



*T.C*

***SAKARYA ÜNİVERSİTESİ***

**BİLGİSAYAR VE BİLİŞİM BİLİMLERİ FAKÜLTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

# **BSM 310 – YAPAY ZEKA**

**Grup üyeleri:**

**G1712.10308 MÜCAHİT AKTÜRK  
G1712.10044 NİSANUR KARATEPE  
G1612.10560 NERMİN KAYA**

**Sakarya  
2020**

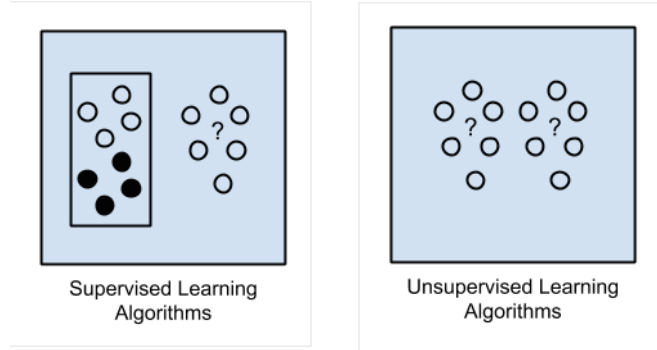
# K-Means (K-Ortalama Kümeleme)

## Supervised ve Unsupervised Learning:

Makine öğrenmesi temelde bu iki gruba göre ayrılır:

- Supervised (Gözetimli) .
- Unsupervised (Gözetimsiz) öğrenme .

**Supervised Learning:** Sınıf sayınız belli, Eğitilmiş veriniz var. Yani elinizdeki verilerin hangi girdi değerinde hangi çıktıyı vereceği işaretlemiş iseniz. Bu tarzda eğitilmiş makine öğrenmeye **Gözetimli Öğrenme** diyoruz. Mesela verdiğiniz resimlerin insanın yüzü olup olmadığına karar veren bir sistem tasarlamak istiyorsanız ilk önce farklı insan yüzlerini verip bunları yüz şeklinde işaretlerseniz, içerisinde yüz olmayan resimleri de tam tersi yönünde işaretlerseniz.



**Unsupervised Learning:** Burada elinizdeki verilerin özelliklerine göre Clustering (kümeleme) yöntemi ile belli gruplar oluşturan öğrenme şeklidir. Bu öğrenme şeklinde bu gruptan yüz olup olmadığını bilemezsiniz, fakat örnek bir resim verdiğinizde bunun hangi kümeleme içerisinde olduğunu bularak benzerlerini getirebilir.

k-means verilen bir veri seti üzerinden belirli sayıda kümeyi (k adet) gruplamak için geliştirilmiş en sade ve basit **Unsupervised “Gözetimsiz”** makine öğrenme algoritmalarından biridir.

## Clustering «Kümeleme» nedir ?

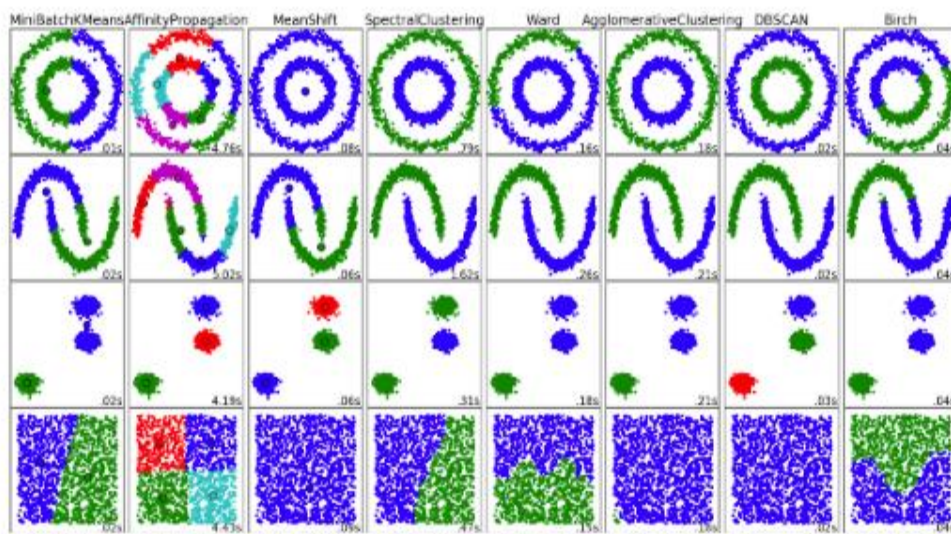
Kümeleme analizi bir veri kümesindeki bilgileri belirli yakınlık kriterlerine göre gruplara ayırma işlemidir. Bu grupların her birine “küme” adı verilir. Kümeleme analizine kısaca “kümeleme” adı verilir.

