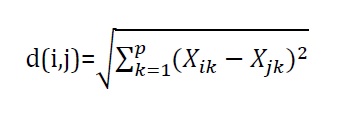
**K ORTALAMAR ALGORİTMASI İLE KÜMELEME**

Bu yöntem hiyerarşik olmayan kümeleme yöntemleri arasında büyük önem taşır. Daha başlangıçta belli sayıdaki küme içi toplam ortalama hatayı minimize etmek amaçlanır. Bu yöntemde başlangıçta k parametresince merkez noktası oluşturarak bu merkez noktaları üzerinden gözlem değerlerinin her bir merkeze olan uzaklığının bulunmasıyla yeni üyelik elde ederek küme için değişimleri hesaplayarak kümeleme işlemini tamamlamak esastır.

* Her bir gözlem noktasının her bir merkeze olan uzaklığının hesaplanmasında kullanılan ***formül 1***:



* N boyutlu uzayda N örnekli kümelerin verildiğini varsayalım. Bu uzay {C1,C2,…Ck} biçiminde k adet kümeye ayrılsın. O zaman ∑nk = N (k=1,2,…k) olmak üzere Ck kümesinin ortalama vektörü (Mk) hesaplamasında kullanılan **formül 2**:
* Burada Xk değeri Ck kümesine ait olan i. Örnektir. Ck kümesi için hata-kare, her bir Ck örneği ile onun merkezi (centroid) arasındaki Öklid uzaklıkları toplamıdır. Bu hataya “küme içi değişme” adı da verilir. Küme içi değişmelerin hesaplanmasında kullanılan **formül 3**:
* K kümesini içeren bütün kümeler uzayı için hata-kare, küme içindeki değişmelerin toplamıdır. Toplam kare hata hesaplanırken kullanılan **formül 4**:

***K Ortalamalar Akış Şeması***

K ortalamalar için takip edilmesi gereken akış şeması aşağıdaki görselde yer almaktadır.

******

***K Ortalamalar Algoritması***

1. Algoritmaya başlamadan önce k küme sayısının belirlenmesi gerkir. Söz konusu k değeri belirlendikten sonra k adet merkez belirlenir ve bu merkezlerin değerlere olan uzaklığına göre her gözlem değerine küme üyeliği atanır.
2. Herbir kümenin merkezi belirlenir. Bu merkezler M1,M2,…Mk biçimindedir.
3. E1,e2,…ek küme içi değişmeler hesaplanır. Bu değişmelerin toplamı olan Ek2 toplam kare hata bulunur.
4. Mk merkez değerleri ile gözlem değerleri arasındaki uzaklıklar hesaplanır. Bir gözlem değeri hangi merkeze yakın ise, o merkez ile ilgili küme içine dahil edilir.
5. Yukarıdaki c ve d adımları, kümelerde herhangi bir değişiklik olmayıncaya dek sürdürülür.

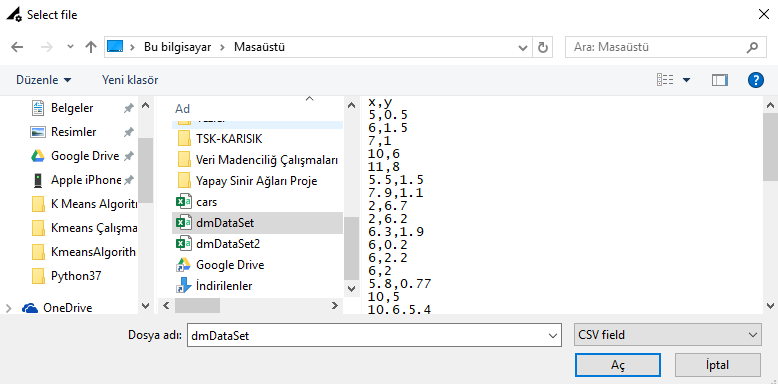
**K ORTALAMALAR UYGULAMASI ÖRNEK**



**Şekil** 1.1 Uygulama Giriş Arayüzü

Şekil 1.1’ de verilen görselde uygulamanın giriş arayüzü yer almaktadır. Bu uygulama Python programlama dilinin Python 3.7. versiyonu ile geliştirilmiştir. Arayüz kütüphanesi olarak Tkinter kütüphanesi kullanılmıştır. Bu uygulama iki boyutlu sayısal veriler üzerinde çalışmaktadır.

