olarak verilen «k» adet kümeye bölümlenektir. Amaç, gerçekleştirilen bölümlenme işlemi sonunda elde edilen kümelerin, küme içi benzerliklerinin maksimum ve kümeler arası benzerliklerinin minimum olmasını sağlamaktır.

Küme benzerliği, kümenin ağırlık merkezi olarak kabul edilen bir nesne ile kümedeki diğer veriler arasındaki uzaklıkların ortalama değeri ile ölçülmektedir [6-7].



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

K-MEANS YÖNTEMİ

K-Means Yöntemi:

Algoritma aşağıdaki şekildedir.

1. Küme sayısı (k) okunur. Bu değer algoritmaya dışarıdan verilir.
2. k adet rastgele veya belirli bir yöntemle küme merkezleri belirlenir.
3. Tüm elemanların merkezlere olan uzaklıkları hesaplanır.
4. Elemanlar yakın oldukları merkezlere göre kümelendir.
5. Dördüncü adımda oluşan kümelerin ortalamaları hesaplanarak yeni küme merkezleri belirlenir.
6. Son bulunan küme merkezleri bir önceki küme merkezlerine eşit oluncaya ya da belirlenen döngü sayısına ulaşılan kadar işlem tekrarlanır.

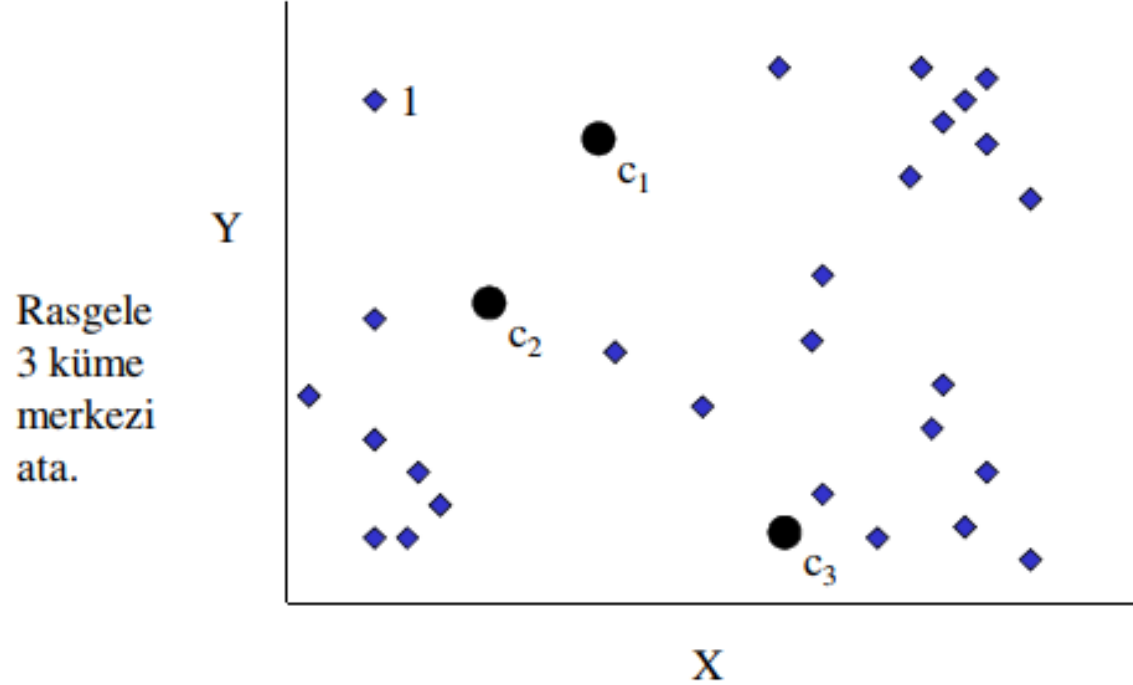


BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

K-MEANS YÖNTEMİ

K-means örnek adım 1



YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

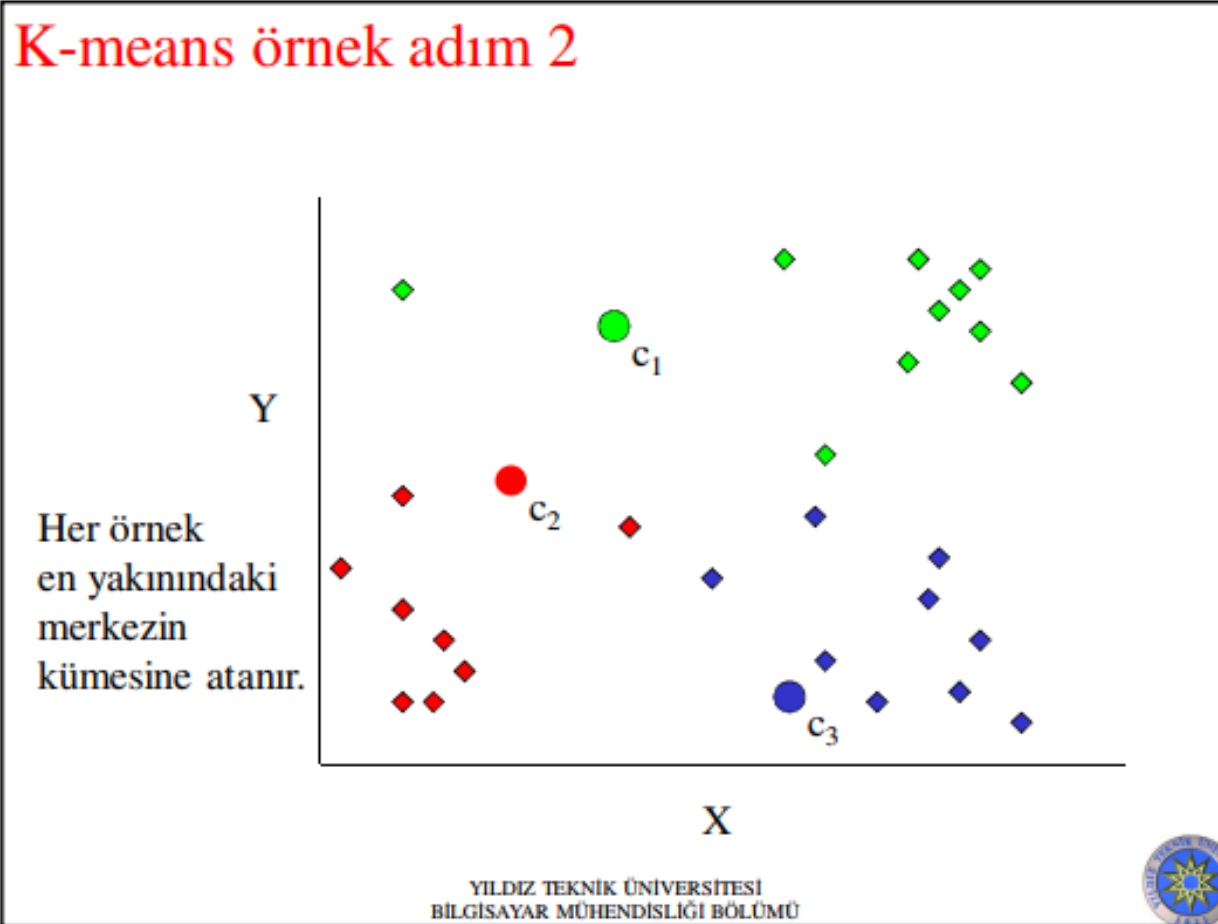




BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

K-MEANS YÖNTEMİ



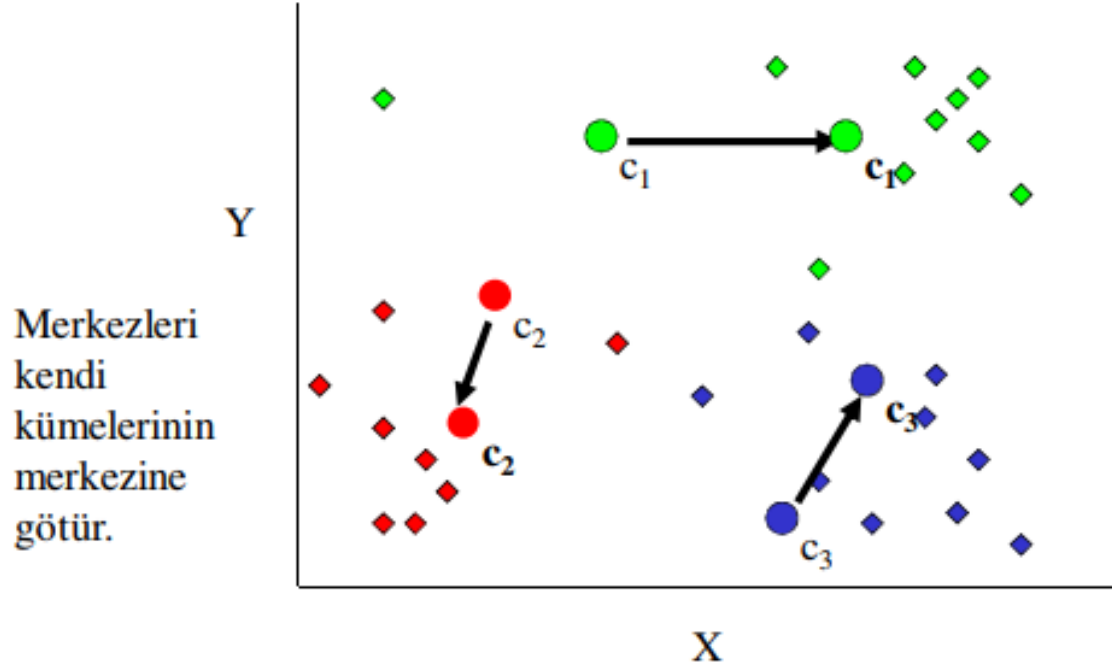


BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

K-MEANS YÖNTEMİ

K-means örnek adım 3





BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

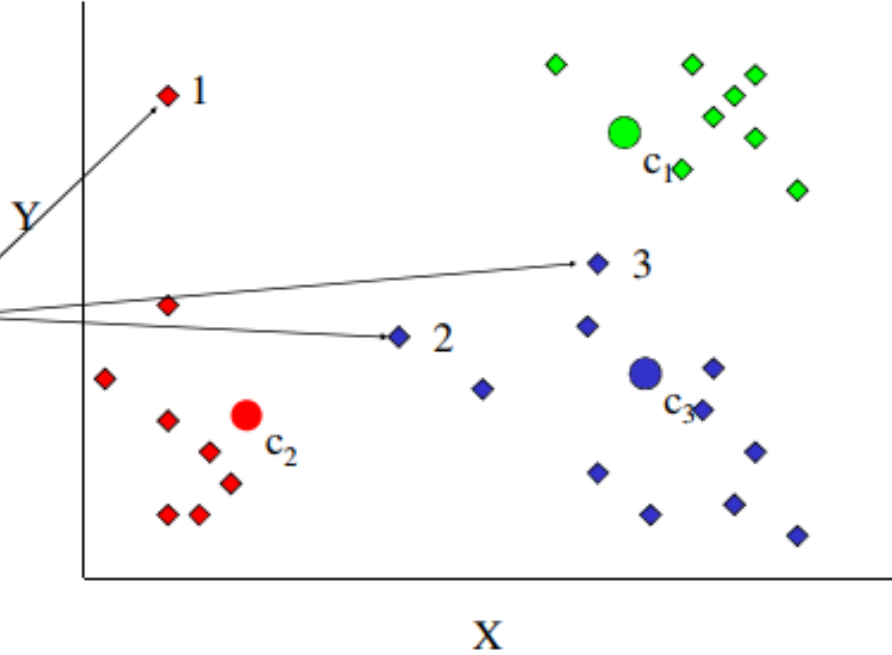
KÜMELEME ANALİZİ

K-MEANS YÖNTEMİ

K-means örnek adım 4

Her örneği yeniden en yakınındaki merkezin kümesine ata.

Q: Hangi örneklerin kümesi değişti?



YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



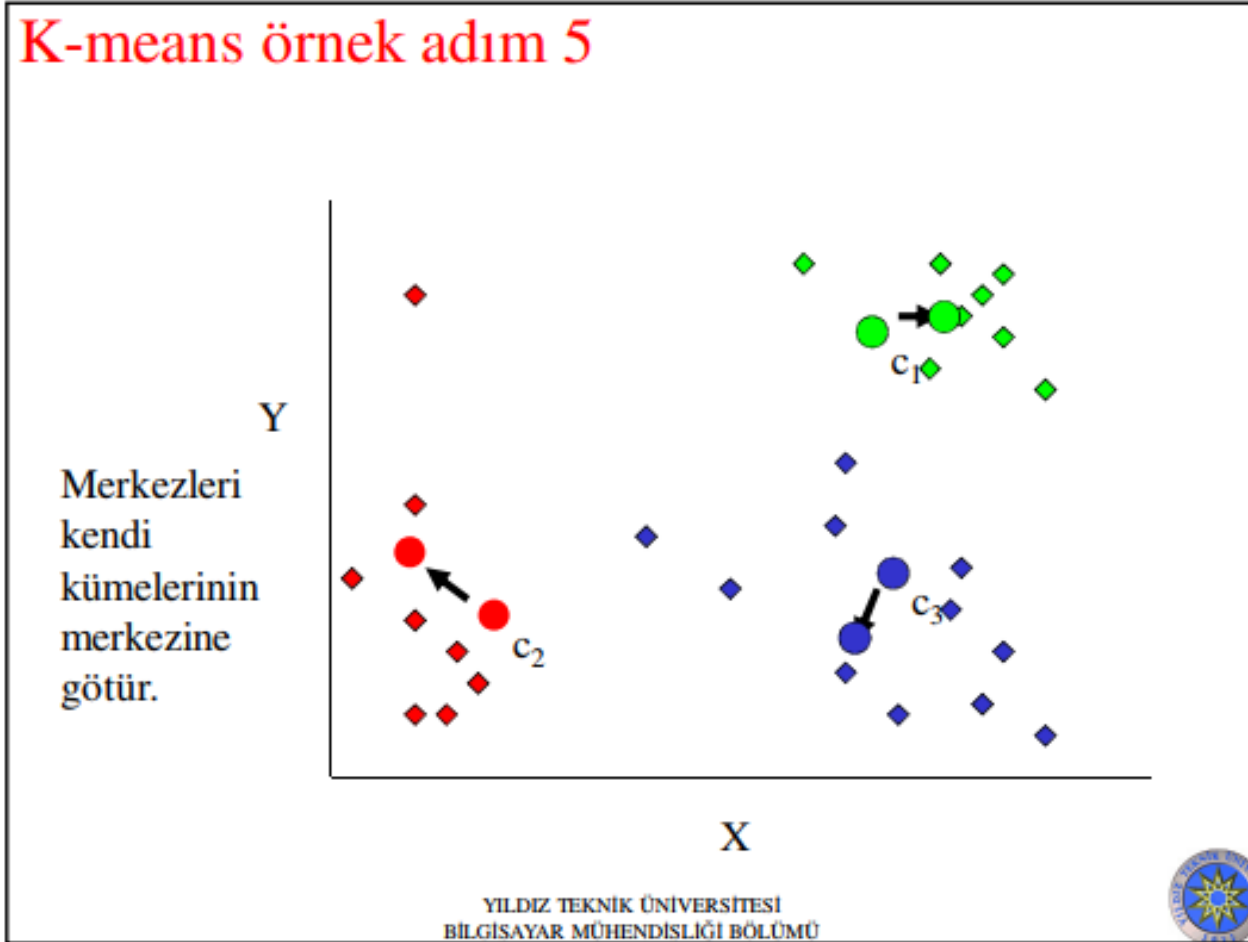


BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

K-MEANS YÖNTEMİ

K-means örnek adım 5





BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

K-MEANS YÖNTEMİ

Matlab Hazır Fonksiyonları ile oluşturulabilecek kod:

```
clc; clear;
% iLK VERÝ SETÝ*****
% X = [randn(100,2)*0.75+ones(100,2);randn(100,2)*0.5-ones(100,2)];

% ÝKÝNCÝ VERÝ SETÝ*****
load('data.mat');

figure;hold on;
plot(X(:,1),X(:,2),'.');%grafik üzerinde veri seti gösteriliyor.

opts = statset('Display','iter');
%Veri seti 6 kümeye bölünüyor. Uzaklık hesabı cityblock yöntemiyle
%yapılıyor.
[idx,C] = kmeans(X,6,'Distance','cityblock','Replicates',5,'Options',opts);
% C matrisi küme merkezlerinin koordinatlarını verir. idx ise her bir
% veri seti vektörünün hangi kümeye ait olduğunu gösterir.