Kümeleme işleminde küme içindeki elemanların benzerliği fazla, kümeler arası benzerlik ise az olmalıdır. Kümeleme işlemi tamamen gelen verinin özelliklerine göre yapılır.

## Neden Kümeleme?

Büyük veri kümeleriyle çalışırken, bunu analiz etmenin etkili bir yolu, önce verileri mantıksal gruplara, diğer bir deyişle kümelere bölmektir. Bu şekilde çok sayıda yapılandırılmamış veri kümesinden değer elde edebilirsiniz. Verileri belirli bulgular için daha derinlemesine analiz etmeden önce bazı desenleri veya yapıları ortaya çıkarmak için veriler arasında göz atmanıza yardımcı olur. Verileri kümeler halinde düzenlemek, verilerdeki temel yapının belirlenmesine yardımcı olur ve endüstrilerdeki uygulamaları bulur.

Örneğin, kümeleme tıp bilimi alanındaki hastalıkları sınıflandırmak için kullanılabilir ve pazarlama araştırmasında müşteri sınıflandırmasında da kullanılabilir. Bazı uygulamalarda, veri bölümlenme nihai hedeftir, öte yandan, kümeleme de diğer yapay zeka veya makine öğrenimi problemlerine hazırlanmak için bir ön koşuldur. Tekrarlanan kalıplar, temel kurallar ve daha fazlası şeklinde verilerde bilgi keşfi için etkili bir tekniktir.



Clustering Algorithms

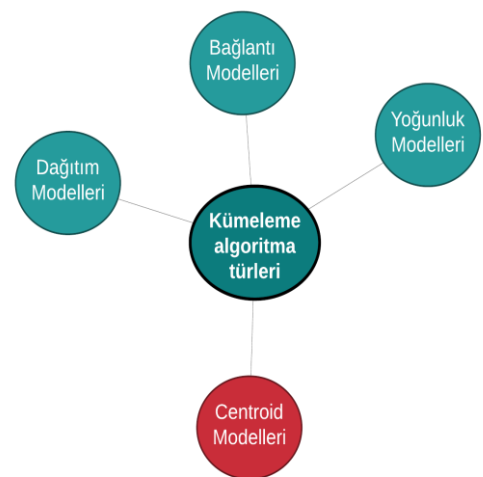
**Clustering «Kümeleme» analizi :** Kümeleme analizi, bireylerin ya da nesnelerin sınıflandırılmasını ayrıntılı bir şekilde açıklamak amacıyla geliştirilmiştir. Bu amaca yönelik olarak, bir örnekte yer alan varlıklar aralarındaki benzerliklere göre gruplara ayrılır, daha sonra bu gruplara dahil edilen bireylerin profili ortaya konur. Bir başka ifade ile kümelemenin amacı, öncelikle ele alınan örnekte gerçekte var olduğu bilinen, varlıklar (birey ya da nesne) arasındaki benzerliklere dayanan az sayıdaki karşılıklı özel grupları oluşturmak, daha sonra bu gruplara giren varlıkların profilini ortaya koymaktır. Diğer bir hedef ise benzer elemanların gruplandırılmasıyla veri setini küçültmektir.

**Clustering «Kümeleme» Yöntemleri :** Kümeleme görevlerinin öznel doğası göz önüne alındığında, farklı kümeleme sorunlarına uyan çeşitli algoritmalar vardır. Her sorunun, iki veri noktası arasındaki benzerliği tanımlayan farklı bir kurallar kümesi vardır, bu nedenle kümelenme hedefine en uygun algoritmayı gerektirir. Bugün, kümelenme için yüzden fazla bilinen makine öğrenme algoritması bulunmaktadır. Kümeleme algoritması veri tabanını alt kümelere ayırır. Her bir kümede yer alan elemanlar dahil oldukları grubu diğer gruplardan ayıran ortak özelliklere sahiptir. Kümeleme modellerinde amaç, küme üyelerinin birbirlerine çok benzediği, ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere bölünmesidir.

