

Clustering(kümeleme)

Bir nesne kümesini gruplama problemidir. Bu problemlerde nesnelerin aynı kümede yer alması için birbirine daha benzer olması gerekmektedir.

K-Means Clustering algoritmasını kullanacağız. K ortalımalı kümeleme demektir.

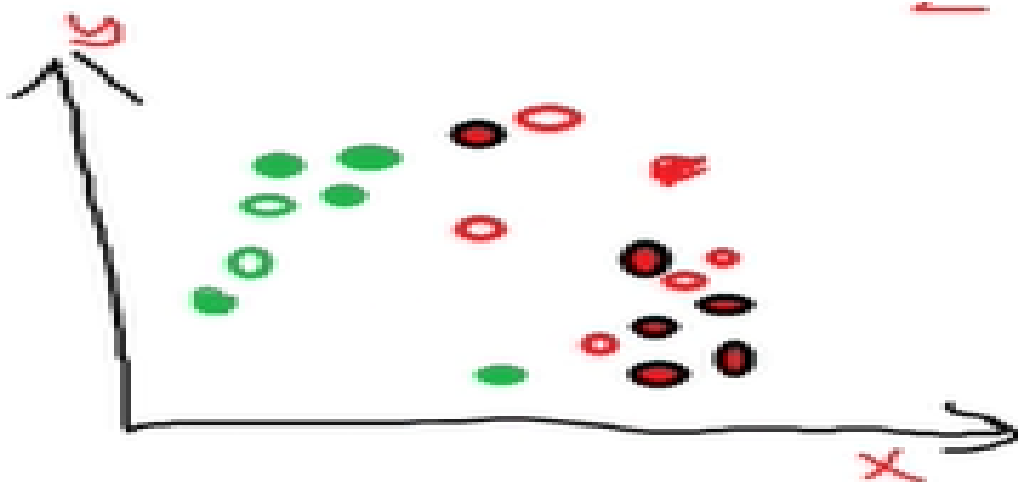
Elimizde bir tane veri seti var ve buradaki veri setini kümelemek istiyoruz. Nasıl?

İlk önce K değerinin ne olduğunu belirlemeliyiz. K değeri bizim kaç tane kümemizin olduğunu ifade eder. Biz K değerini 2 vererek kullanacağız. Sayısal bir veri seti olarak ele alıyoruz. Bu veri setinde label'leri bilmediğimiz ve kümeleme yapmadığımız bir tane data var. Bu datayı K-Means Kümeleme algoritmasını kullanarak 2 tane kümeye ayırmaya çalışacağız.

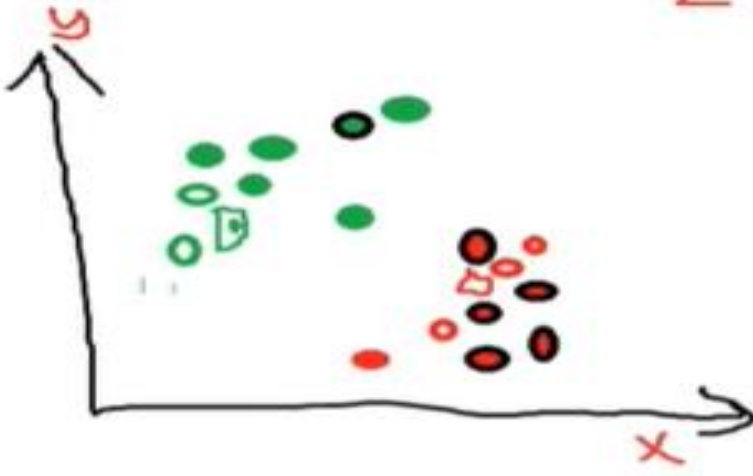
1. **Adım:** olarak veri setinde bir tane centroid seçmek gerek, random yani rast gele seçeceğiz. 1.si kırmızı 2.si yeşil olucak.



2. **Adım:** Euclidean distance bakmamız gerekiyor. Uzaklıklara göre kümelerini seçtik.



3. **Adım:** Mean recompute centroid; aynı renkte olanların ortalamasını al ve yeni bir centroid hesapla.
4. Kümelerin içindeki garip şekil çizip yeni centroid'ler oluşturduk ve tekrardan euclidean distance'leri hesaplıyoruz.



5. **Adım:** Tekrardan 3. Adıma geri döndük, centroid'i recompute yapıcaz. Daha önceki centroidleri silip, tekrardan centroidleri hesaplıyoruz. Kırmızı grubun centroid'i tam ortasına gelicek, yeşil grubun centroid'i de tam ortasına gelicek. Artık hiçbir euclidean distance sonucuna göre bunlar renk değiştirmicek, yeşiller yeşil, kırmızılar kırmızı kalıcak.