plot(C(:,1),C(:,2),'kx','MarkerSize',15,'LineWidth',2)
```



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

K-MEANS YÖNTEMİ

- Aşağıdaki kodları inceleyiniz.
- Data.mat (700x2 boyutlarında veri seti)
- kMeans2.m (Algoritma ile kendi kodlarımız ile uygulama kodu)
- kMeansMatlabHazirKod.m (Matlab hazır fonksiyonları ile uygulama kodu)

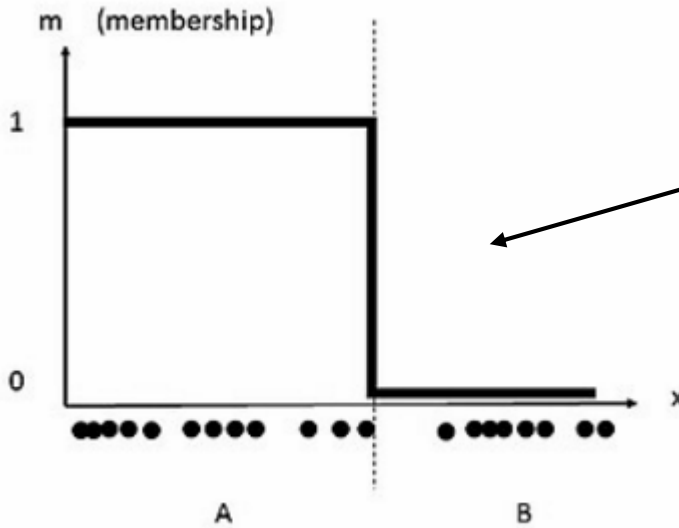


BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

Elimizde bir boyutlu n adet verimiz olduğunu varsayalım. Bu veriler A kümesi ve B kümesi olarak iki kümeye ayrılınsın.



Kesin kümeleme (K-means)
yapıldığında

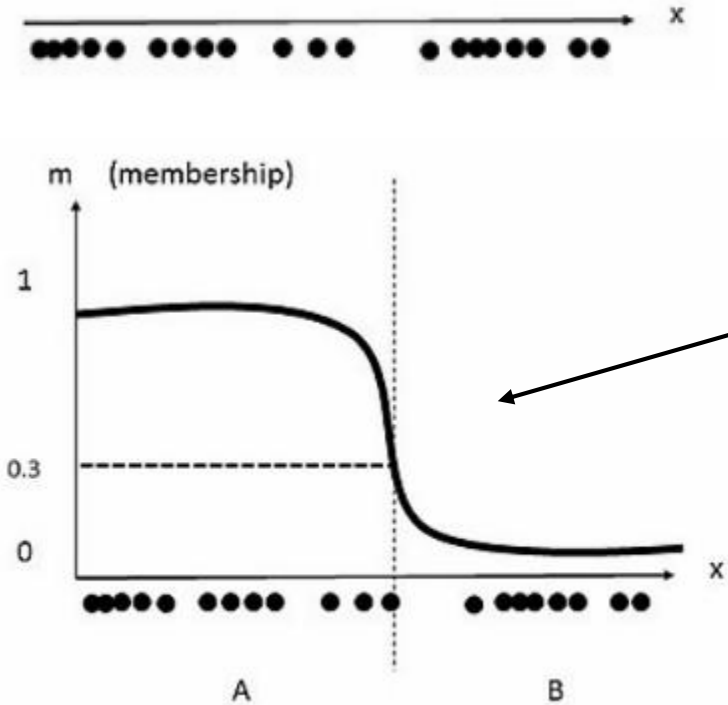


BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

Elimizde bir boyutlu n adet verimiz olduğunu varsayalım. Bu veriler A kümesi ve B kümesi olarak iki kümeye ayrılınsın.



Bulanık kümeleme (C-means)
yapıldığında



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

- Bulanık c-ortalamlar (BCO) kümeleme, veri noktalarını benzerliklerine göre kümeler halinde gruplayan bir veri kümeleme tekniğidir.
- Kümeleme, verilerin benzerliklerine dayanarak gruplar oluşturmayı amaçlar.
- Her veri noktasını tek bir kümeye atayan k-ortalamlar gibi geleneksel kümeleme algoritmalarının aksine, BCO her veri noktasının belirli bir dereceye kadar birden fazla kümeye ait olmasına izin verir. Bu nedenle "bulanık" bir yaklaşım olarak adlandırılır.
- Bu, üyelik derecelerine sahip olabilen kümeler olan bulanık kümeler kullanılarak yapılır.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

- BCO algoritması, küme merkezlerini ve veri noktalarının üyelik derecelerini iteratif olarak güncelleyerek çalışır. Küme merkezleri, o kümeye yüksek derecede üyeliği olan tüm veri noktalarının ortalaması olacak şekilde güncellenir. Veri noktalarının üyelik dereceleri, veri noktaları ile küme merkezleri arasındaki benzerlikle orantılı olacak şekilde güncellenir.
- BCO algoritması tipik olarak veri noktalarının üyelik dereceleri yakınsayana kadar yinelenir. Küme sayısı kullanıcı tanımlı bir parametredir.
- BCO kümeleme, görüntü bölütleme, metin sınıflandırma ve tıbbi teşhis dahil olmak üzere çeşitli uygulamalarda kullanılmıştır.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

BCO kümeleme algoritması (Bezdek, 1981) basit ve kullanışlı bir yöntemdir. Bu yöntemde,

$X = \{x_1, x_2, \dots, x_n\}$ veri setinin kaç kümeye parçalanacağını belirten c sayısının bilindiği veya en azından belirlenebilir olduğu varsayılır.

BCO kümeleme yöntemi, optimum çözüme amaç fonksiyonunun minimumu ile ulaşan bir kısıtlı optimizasyon problemine dayanır. Küme sayısı (c) ve bulanıklık parametresi (m) gibi iki ön bilgi ile bu optimizasyon problemi;



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

Veri seti: $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$

Kümeler: $C = \{\mathbf{c}_1, \dots, \mathbf{c}_c\}$

Amaç Fonksiyonu= $\sum_{j=1}^k \sum_{\mathbf{x}_i \in C_j} u_{ij}^m (\mathbf{x}_i - C_j)^2$ Şeklinde ifade edilebilir.

$u_{i,j}$ x_i verisinin c_j kümesine üyelik derecesi

C_j j kümesinin merkezi

m Bulanıklık katsayısı

$u_{i,j}$ ve m değerleri dışında algoritma K-Ortalamalar algoritması ile aynıdır.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

Veri setinin her bir i . elemanının j kümesine üyelik derecesi aşağıdaki şekilde hesaplanır. Burada (m) bulanıklık katsayısına dikkat ediniz.

$$u_{ij}^m = \frac{1}{\sum_{l=1}^k \left(\frac{|x_i - c_j|}{|x_i - c_l|} \right)^{\frac{2}{m-1}}}$$

Yapılan çalışmalarda genel olarak $m=2$ olarak kullanılmaktadır.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

	KÜME1	KÜME2	KÜME3	KÜME4
1	0.9887	0.0023	0.0047	0.0043
2	0.8396	0.0326	0.0712	0.0565
3	0.6519	0.0719	0.1028	0.1735
4	0.0298	0.8540	0.0635	0.0527
5	0.0044	0.0049	0.0023	0.9884
6	0.0645	0.0437	0.8645	0.0272
7	0.0942	0.0950	0.0497	0.7611
8	0.0081	0.9599	0.0148	0.0173
9	0.9360	0.0128	0.0280	0.0232
10	0.0641	0.0355	0.8766	0.0238

- Yan tarafta örnek bir 10 elemanlı veri setinin 4 kümeye bölündüğünde oluşan üyelik değerleri tablosu verilmiştir.
- Dikkat edilirse tablodaki satırların toplamı bir (1)'dir. Bir veri, belirli bir eşik değerden daha yüksek değerde üyelik değeriyle bağlı olduğu kümenin elemanı olarak kabul edilebilir.
- Ancak diğer kümelerin de az da olsa elemanıdır.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

Her bir kümenin merkez inin hesaplanması ise aşağıdaki şekilde gerçekleştirilir.

$$C_j = \frac{\sum_{x \in C_j} u_{ij}^m x}{\sum_{x \in C_j} u_{ij}^m}$$

C_j j kümesinin merkezi

Başlangıçta küme merkezleri rastlantısal olarak belirlenir.

Durma koşulları sağlanıncaya kadar küme merkezleri ve küme elemanları iteratif olarak hesaplanır.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

Matlab'da fcm() fonksiyonu kullanımı

```
[center, U, obj_fcn] = fcm(data, cluster_n, seçenekler)
```

Burada:

- "data", kümeleme yapılacak veri setini temsil eder.
- "cluster_n", oluşturulacak küme sayısını belirtir.
- "center", hesaplanan küme merkezlerini içeren bir matristir.
- "U", her bir veri noktasının her bir kümeyle ilişkisini temsil eden üyelik değerlerini içeren bir matristir.
- "obj_fcn", optimizasyon fonksiyonunun değerini temsil eder.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

Matlab'da fcm() fonksiyonu kullanımı

[center, U, obj_fcn] = fcm(data, cluster_n, seçenekler)

Seçenekler bölümünde 4 farklı parametre verilebilir. Verilmez ise varsayılan değerler otomatik olarak kullanılır.

Seçenek 1.

Bulanık bölümlenme matrisi için üs değeridir. Bu seçenek, Fuzzy C-Means (FCM) algoritmasında üyelik derecesinin ağırlıklandırma katsayısı olarak bilinen parametredir. FCM algoritmasının bulanıklık seviyesini kontrol eder. Bu parametre genellikle bulanıklık parametresi veya fuzziness exponent olarak da adlandırılır ve genellikle 2 olarak seçilir. **m değeri ne kadar büyükse, bulanıklık o kadar fazla olur; m değeri ne kadar küçükse, üyelik dereceleri o kadar keskin olur.**

m = 1 seçilirse; Bu durumda, FCM algoritması sert k-means kümelemeye benzer şekilde çalışır. Her veri noktası yalnızca bir kümeye %100 üyelik derecesine sahip olur.

Varsayılan Değer:2



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

Matlab'da fcm() fonksiyonu kullanımı

[center, U, obj_fcn] = fcm(data, cluster_n, seçenekler)

Seçenek 2.

Pozitif bir tamsayı olarak belirtilen maksimum yineleme sayısı.

Varsayılan Değer=100

Seçenek 3.

Pozitif skaler olarak belirtilen, iki ardışık iterasyon arasında amaç fonksiyonundaki minimum gelişme.

Varsayılan Değer=1e-5

Seçenek 4.

Her iterasyondan sonra amaç fonksiyonu değerinin görüntülenip görüntülenmeyeceğini belirten bilgi görüntüleme bayrağı,

true - Amaç fonksiyonunu göster. (Varsayılan)

false - Amaç fonksiyonunu görüntüleme.



BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

Örnek Kodlar -1

```
clear;clc;close all;
fcmdata = simplecluster_dataset;
fcmdata=fcmdata';
[centers,U] = fcm(fcmdata,4,[2.5 200 1e-5 1]);
maxU = max(U);
index1 = find(U(1,:) == maxU);%Kümeler arasında en yüksek üyelik değeri olan küme belirleniyor..
index2 = find(U(2,:) == maxU);
index3 = find(U(3,:) == maxU);
index4 = find(U(4,:) == maxU);
plot(fcmdata(index1,1),fcmdata(index1,2),'ob')
hold on
plot(fcmdata(index2,1),fcmdata(index2,2),'or')
plot(fcmdata(index3,1),fcmdata(index3,2),'oy')
plot(fcmdata(index4,1),fcmdata(index4,2),'oc')

plot(centers(1,1),centers(1,2),'xc','MarkerSize',15,'LineWidth',3)
plot(centers(2,1),centers(2,2),'xy','MarkerSize',15,'LineWidth',3)
plot(centers(3,1),centers(3,2),'xr','MarkerSize',15,'LineWidth',3)
plot(centers(4,1),centers(4,2),'xb','MarkerSize',15,'LineWidth',3)
hold off
```

