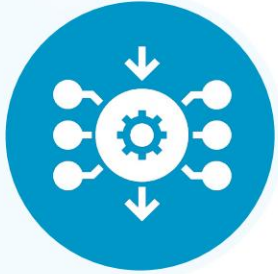


Etiketsiz bir veri kümesinden öğrenme: denetimsiz öğrenme



Denetimsiz öğrenme nasıl çalışır?

- Denetimsiz öğrenme, keşfetmek için ML algoritmaları nı kullanı r insan denetimi olmadan verilerdeki gizli kalı plar . Benzerlikleri ve zı tlı kları bulmak için veri analizinde etkili bir araçtı r.
- Örneğin, hangi grubun daha fazla harcama yaptı ğı nı analiz etmek için müşterileri aradı kları ürün türlerine ve çevrimiçi olarak geçirdikleri süreye göre gruplandı rmak.

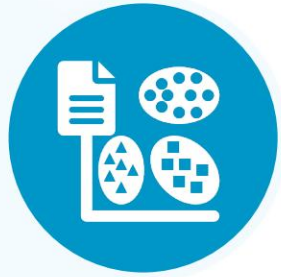
En yaygın denetimsiz öğrenme tekniklerinden biri: kümeleme

Kümeleme, farklı veri noktalarını etiketler yerine özelliklerindeki benzerliklere göre gruplanması işlemidir.

- Örneğin, benzer şekle, boyuta ve renge sahip meyveler aynı kümenin parçasıdır ve dolayısıyla aynı tür meyvelerdir, örneğin kavun veya yaban mersini.



Bilgisayarların kümeleme yapmasını sağlamak için K-Means algoritmasını kullanıyoruz



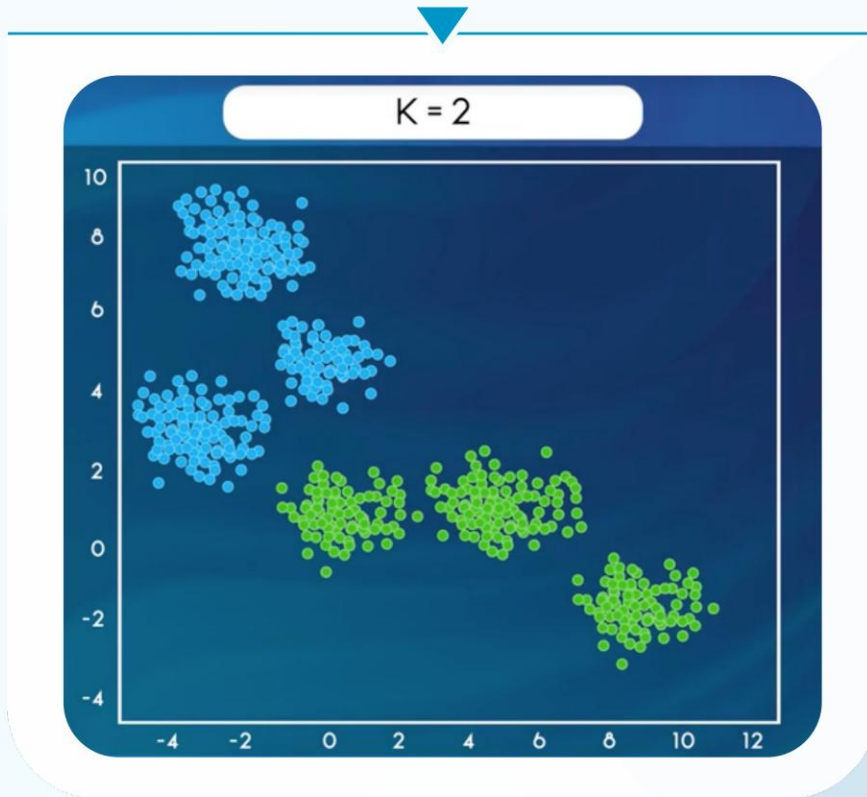
K-arama algoritması nedir?

K-means, veri noktalarını "K" sayı da gruba bölündüğü bir kümeleme yaklaşımıdır.