### Clustering «Kümeleme» Algoritma Türleri :

#### ➤ Bağlantı Modelleri :

Adından da anlaşılacağı gibi, bağlantı modelleri veri noktalarını veri noktalarının yakınlığına göre sınıflandırma eğilimindedir. Birbirine daha yakın olan veri noktalarının, daha uzağa yerleştirilenlere göre daha benzer özellikleri tasvir ettiği fikrine dayanır. Algoritma, belirli noktalarda birbiriyle birleşebilecek kapsamlı bir küme hiyerarşisini destekler. Veri kümesinin tek bir bölünmesi ile sınırlı değildir.

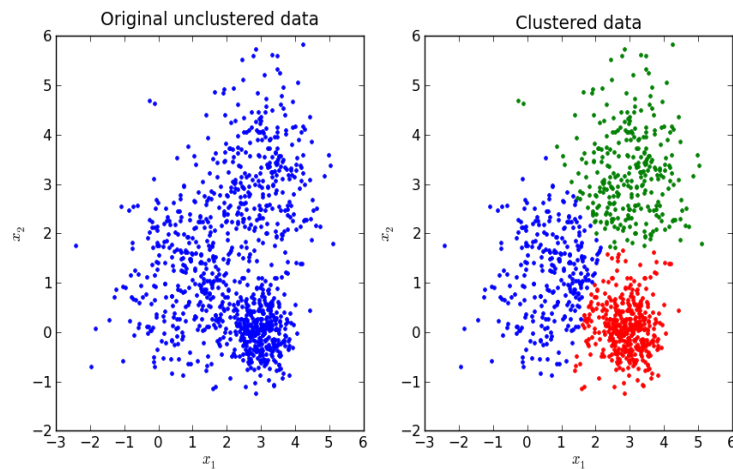


Mesafe fonksiyonu seçimi öznel ve her kümeleme uygulamasına göre değişebilir. Bağlantı modellerinde bir kümelenme sorununu ele almak için iki farklı yaklaşım vardır. Birincisi, tüm veri noktalarının ayrı kümeler

halinde sınıflandırıldığı ve daha sonra mesafe azaldıkça toplandığı yerdir. İkinci yaklaşım, tüm veri kümesinin bir küme olarak sınıflandırıldığı ve daha sonra mesafe arttıkça birden çok kümeye ayrıldığı yerdir. Model kolayca yorumlanabilse de, daha büyük veri kümelerini işlemek için ölçeklenebilirlikten yoksundur.

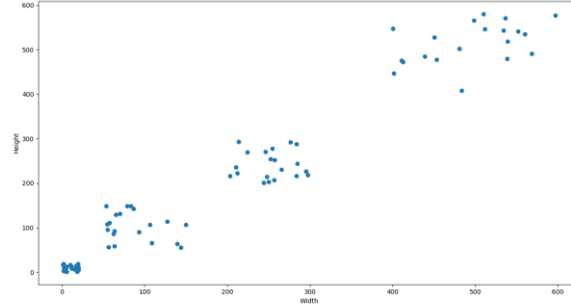
- **Dağıtım Modelleri:** Dağıtım modelleri, bir kümedeki tüm veri noktalarının aynı dağılıma, yani Normal dağılım veya Gauss dağılımına ait olma olasılığını temel alır. Küçük bir dezavantaj, modelin aşırı uyumdan muzdarip olma eğiliminde olmasıdır. Bu modelin iyi bilinen bir örneği, beklenti maksimizasyon algoritmasıdır.
- **Yoğunluk Modelleri :** Bu modeller, veri alanında çeşitli veri noktalarının yoğunluğunu araştırır ve farklı yoğunluk bölgelerini izole eder. Daha sonra kümeler ile aynı bölgedeki veri noktalarını atar. DBSCAN ve OPTICS yoğunluk modellerinin en yaygın iki örneğidir.
- **Centroid Modelleri :** Centroid modelleri, veri noktaları arasındaki benzerliğin, kümenin sentroidine olan yakınlıklarına dayanılarak elde edildiği yinelemeli kümeleme algoritmalarıdır. Centroid (kümenin merkezi), veri noktalarının mesafesinin merkezle minimum olmasını sağlayarak oluşturulur. Bu tür kümelenme sorunlarının çözümü genellikle birden fazla denemede tahmin edilmektedir. Centroid modellerine bir örnek K-ortalama algoritmasıdır.