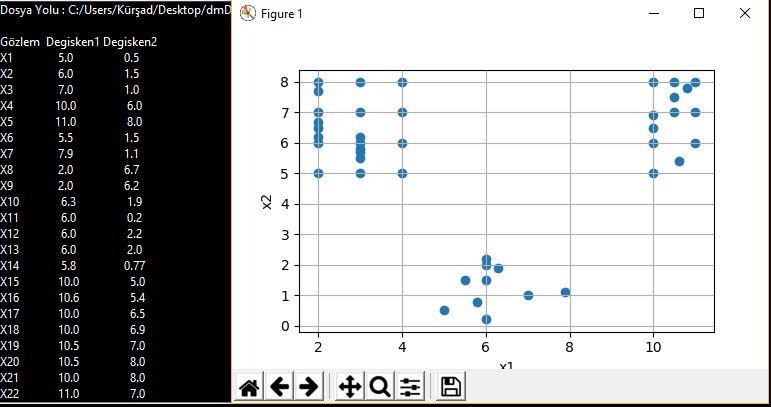
Öncelikle yapılması gereken “Dosya yolunu belirleyiniz” butonu kullanılarak uygulama içerisine gözlem değerlerinin import edilmesi gerekmektedir. Dosya formtı “.csv” formatında olmalıdır ve yukarıda da belirtildiği üzere bu uygulama iki boyutlu veriler üzerinde çalışmaktadır.



**Şekil** 5.2 Dosya Seçme Arayüzü

**Şekil 5.2 de yer alan arayüzde uygulama içerisine verilen import edilişi gösterilmektedir.**

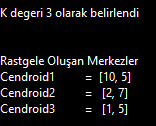
Veriler import edildikten sonra gözlem değerleri iki boyutlu grafik üzerinde yerleşimleri görsel olarak görülebilmektedir.



**Şekil** 5.3 Gözlem Değerlerini Gösteren Pyplot Grafiği

Şekil 5.3’de gözlem değerleri import edildiğinde uygulama giriş arayüzünde yer alan listbox üzerinde verilerin gösterildiği ve iki boyutlu uzay üzerinde gözlem değerlerinin görünümü verilmektedir.

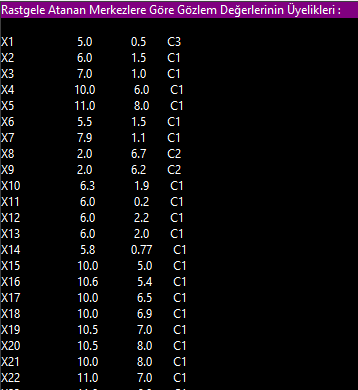
**Bu aşamadan sonra kullanıcının k parametresini belirlemesi gerekmektedir bu k parametresine göre k adet merkez noktası rastgele olarak oluşturulmalıdır. Bu uygulamada k parametresi 3 olarak belirlenmiştir.**



**Şekil** 5.4

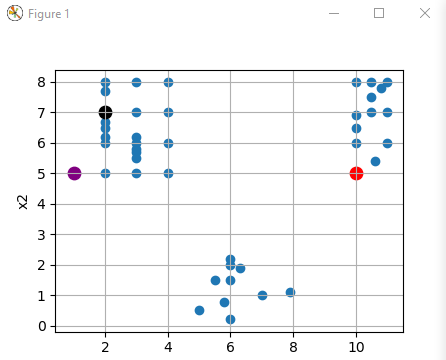
Şekil 5.4’de görüldüğü üzere 3 adet rastgele merkez noktaları oluşturulmuştur.