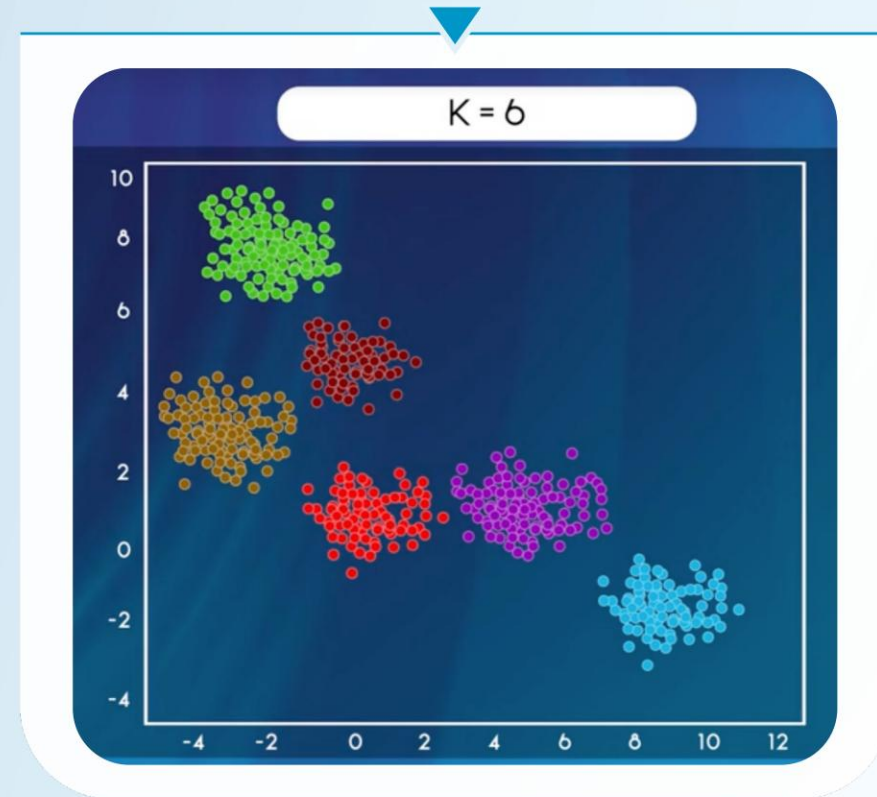
K-means algoritması belirli bir merkeze yakını olan veri noktalarını bir arada gruplandırır.

K-ortalama algoritması

Az sayı da
K = daha büyük kümeler



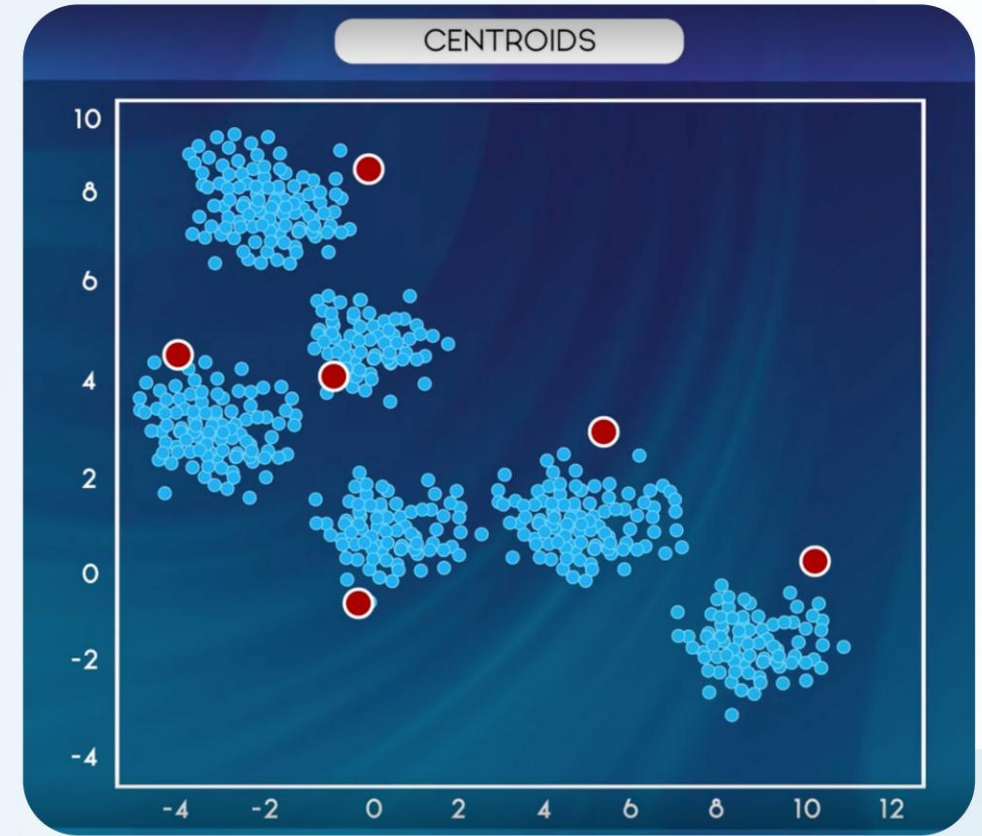
Çok sayı da
K = daha küçük kümeler



K-aracı nda merkez noktaları n konumları nı optimize edilmesi: merkezoidler

İlk olarak, rastgele seçilen bir grup küme merkez noktası , merkezoidler, her küme için başlangı ç noktaları olarak kullanı lı r.

Daha sonra ağı rlı k merkezlerinin konumları nı optimize etmek için yinelemeli hesaplamalar gerçekleştirilir.



Kümeler için performans değerlendirme metrikleri:

siluet katsayı s

Siluet katsayı s nedir?

- Siluet katsayı s veya siluet puanı , kümeleme tekniğinin iyiliğini hesaplamak için kullanılan bir ölçümdür.
- Bir noktayı n aynı kümedeki diğer noktalara olan ortalama uzaklığı a nı , aynı noktayı n en yakını kümedeki noktalara olan ortalama uzaklığı b yla karşılaştırır.

$$s = \frac{b-a}{\max(a,b)}$$

a : The mean distance between a sample and all other points in the same cluster.

b : The mean distance between a sample and all other points in the next nearest cluster.