Mustafa AYGÜN

383679@ogr.ktu.edu.tr

K-means en sık kullanılan kümeleme algoritmalarındandır. Uygulanması kolaydır. Büyük ölçekli verileri hızlı ve etkin şekilde kümeleyebilir.

K-Ortalamalar Algoritması

Bilgisayar Bilimlerine Giriş

İçindekiler

[1. K-Ortalamalar Algoritması 2](#_Toc532772948)

[2. Kümeleme Yöntemi 3](#_Toc532772949)

[3. Şekil Dizini 4](#_Toc532772950)

[4. Tablo Dizini 5](#_Toc532772951)

[5. Kaynakça 6](#_Toc532772952)



Şekiller İndeksi

[Şekil 1. Kümeleme Sonucu 2](#_Toc532778694)

[Şekil 2.Result of clustering 2](#_Toc532778695)

Tablo İndeksi

Tablo 1. Küme Merkezleri. 3

* 1. K-Ortalamalar Algoritması

En eski [kümeleme](http://mustafaakca.com/tag/kumeleme/) algoritmalarından olan [k-means](http://mustafaakca.com/tag/k-means/), 1967 yılında [J.B. MacQueen](http://mustafaakca.com/tag/j-b-macqueen/)  tarafından geliştirilmiştir (MacQueen, 1967)

[K](http://mustafaakca.com/tag/k-means/)-[Means](http://mustafaakca.com/tag/kumeleme/" \o "Kümeleme) Algoritması [Data Mining Dünyasında En Çok Kullanılan Algoritmalar](http://mustafaakca.com/data-mining-dunyasinda-top-10-algoritma/)ın başında yer almaktadır.  [Kümeleme algoritmaları ile Sınıflandırma algoritmaları](http://mustafaakca.com/k-means-kumeleme-algoritmasi/www.mustafaakca.com/kumeleme-ile-siniflandirma-arasindaki-fark)arasında birtakım farklılıklar bulunmaktadır. [K-means](http://mustafaakca.com/tag/k-means/) algoritması bir [kümeleme](http://mustafaakca.com/tag/kumeleme/) algoritmasıdır. [Kümeleme](http://mustafaakca.com/tag/kumeleme/) algoritmaları otomatik olarak verileri daha küçük kümelere ya da alt kümelere ayırmaya yarayan algoritmalardır. Algoritma istatistiksel olarak benzer nitelikteki kayıtları aynı gruba sokar. Bir elemanın yalnızca bir kümeye ait olmasına izin verilir.  Küme merkezi kümeyi temsil eden değerdir.

Algoritmanın isminde yer alan “k” harfi, aslında küme sayısını belirtir: Algoritma, hata hesaplamada yaygın olarak kullanılan Karasel Hata Fonksiyonunu en aza indirgeyecek “k” küme sayısını da arar. Verilen “n” sayıdaki veri seti “k” tane kümeye bu hata fonksiyonunu en aza indirgeyecek şekilde yerleştirilir. Bu nedenle küme benzerliği kümedeki değerlerin ortalamaya yakınlıkları ile ölçülür. Bu da kümenin ağırlık merkezidir. Kümenin merkezinde yer alan değer kümenin temsilci değeridir ve medoid olarak adlandırılır.

K-ortalama kümeleme ya da K-means kümeleme (K-means clustering) yöntemi N adet veri nesnesinden oluşan bir [veri kümesini](https://tr.wikipedia.org/wiki/Veri_k%C3%BCmesi) giriş parametresi olarak verilen K adet kümeye bölümlemektir. Amaç, gerçekleştirilen bölümleme işlemi sonunda elde edilen kümelerin, küme içi benzerliklerinin maksimum ve kümeler arası benzerliklerinin ise minimum olmasını sağlamaktır.

K-means en sık kullanılan [kümeleme](https://tr.wikipedia.org/wiki/K%C3%BCmeleme) algoritmalarındandır. Uygulanması kolaydır. Büyük ölçekli verileri hızlı ve etkin şekilde kümeleyebilir. “K” algoritmaya başlamadan önce ihtiyaç duyulan sabit küme sayısını ifade etmektedir. Tekrarlı bölümleyici yapısı ile K-means algoritması, her verinin ait olduğu kümeye olan uzaklıkları toplamını küçültmektedir. K-means algoritması karesel hatayı en küçük yapacak olan K adet kümeyi tespit etmeye çalışmaktadır.