Şekil 5.5’de k parametresi 3 olarak belirlendikten sonra 3 adet merkez oluşturularak her bir gözlem değerinin bu merkezlere olan uzaklıkları hesaplanmıştır sonucunda da en yakın merkez noktasının indeksi alınarak ilk küme üyeliği olarak belirlenmiştir. Ardın dan uygulama giriş arayüzünde yer alan listboxa bu ilk sınıf üyelikleri bastırılmıştır.

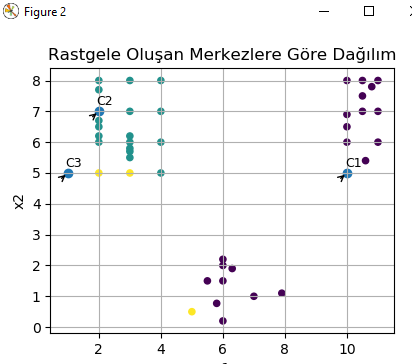


**Şekil** 5.5 İlk Sınıf Üyelikleri (ilk22 gözlem değeri)

Bu yeni üyeliklerinin oluşumuyla gözlem değerleri yakın buldukları merkez noktalarının küme üyeliklerini alırlar. Şekil 5.6’da rastgele oluşan merkez noktalarının iki boyutlu uzayda yerleşimini gösteren görsel yer almaktadır.



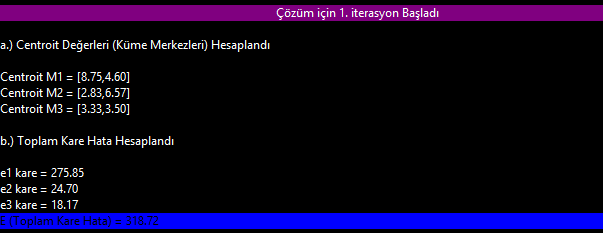
**Şekil** 5.6 Centroidlerin İki Boyutlu Uzayda Yerleşimi



**Şekil** 5.7 Gözlem Noktalarının İlk Küme Üyelikleri

Şekil 5.6’ da her bir gözlem değerinin merkezlere olan uzaklıklarının karşılaştırılması ile yakın oldukları merkezlerin üyeliklerini aldığı grafik yer almaktadır.

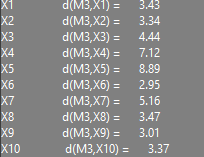
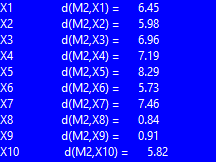
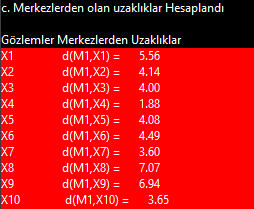
**Rastgele oluşturulan merkezlere her bir objenin uzaklığı bulunup en yakın merkez noktasının küme üyeliği her bir gözlem değerine ayrı ayrı verildikten sonra K Ortalamalar uygulamasının esas çözümlemesi başlamaktadır.**



**Şekil** 5.8 Yeni Merkez Noktaları ve Toplam Kare Hata

Şekil 5.5’de yer alan gözlem değerlerine ait ilk sınıf üyeliklerinden yararlanılarak x1 ve x2 gözlem değerleri güncel haline geçmiş bulunmaktadır.

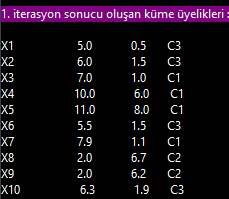
Şekil 5.8’ de oluşan yeni gözlem değerleri üyeliklerinden yararlanılarak yani güncel verilerden yararlanılarak Formül 1 adlı şekilde formül ile yeni merkez noktaları hesaplanmıştır. Toplam kare hata hesabı ise formül 3 ve formül 4 yardımı ile hesaplanmıştır. 1. İterasyon sonucunda toplam kare hatanın oldukça yüksek olduğu görülmektedir. Amaç bu toplam kare hatayı olabildiğince indirgemektir.



**Şekil** 5.9 Merkezlerden Uzaklıklar

Yeni belirlenen merkezler referans alınarak her bir gözlem değerinin her bir merkez noktasına olan uzaklıkları şekil 5.9’da yer almaktadır. Merkez noktası ile arasında en kısa mesafe olan sonuç satıra bakılarak ait merkezin indeksi ilgili gözlem noktasının yeni sınıf üyeliği olarak nitelendirilmektedir.

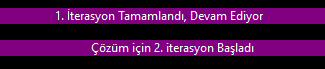
X1 gözlem noktasının m1,m2,m3 merkez noktalarına olan uzaklıklarını kıyasladığımızda yeni sınıf üyeliğinin “3” değeri olması beklenmektedir.



**Şekil** 5.10 Yeni Sınıf Üyelikleri

Her bir gözlem noktası için merkez noktaları arasındaki uzaklıklar karşılaştırıldığında yapılan hesaplamalar sonucu şekil 5.10’da yer alan sonuçlar belirlenmiştir.

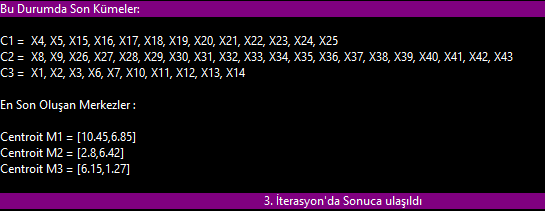
Bu sonuçları yorumlayacak olursak ilk 5 gözlem değeri için: “ 3 , 1 , 1 , 1 , 1” şeklinde olan ilk küme üyelikleri yapılan yeni hesaplamalar sonucunda “ 3 , 3 , 1 , 1 , 1” olarak belirlenmiştir.

****

**Şekil** 5.11 1. İterasyon Son

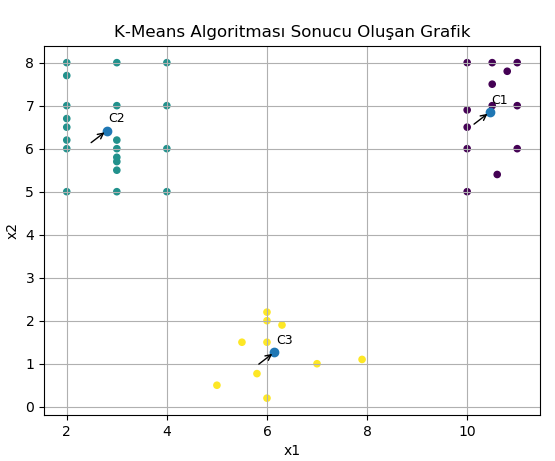
**Birinci iterasyon burada tamamlanmıştır. Yukarıda gözlemlenen sınıf üyeliği değişikliği ve ya merkez noktası değişikliği duruncaya, sabitleninceye dek bu iterasyon aşamaları tekrarlanmaya devam eder.**

1. İterasyonu yorumladığımızda rastgele belirlenen merkez noktaları : m1=[10,5] , m2=[2,7] , m3=[1,5] şeklindedir , 1. İterasyonda ise oluşan merkezler : m1=[8.75,4.60] , m2=[2.83 , 6.57] , m3=[3.33 , 3.50] şeklindedir. Bu durumda merkezlerde gelişen değişiklikten dolayı iterasyon devam etmelidir. Ta ki bu değişiklik duruncaya dek.



**Şekil** 5.12 Oluşan Son Merkez Noktaları

Şekil 12’de yapılan daha önceki iterasyonlardan sonra oluşan merkez noktalarına bakılara 3. İterasyonda artık bir değişiklik olmadığı yani centroidlerin hareket etmediği gözlenmektedir. Bu nedenle algoritma 3. İterasyonda sonuca ulaşmıştır.



**Şekil** 5.13 Son Grafik

Şekil 13’de oluşan son üyelikler, merkez noktaları ve gözlem değerlerini içeren grafik yer almaktadır. Sonuçla k means uygulaması beklenen gruplama işlemini başarı ile tamamlamıştır.