

## K-Means Algoritması

K-means algoritması bir kümeleme algoritmasıdır. **Kümeleme** algoritmaları otomatik olarak verileri daha küçük kümeler ya da alt kümeler ayırmaya yarayan algoritmalar. 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 Karasal Hata Fonksiyonunu en aza indirgiyecek “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.

### YİĞİNSAL (AGGLOMERATIVE) KÜMELEME ALGORİTMASI

Hiyerarşik kümelemede yukarıdan aşağıya doğru önceden belirlenmiş sıraya sahip kümeler oluşturulur. Agglomerative Clustering ‘de bunun bir türüdür. K-Means’ten farkı, kaç küme oluşturulacağı belirtilmez. Başlangıçta bütün örnekler (samples) birer küme olarak kabul edilir. Daha sonra her bir küme kendisine en yakın diğer bir küme ile birleşir. En sonunda ortaya dendrogram şeması ve tek bir küme çıkar. **Dendrogram**, kümeler arasındaki hiyerarşik yapıyı göstermek için kullanılan ağaç şemasıdır. Dendrogram şeması yorumlanarak, kaç tane küme oluşturulacağı belirlenir.

### Expectation-maximization

İstatistikte, Beklenti maksimizasyon algoritması, gözlemlenemeyen gizli değişkenlere bağlı istatistiksel modellerin parametrelerinin en büyük olabilirlik ya da en büyük artçıl tahminlerinin bulunması için kullanılan bir yinelemeli arama yöntemidir. Beklenti maksimizasyonu, beklenti (B) adımı ve maksimizasyon (M) adımı olarak iki adımın art arda tekrarlanmasıyla gerçekleşir. B-adımı parametrelerin o anki tahminlerini kullanarak bir log-olabilirlik beklentisi fonksiyonu oluşturur. M adımı parametre değerlerini log-olabilirlik beklentisini maksimize edecek şekilde günceller. Yani bu iki adımın her biri diğerinin girdisini hesaplayarak birbirini besler. Beklenti maksimizasyon adımları tahmindeki hata miktarı belirli bir oranın altına düşene kadar yinelenir.

### Kaynaklar:

- 1- [https://tr.wikipedia.org/wiki/Beklenti\\_maksimizasyon](https://tr.wikipedia.org/wiki/Beklenti_maksimizasyon)
- 2- <https://medium.com/@ekrem.hatipoglu/machine-learning-clustering-k%C3%BCmeleme-kmeans-algorithm-part-13-be33aeef4fc8>
- 3- <https://www.datanovia.com/en/lessons/agglomerative-hierarchical-clustering/>