K-means ile küme içi benzerlik büyük, kümeler arası benzerlik ise küçük olduğu sürece kümelenmenin doğruluğundan söz edilebilir. Problem [NP-hard](https://tr.wikipedia.org/wiki/NP-hard) olmasına rağmen K-means algoritması bir iteratif (tekrarlayıcı) yaklaşım ile genelde iyi bir çözüm verir.

* 1. Kümeleme Yöntemi

K-means algoritmasının çalışma mekanizmasına göre öncelikle her kümenin merkez noktasını veya ortalamasını temsil etmek üzere K adet nesne rastgele seçilir. Kalan diğer nesneler, kümelerin ortalama değerlerine olan uzaklıkları dikkate alınarak en benzer oldukları kümelere dahil edilir. Daha sonra, her bir kümenin ortalama değeri hesaplanarak yeni küme merkezleri belirlenir ve tekrar nesnelerin merkeze uzaklıkları incelenir. Herhangi bir değişim olmayıncaya kadar algoritma tekrarlamaya devam eder.

Burada en önemli iki amaç şudur:

Küme içindeki değerler birbirlerine en çok benzemeli,

Kümeler birbirine mümkün olduğunca benzememeli

Bu istekleri gerçekleştirmek için algoritma tarafında sırasıyla şu adımlar gerçekleştirilmelidir:

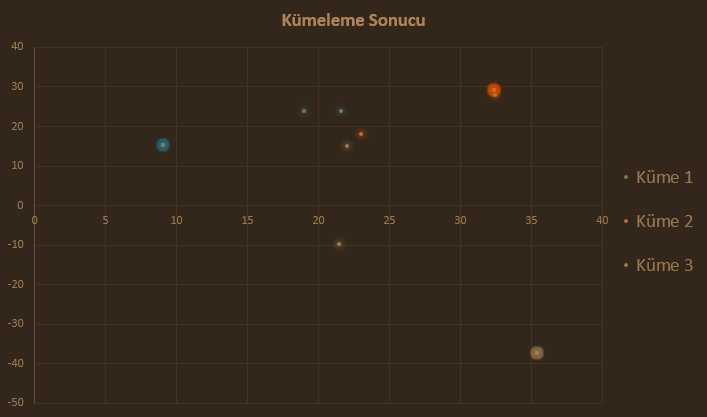
1- Sınıf merkezlerinin belirlenmesi

2- Örneklerin mesafelere göre sınıflandırılması

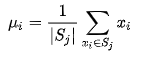
3- Yapılan sınıflandırma sonrasında yeni merkezlerin belirlenmesi

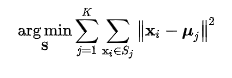
4- İstenilen hale gelinceye kadar 2. ve 3. adımların algoritmik olarak tekrarlanması.

* 1. Şekil Dizini

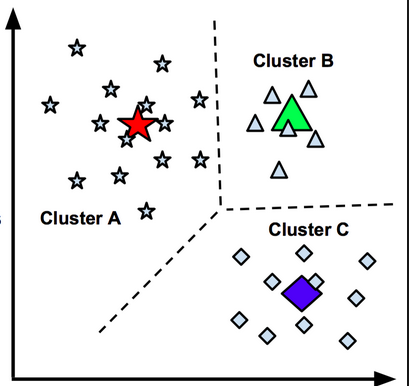


Şekil 1. Kümeleme Sonucu

Her bir veri n-boyutlu reel vektör olmak üzere bir {x1, x2, …, xN} veri kümesi ve K bölünecek küme sayısı olarak verilsin. K-means kümeleme, karesel hatayı en aza indirgemek için N tane veriyi K adet S = {S1, S2, …, SK} kümeye bölümlemeyi amaçlar. Başka bir deyişle,

burada {\displaystyle \mu \_{i}} *Sj* 'deki noktaların ortalaması olmak üzere

bulmaktır.



Şekil 2.Result of clustering

* 1. Tablo Dizini



Tablo 1. Küme Merkezleri.

* 1. Kaynakça

<http://mustafaakca.com/k-means-kumeleme-algoritmasi/>

<https://tr.wikipedia.org/wiki/K-means_kümeleme>

<https://medium.com/@dilekamadushan/introduction-to-k-means-clustering-7c0ebc997e00>