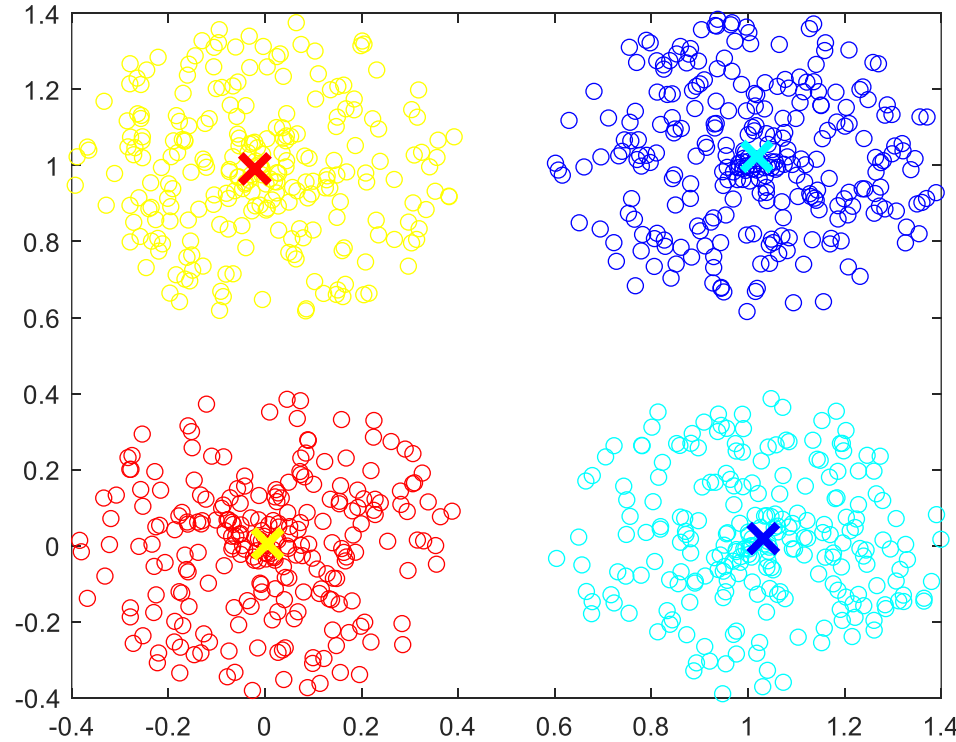


BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ

FUZZY C-MEANS (FCM) YÖNTEMİ

Örnek Kodlar -1





BULANIK MANTIK ile KONTROL ve UYARLAMA SİSTEMLERİ

KÜMELEME ANALİZİ - KAYNAKLAR

- [1] http://www.academia.edu/9860553/Veri_Madencili%C4%9Fi__K%C3%BCmeleme_ve_K%C3%BCmeleme_Y%C3%B6ntemleri
- [2] <http://mnemstudio.org/neural-networks-kohonen-self-organizing-maps.htm>
- [3] http://www.statistics4u.com/fundstat_eng/cc_ann_kohonen.html
- [4] <http://bilgisayarkavramlari.sadievrenseker.com/2008/12/15/self-organizing-maps-ozduzenleyici-haritalar/>
- [5] **Dinçer E.**, 2006. Veri Madenciliğinde K-means Algoritması ve Tıp Alanında Uygulanması, p. 24-64.
- [6] **Han J., Kamber M.**, 2001. Data Mining Concepts and Techniques, Morgan Kauffmann Publishers Inc.
- [7] **Berkhin P.**, 2002. Survey of Clustering Data Mining Techniques, San Jose, California, USA, Accrue Software Inc.
- [8] Bezdek, J.C. (1981) "Pattern Recognition with Fuzzy Objective Function Algorithms," Plenum Press, New York.