**K-Means Algoritması :** K-Means, çok çeşitli veri bilimi ve makine öğrenimi problemlerini anlamak ve uygulamak için çok kolay olduğu göz önüne alındığında, en popüler kümeleme algoritmasıdır .

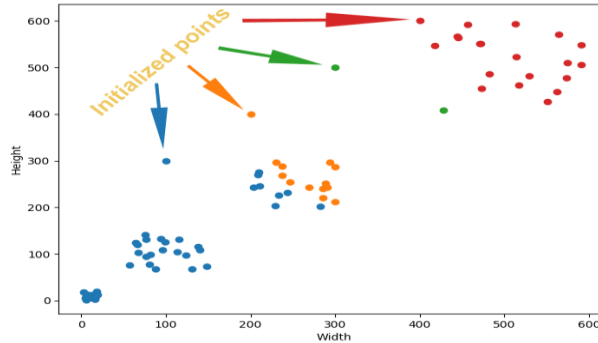


## ***K-Means algoritmasını kümeleme probleminize nasıl uygulanır ?***

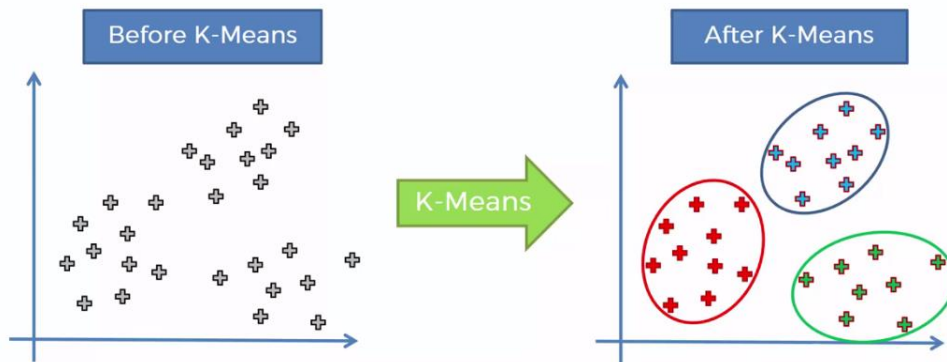
Her zamanki gibi, sorunun ilk kısmı mümkün olduğunca fazla veri toplamaktır.



Bir sonraki adım, düzlemde rastgele seçilen "K" sayısını ile temsil edilen rastgele kümeler seçmektir. Daha sonra, her kümeye bir merkez, yani belirli bir kümenin merkezi atanır. Varyasyonu azaltmak için sentroidleri birbirinden mümkün olduğunca uzakta tanımlamak önemlidir. Tüm sentroidler tanımlandıktan sonra, her veri noktası, sentroidi en yakın mesafede olan kümeye atanır. Tüm veri noktaları ilgili kümelerle atandıktan sonra, her küme için sentroid yeniden atanır. Bir kez daha, tüm veri noktaları yeni tanımlanmış sentroidlerden olan uzaklıklarına göre belirli kümelerde yeniden düzenlenir. Bu işlem, sentroidler pozisyonlarından hareket etmeyi bırakana kadar tekrarlanır.



K-Means algoritması yeni verilerin gruplanmasında harikalar yaratıyor. Bu algoritmanın pratik uygulamalarından bazıları sensör ölçümleri, ses algılama ve görüntü segmentasyonundadır.



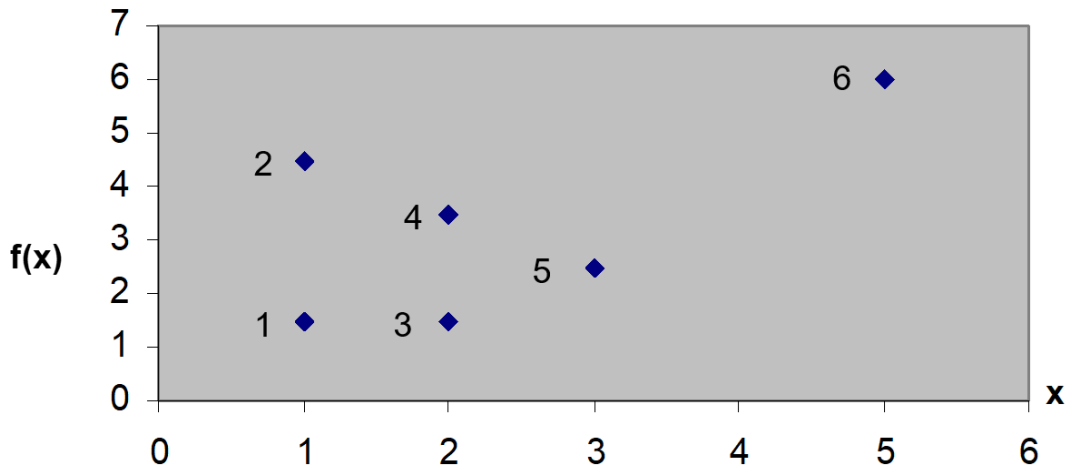
### Örnek:

Başlangıç değerleri :

#### • K-Means Input Values

Instance	X	Y
1	1.0	1.5
2	1.0	4.5
3	2.0	1.5
4	2.0	3.5
5	3.0	2.5
6	5.0	6.0

Değerlerin koordinat düzlemindeki yerleşimleri:

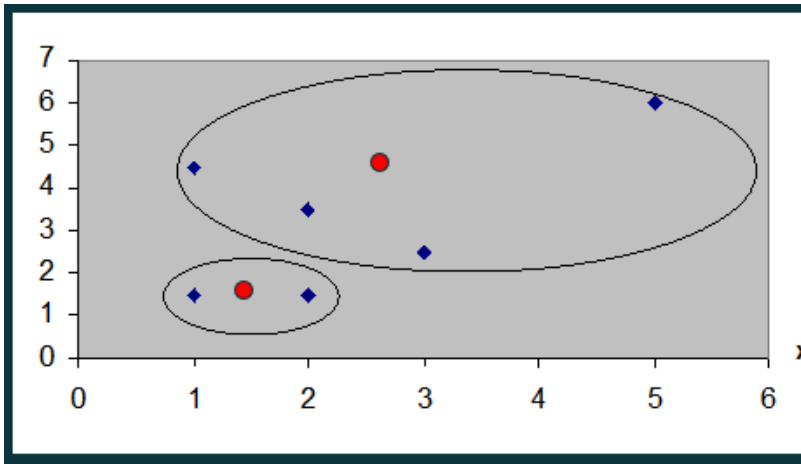


Değerlerin K-Means Algoritmasıyla sonuçlarının hesaplanması:

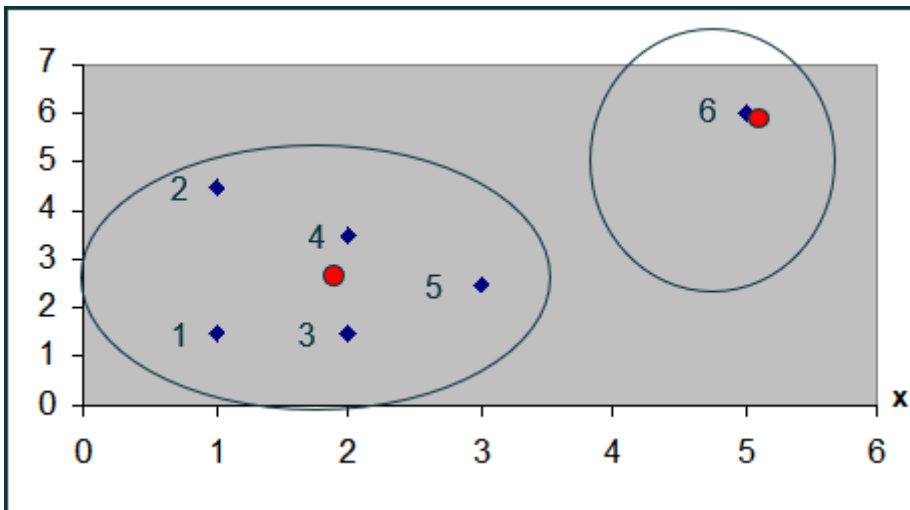
• Several Applications of the K-Means Algorithm ( $K = 2$ )

Outcome	Cluster Centers	Cluster Points	Squared Error
1	(2.67,4.67)	2, 4, 6	14.50
	(2.00,1.83)	1, 3, 5	
2	(1.5,1.5)	1, 3	15.94
	(2.75,4.125)	2, 4, 5, 6	
3	(1.8,2.7)	1, 2, 3, 4, 5	9.60
	(5,6)	6	

2. İterasyon sonucunda kümelerin şekli:



3. İterasyon sonunda kümelerin şekli:





## **K-Means Algoritması Nerelerde Uygulanabilir?**

Bir algoritmanın nasıl çalıştığı kadar neye hizmet ettiği de önemlidir. K-means algoritmasının günlük hayatta kullanılabileceği bazı yerler şu şekildedir:

- ✓ **Belge Sınıflandırması** : Belgeleri etiketlere, konulara ve belgenin içeriğine göre birden fazla kategoride kümeleyin. Bu çok standart bir sınıflandırma problemidir ve k-means aracı bu amaç için oldukça uygun bir algoritmadır.
- ✓ **Suç Yerlerinin Belirlenmesi** : Bir şehirdeki belirli bölgelerde mevcut olan suçlarla ilgili veriler, suç kategorisi, suç alanı ve ikisi arasındaki ilişki, bir şehirdeki ya da bölgedeki suça eğilimli alanlara ilişkin kaliteli bilgiler verebilir.
- ✓ **Müşteri Segmentasyonu**: Kümeleme, pazarlamacıların müşteri tabanını geliştirmelerine, hedef alanlarda çalışmasına ve müşterileri satın alma geçmişine, ilgi alanlarına veya etkinlik izlemeye göre segmentlere ayırmasına yardımcı olur. Sınıflandırma, şirketin belirli kampanyalar için belirli müşteri kümelerini hedeflemesine yardımcı olur.
- ✓ **Oyuncu Analizi** : Oyuncu istatistiklerini analiz etmek, spor dünyasının her zaman kritik bir unsuru olmuştur ve artan rekabetle birlikte, makine öğrenmenin burada oynayacağı kritik bir rol vardır.
- ✓ **Dolandırıcılık Tespiti** : Makine öğrenimi sahtekarlık tespitinde önemli bir rol oynar ve otomobil, sağlık ve sigorta sahtekarlığı tespitinde sayısız uygulamaya sahiptir. Sahte iddialarla ilgili geçmiş verileri kullanarak, yeni iddiaları , sahte kalıpları belirten kümelerle yakınlığına dayanarak izole etmek mümkündür.
- ✓ **Çağrı Kaydı Detay Analizi**: Bir çağrı detay kaydı (CDR), telekom şirketleri tarafından bir müşterinin araması, SMS ve internet etkinliği sırasında elde edilen bilgilerdir. Bu bilgiler, müşteri demografisiyle birlikte kullanıldığında, müşterinin ihtiyaçları hakkında daha fazla bilgi sağlar.

## Kaynaklar

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

<https://medium.com/datadriveninvestor/k-means-clustering-b89d349e98e6>

<https://medium.com/pursuitnotes/k-means-clustering-model-in-6-steps-with-python-35b532cfa8ad>

<https://medium.com/@adityakaushik471/kmeans-algorithm-25634378261b><https://medium.com/@EbubekirBbr/makine-%C3%B6%C4%9Frenmesi-nedir-dbb55d007da0>

<https://medium.com/deep-learning-turkiye/makine-%C3%B6%C4%9Frenmesi-e%C4%9Flencelidir-b9d50aad3a62>

<https://medium.com/machine-learning-for-humans/why-machine-learning-matters-6164faf1df12>

<https://www.bilgiinegi.com/kumeleme-clustering-analizi-nedir.htm>

<https://www.mygreatlearning.com/blog/clustering-algorithms-in-machine-learning/>