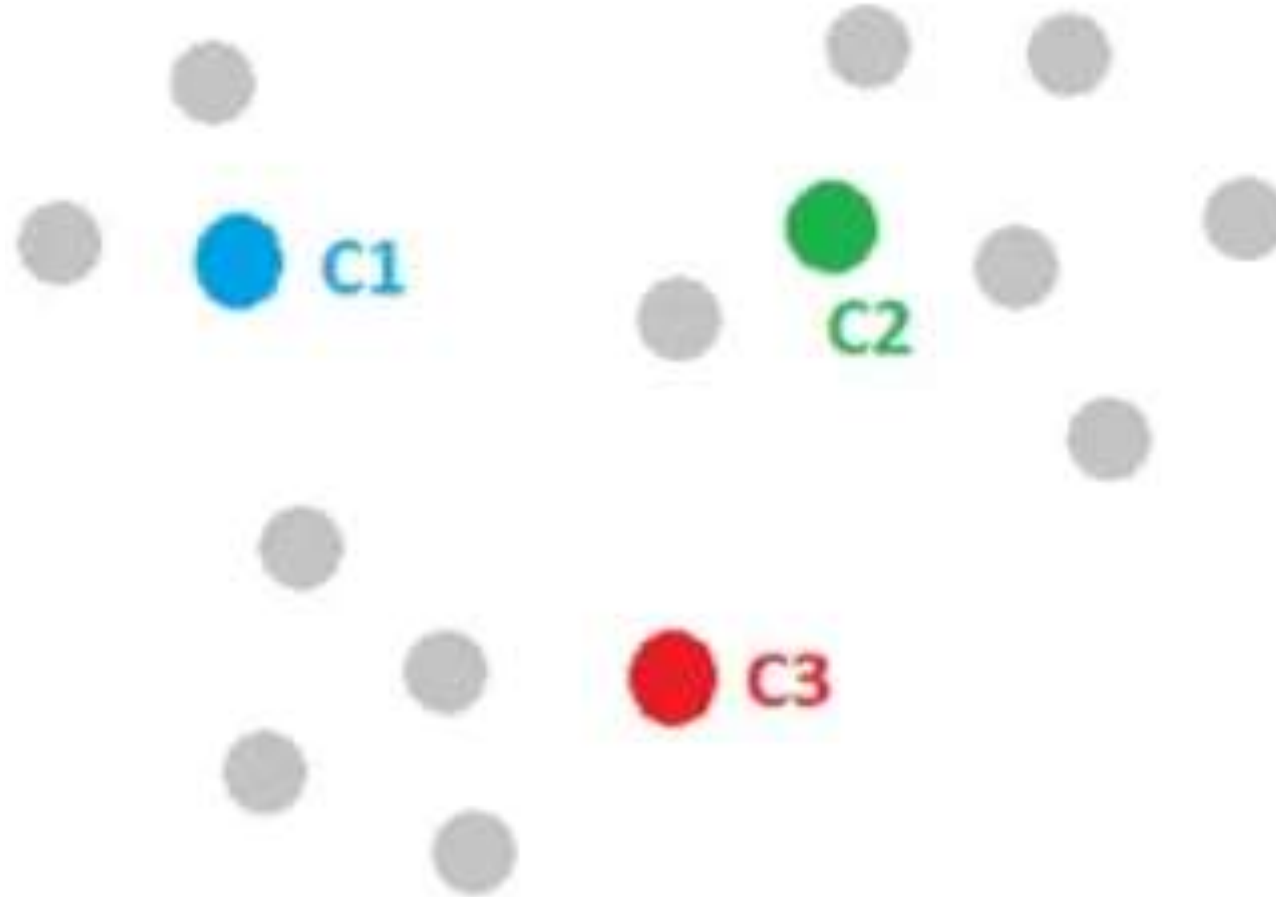


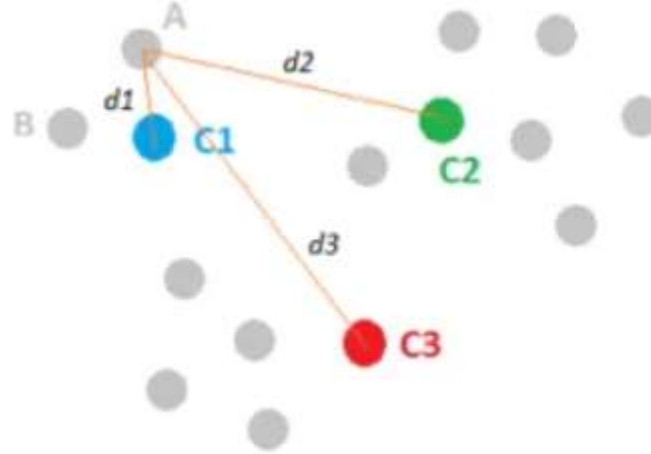
KÜMELEME ALGORİTMASI

- Kümeleme (Clustering) , makine öğrenmesi konseptlerinden biri olan Denetimsiz Öğrenme (Unsupervised Learning) için önemli bir kavramdır. Kümeleme algoritmaları basitçe veri kümesindeki elemanları kendi arasında gruplamaya çalışır. Burada kaç grup olacağı bizim inisiyatifimizde olan bir bilgi de olabilir veyahut en uygun küme sayısını algoritmanın kendisi de belirleyebilir.

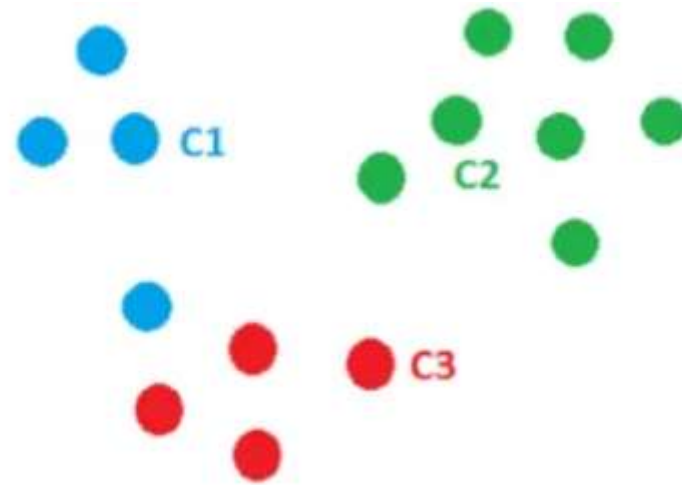
İlk adımda kümeler için bir merkez noktası rastgele olarak belirlenir.



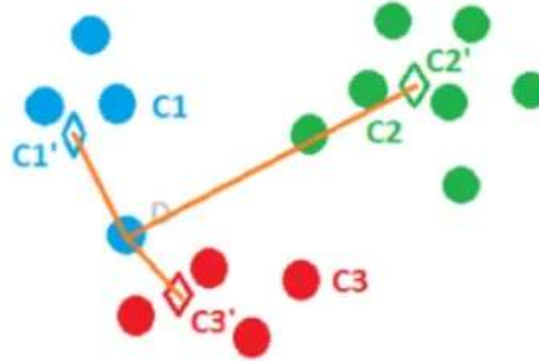
Merkez eleman dışındaki elemanlar en yakın merkez noktaya göre kümelenir. Bunun için her bir elemanın merkez noktalara olan uzaklıkları ölçülür.



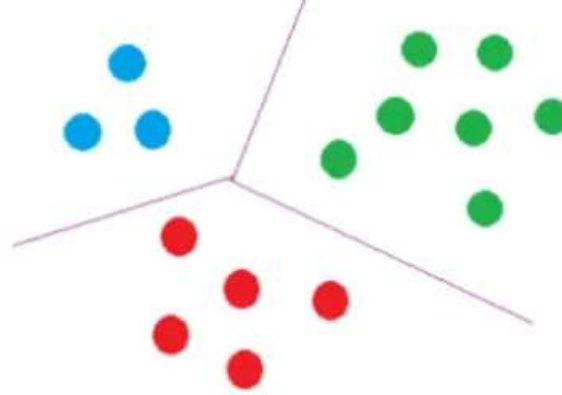
Kümelenen elemanlarımız böyle bir hal alır.



Ardından aynı kümeler içerisinde tekrar bir merkez noktası hesaplaması yapılır.



Ardından merkez noktası stabil hale gelinceye dek ilk iki adım tekrarlanır.



- Bunun için önce bir kümeleme algoritmasının başarı kriterini konuşmamız gerekir. Şayet bir kümeleme algoritması küme elemanları arasındaki mesafeyi ne kadar minimum ve her bir küme arasındaki mesafeyi de ne kadar maksimum tutabiliyorsa o kadar başarılı bir kümeleme yapmıştır diyebiliriz. Bu kriteri küme sayısını olabildiğince artırarak gerçekleştirebilir miyiz?.

- Aslında cevap evet. Küme sayısı arttıkça küme içindeki her bir eleman arası mesafe azalmış olacaktır ve WCSS değeri de bununla beraber azalmış olacaktır. O halde neden WCSS değerini 0 yapana kadar küme eklemiyoruz düşüncesi aklımıza gelebilir. Bunun sonucunda da her bir eleman kendi başına birer küme olarak karşımıza çıkacak ve modelimiz overfittinge yani ezberleme durumuna uğrayacaktır. Peki o zaman optimum noktayı nasıl belirleyebiliriz?. Bunun için WCSS grafiğinde dirsek noktası diye tabir edilen bir değere bakmamız yeterlidir. Şöyle ki:

