

2018
Recep KARADEMİR
2015141003

K-En Yakın Komşu Algoritması ile Iris Veri Setini
Sınıflandırma

Özet ve Açıklamalar

İRİS : Süsengiller (Iris), Asparagales takımına ait bir çiçekli bitki familyasıdır.

Iris familyasındaki üç türün temel özelliklerinden oluşan veriler yardımıyla, en iyi tür tahmini yapabilen K (KNN) tespiti yapabileceğimiz program yazılmıştır.

Bu proje C++ programlama dilinde, K-En Yakın Komşu algoritması kullanılarak hazırlanmıştır.

Program iris.txt içindeki iris çiçeği verilerini üç sete ayırarak işlem yapmaktadır.

İlk iki seti programı eğitime amaçlı kullanan, üçüncü seti ise koruyarak test amaçlı tutan bir program yazılmıştır.

Okuduğu veri setini istersek karıştırarak işleyebiliriz (her veri setinde her iris sınıfından örnekler bulunması için).

Karıştırılan veya karıştırılmayan veri seti için setlere ayırma şöyle yapılır :

İlk seti eğitim seti olarak saklar. İlk set toplam veri setinin %50 sini oluşturacak şekilde ayrılır.

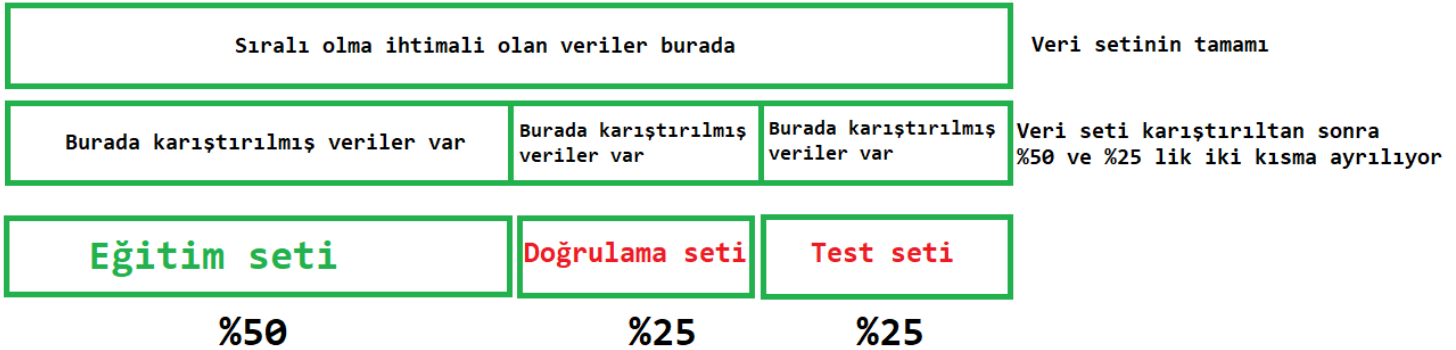
Rastgele karıştırılan verinin ilk % 50 lik kısmı program tarafından eğitim seti olur.

ikinci seti toplam veri setinin %25'ini (%50 lik kısmın üstündeki %25 lik kısım) ve

üçüncü seti toplam veri setinin %25(%75 den sonraki kalan %25 lik kısım) 'ini oluşturacak şekilde ayırır.

İlk %50 lik set eğitim setimizdir. İkinci %25 lik set doğrulama setimizdir. Üçüncü %25 lik set test setimizdir.

Görselleştirmek gerekirse :



Doğrulama setinin ilk indisi, 150 adet veri içeren veri setinde 75.indis olacaktır.

Test setinin ilk indisi , 150 adet veri içeren veri setinde 113.indis olacaktır.

Eğitim seti [0..74] verileridir. Doğrulama seti [75..112] verileri olacaktır. Test seti [113..149] indislerdeki veriler olur.

Veriler txt içerisinde aralarında boş satır olmayacak şekilde bulunmalıdır.

Her satırdaki iris özellikleri aralarına virgül koyularak ayrılmış olmalıdır.

Örnek :

5.1,3.5,1.4,0.2,Iris-setosa Aralarda virgül var. Doğru veri seti böyle kaydedilmiş olmalı.

6.0,2.2,4.0,1.0,Iris-versicolor

4.7,3.2,1.3,0.2,Iris-setosa

7.3,2.9,6.3,1.8,Iris-virginica

Okunan txt verilerinden Iris adları bellekte işlem kolaylığı olması amaçlı 0 ,1 ,2 olacak şekilde kodlanarak tutuluyor.

0 kodu iris-setosa , 1 kodu iris-versicolor , 2 kodu iris-virginica olarak bellekte tutulacaklar.

Program içerisinde iris adlarının kodları yerine adların kendilerini metin olarak göreceğiz.

Veri setinin mantıklı ve yüksek tahmin oranında işlenmesi için verilerin değerleri korunarak

bütün veri seti 3 defa rasgele karıştırılmıştır.

Örnek 1.0) Veriler sıralı. Karıştırılması gerekli.

```
1 : 5.1000 3.5000 1.4000 0.2000 Iris-setosa
2 : 4.9000 3.0000 1.4000 0.2000 Iris-setosa
3 : 4.7000 3.2000 1.3000 0.2000 Iris-setosa
4 : 4.6000 3.1000 1.5000 0.2000 Iris-setosa
5 : 5.0000 3.6000 1.4000 0.2000 Iris-setosa
6 : 5.4000 3.9000 1.7000 0.4000 Iris-setosa
7 : 4.6000 3.4000 1.4000 0.3000 Iris-setosa
8 : 5.0000 3.4000 1.5000 0.2000 Iris-setosa
```

Solda veri setinden birkaç kesit koydum.

```
49 : 5.3000 3.7000 1.5000 0.2000 Iris-setosa
50 : 5.0000 3.3000 1.4000 0.2000 Iris-setosa
51 : 7.0000 3.2000 4.7000 1.4000 Iris-versicolor
52 : 6.4000 3.2000 4.5000 1.5000 Iris-versicolor
53 : 6.9000 3.1000 4.9000 1.5000 Iris-versicolor
54 : 5.5000 2.3000 4.0000 1.3000 Iris-versicolor
55 : 6.5000 2.8000 4.6000 1.5000 Iris-versicolor
56 : 5.7000 2.8000 4.5000 1.3000 Iris-versicolor
57 : 6.3000 3.3000 4.7000 1.6000 Iris-versicolor
58 : 4.9000 2.4000 3.3000 1.0000 Iris-versicolor
59 : 6.6000 2.9000 4.6000 1.3000 Iris-versicolor
```

```
97 : 5.7000 2.9000 4.2000 1.3000 Iris-versicolor
98 : 6.2000 2.9000 4.3000 1.3000 Iris-versicolor
99 : 5.1000 2.5000 3.0000 1.1000 Iris-versicolor
100 : 5.7000 2.8000 4.1000 1.3000 Iris-versicolor
101 : 6.3000 3.3000 6.0000 2.5000 Iris-virginica
102 : 5.8000 2.7000 5.1000 1.9000 Iris-virginica
103 : 7.1000 3.0000 5.9000 2.1000 Iris-virginica
104 : 6.3000 2.9000 5.6000 1.8000 Iris-virginica
105 : 6.5000 3.0000 5.8000 2.2000 Iris-virginica
```

Örnek 1.0) Karıştırılmış veri setinin görüntüsü :

```
1 : 4.4000 3.0000 1.3000 0.2000 Iris-setosa
2 : 5.6000 2.8000 4.9000 2.0000 Iris-virginica
3 : 4.8000 3.0000 1.4000 0.1000 Iris-setosa
4 : 7.7000 2.8000 6.7000 2.0000 Iris-virginica
5 : 4.8000 3.4000 1.6000 0.2000 Iris-setosa
6 : 5.0000 2.3000 3.3000 1.0000 Iris-versicolor
7 : 4.4000 2.9000 1.4000 0.2000 Iris-setosa
8 : 5.1000 3.8000 1.6000 0.2000 Iris-setosa
9 : 5.8000 2.7000 3.9000 1.2000 Iris-versicolor
10 : 5.4000 3.7000 1.5000 0.2000 Iris-setosa
```



Veri setinden bir kesit.

Karıştırma işlemi verilerin sadece satırlar arasında yer değiştirmesi olarak yapılmıştır.

Her bir satırdaki iris çiçeği özellikleri korunmuştur. iris.txt deki veri setini , programımızın içinde tutan çok boyutlu dizi karıştırılmıştır.

Program, verilerin güncel halini, bellekteki tuttuğu diziyi ekrana yazdırarak gösterir.

Txt den okunan verilerin değerlerinin tahmine aynı oranda etki etmesini, normalizasyon yaparak sağlarız.

Normalizasyon yaparsak daha yüksek doğrulukta tahmin sonuçları da elde edebiliriz.

Bazen normalizasyon tahmin oranını düşürebilir.

Bu nedenle veri özelliklerinde normalizasyon yapılacaksa, özelliklerin birbirleriyle eşit veya benzer bir öneme sahip olması gerekir.

Örnek : Aşağıdaki iris verisi verilsin.

5.1 , 3.5 , 1.4 , 0.2 , Iris-setosa

Bu veride 5.1 in 0.2 ile tahmine etkisi aynı oranda olmayacaktır. Farklar yardımıyla tahmin yapılırken 5.1 için -1 lik bir fark değeri = 4.1 yaparken 0.2 için aynı -1 lik bir fark değeri = -0.8 yapacaktır.

0.2 özelliği bu değişimden %400 etkilenirken , 5.1 özelliği bu değişimden %19 etkilenir.

Bu farkı kapatmak özelliklerin birbirleriyle eşit veya benzer öneme sahip olduğunda kritik önem taşır.

Doğrudan eğitim setinden mesafe ölçümlerinin hesaplanmasında bir büyük dezavantaj, değişkenlerin farklı ölçüm ölçeklerine sahip olması ya da sayısal ve kategorik değişkenlerin bir karışımı olması durumunda ortaya çıkmaktadır. Örneğin, bir değişken dolar cinsinden yıllık geliri temel alırken diğeri yıllara dayanıyorsa, gelirin hesaplanan mesafe üzerinde çok daha fazla etkisi olacaktır. Bir çözüm, aşağıda gösterildiği gibi bütün veri setini standartlaştırmaktır.

Programımızda kullanılan standartizasyon formülü şöyledir:

Bütün veri setindeki verilerin minumum değeri = veriMin

Bütün veri setindeki verilerin maximum değeri = veriMax

Normalize olacak veri = veri

normalize olmuş hali = normVeri

$$\text{normVeri} = \frac{\text{veri} - \text{veriMin}}{\text{veriMax} - \text{veriMin}}$$

Her bir iris özelliği için veri min ve veri max değeri değişiklik gösterir.

Program bu değerleri veri normalizasyonunda tespit eder ve her bir farklı veri niteliği için o niteliğe özel min ve max değerlerini kullanarak normalizasyon yapar. **Örn:** sepalL min ile sepalW verisini normalize edilmeye çalışılmamalıdır.

K deęerinin incelenmesi

Hata oranının artması, veri setindeki özellik deęerlerinin birbirine yakın olması veya k deęerinin bir iris türü adının (her tür için 50 tane var) adet sayısına(50ye yakın) yakın veya üzeri bir deęer olmasına baęlıdır.

K için 37-38-39-40-41 gibi üzeri her deęer programın tahmininde hata sayısının çoęalmasına neden olur.

Bunun sebebi tür adetinden fazla k deęeri girildiğinde komşuluk farkının başka türleri de kapsamasıdır.

Bu bilgi detaylı olarak "En iyi K tespiti" başlığında incelenecektir.

Yukarıda anlatmak istediğim iris veri setine uygulanırsa:

Iris tür adı adet sayısı = Iris-setosa 50 tane , Iris-versicolor 50 tane , Iris-virginica 50 tane

K deęeri 50 ye yaklaştıkça hata oranı artar.

K deęeri 50 yi geçince hata oranı yüksek bir hızla artar.

Veri setindeki iris özellikleri yakın(benzer) deęerler içerirse de program tür ayrımı yapmakta zorlanacaktır.

Ayrım yapmak zorlaşınca da tahmin oranı aşağı düşecektir.

Örnek :

5.1 , 3.7 , 1.5 , 0.4 , Iris-setosa

5.2 , 2.7 , 3.9 , 1.4 , Iris-versicolor

Bu sütundaki bilgi programda tür farkı için kullanılmaz. Aranan bilgi fark tespiti için kullanılamaz. Doğrulama setiyle tahmin verilerinin kıyaslanması için saklanır.

Bingo! Çok yüksek fark. Hatalı tahmin oranını azaltacak. Veri türünün fark tespiti için önemli.

Bingo! İşte yüksek fark. Bu tarz farklılıklar veri türlerini doğru tahmin etmemize yarar.

Bu fark ayırt etmede biraz yardımcı olacak, hatalı tespiti azaltacaktır.

Deęerler çok yakın ayırt edici bir fark için yardımcı olmaz.



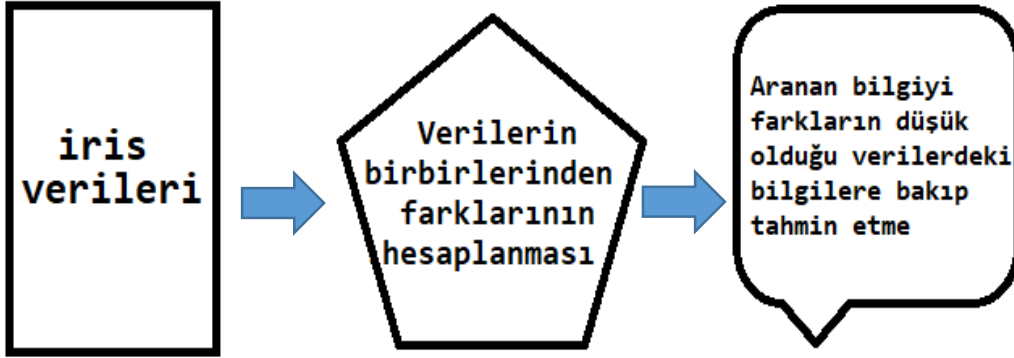
Yüksek farklar olmasaydı veri türlerini ayırmak çok zor olurdu ve doğru tahmin oranı da düşük olurdu.

Programın incelenmesi

Verilerin farklarına bakılarak doğru tahmin nasıl yapılıyor ?

Temel olarak şu şekilde program anlatılabilir.

TXT VERİ



Doğrulama setindeki verilerin her biri için eğitim setindeki bütün değerlerle farkı alınıp saklanıyor. **150 adet veri için:**

Görselden kolay anlaşılacağından, önce verilerin her zaman farklı değerlerle aşağıdaki gibi dizileceğini söyleyelim.

| | | | | | | | |
|-----|---|--------|--------|--------|--------|-----------------|---|
| 1 | : | 0.6111 | 0.5000 | 0.6949 | 0.7917 | Iris-virginica | <-- 0. indis 1.eğitim seti verisi |
| 2 | : | 0.5556 | 0.2083 | 0.6610 | 0.5833 | Iris-versicolor | <-- 1. indis 2.eğitim seti verisi |
| 3 | : | 0.5556 | 0.2083 | 0.6780 | 0.7500 | Iris-virginica | . |
| 4 | : | 0.6111 | 0.4167 | 0.8136 | 0.8750 | Iris-virginica | . |
| 5 | : | 0.3333 | 0.1667 | 0.4576 | 0.3750 | Iris-versicolor | . |
| 6 | : | 0.1389 | 0.5833 | 0.1525 | 0.0417 | Iris-setosa | . |
| 7 | : | 0.6389 | 0.3750 | 0.6102 | 0.5000 | Iris-versicolor | <-- 6. indis 7.eğitim seti verisi |
| . | | | | | | | |
| . | | | | | | | |
| 70 | : | 0.5000 | 0.4167 | 0.6102 | 0.5417 | Iris-versicolor | <-- 69. indis 70.eğitim seti verisi |
| 71 | : | 0.2222 | 0.7500 | 0.0847 | 0.0833 | Iris-setosa | <-- 70. indis 71.eğitim seti verisi |
| 72 | : | 0.1667 | 0.4583 | 0.0847 | 0.0000 | Iris-setosa | . |
| 73 | : | 0.3056 | 0.7917 | 0.1186 | 0.1250 | Iris-setosa | . |
| 74 | : | 1.0000 | 0.7500 | 0.9153 | 0.7917 | Iris-virginica | . |
| 75 | : | 0.5556 | 0.5417 | 0.6271 | 0.6250 | Iris-versicolor | <-- 74. indis 75.eğitim seti son verisi |
| 76 | : | 0.2222 | 0.7500 | 0.1525 | 0.1250 | Iris-setosa | <-- 75.indis 1. doğrulama seti verisi |
| 77 | : | 0.7500 | 0.5000 | 0.6271 | 0.5417 | Iris-versicolor | <-- 76.indis 2. doğrulama seti verisi |
| 78 | : | 0.4722 | 0.2917 | 0.6949 | 0.6250 | Iris-versicolor | . |
| 79 | : | 0.4722 | 0.4167 | 0.6441 | 0.7083 | Iris-virginica | . |
| . | | | | | | | |
| 109 | : | 0.0833 | 0.5000 | 0.0678 | 0.0417 | Iris-setosa | . |
| 110 | : | 0.6667 | 0.4583 | 0.6271 | 0.5833 | Iris-versicolor | . |
| 111 | : | 0.3889 | 0.3750 | 0.5424 | 0.5000 | Iris-versicolor | . |
| 112 | : | 0.7222 | 0.4583 | 0.6610 | 0.5833 | Iris-versicolor | . |
| 113 | : | 0.5833 | 0.3333 | 0.7797 | 0.8333 | Iris-virginica | <-- 112.indis 38. doğrulama seti son veri |
| 114 | : | 0.5556 | 0.3750 | 0.7797 | 0.7083 | Iris-virginica | <-- 113.indis 1. test seti verisi |
| 115 | : | 0.6389 | 0.4167 | 0.5763 | 0.5417 | Iris-versicolor | . |

İlk (0 ve 74 dahil) 0 dan 74. indise kadar olan veriler eğitim verisi olarak kabul görülür. **150 adetlik veri seti için:**
 75. indisteki 1.doğrulama seti verisi ile ilk 74 indise kadar ki verilerin özellik farkları çok boyutlu bir dizide tutulur.
 76. indisteki 2.doğrulama seti verisi ile ilk 74 indise kadar ki verilerin özellik farkları çok boyutlu bir dizide tutulur.

.

112. indisteki 38.doğrulama seti verisi ile ilk 74 indise kadar ki verilerin özellik farkları çok boyutlu bir dizide tutulur.
 Aşağıdaki tabloyu farkların tutulduğu çok boyutlu bir dizi olarak düşünebilirsiniz.

NOT : X'a kadar , X'e kadar: X numarası dahil sayılır. Aşağıdaki veri hesap değerleri programda sürekli değişir.

Eğitim setindeki veri adeti kadar genişlikte ↓

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | 72 | 73 | 74 |
|----|---|---|---|---|---|---|---|---|-------|----|----|----|
| 0 | | | | | | | | | | | | |
| 1 | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | |
| 6 | . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . | . |
| 36 | | | | | | | | | | | | |
| 37 | | | | | | | | | | | | |

↑
Doğrulama seti veri adeti kadar

Kırmızı alandaki değer ilk doğrulama verisinin özelliklerinin ilk eğitim seti verisinin özelliklerinden farkını tutar.

Fark tutması şöyle bir işlemle yapılır: **FARKLAR NEGATİF ÇIKARSA POZİTİFE ÇEVİRİLİRLER**. Uzaklık negatif olamaz.

$$0.5000 - 0.7500 = +0.2500 \quad 0.7917 - 0.1250 = 0.6667$$

| | | | | | | |
|-----------------------------|---|--------|--------|--------|--------|----------------|
| 0.indis 1.eğitim verisi | : | 0.6111 | 0.5000 | 0.6949 | 0.7917 | Iris-virginica |
| 75.indis 1.doğrulama verisi | : | 0.2222 | 0.7500 | 0.1525 | 0.1250 | Iris-setosa |

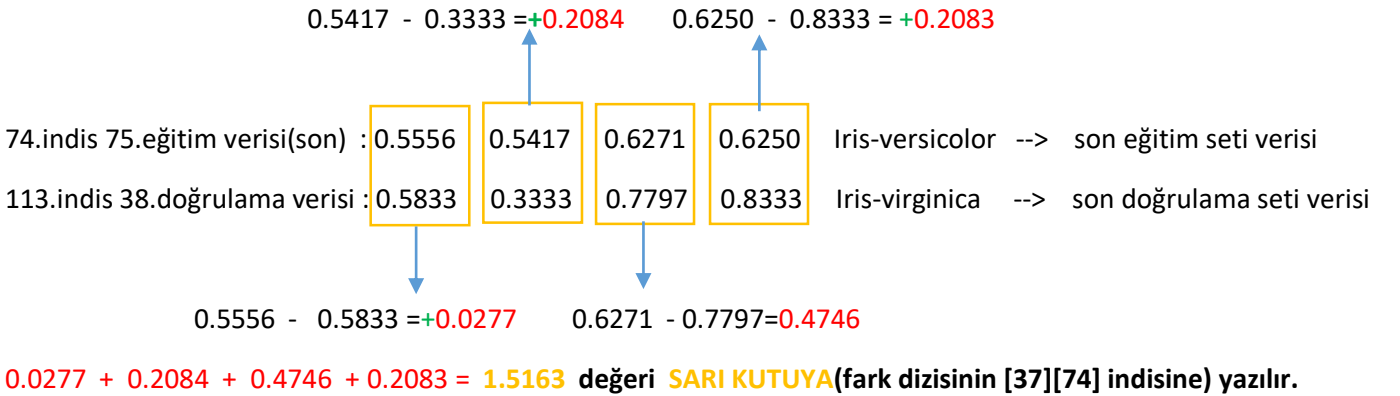
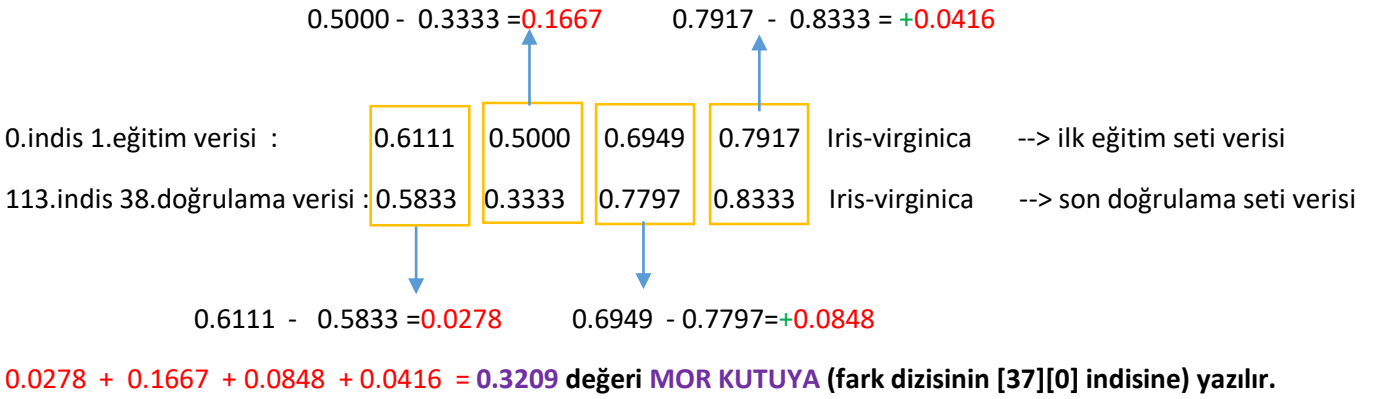
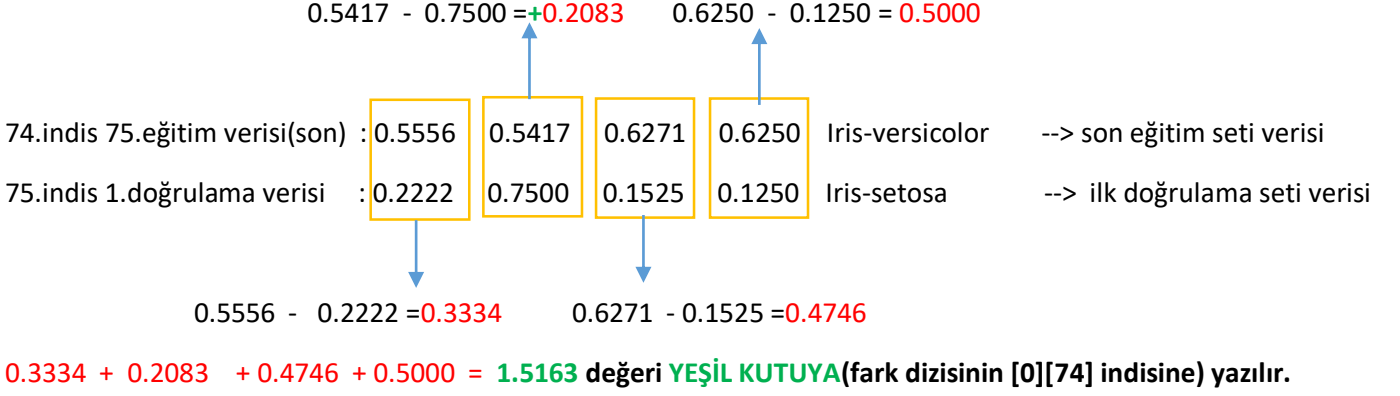
$$0.6111 - 0.2222 = 0.3889 \quad 0.6949 - 0.1525 = 0.5424$$

+0.2500 + 0.3889 + 0.5424 + 0.6667 = 1.848 değeri **KIRMIZI KUTUYA**(fark dizisinin [0][0] indisine) yazılır.

Yeşil alandaki değer ilk doğrulama verisinin özelliklerinin, son eğitim seti verisinin özelliklerinden farkını tutar.

UYARI :Programın her çalışmasında 74.indiste aşağıdaki 0.5556 0.5417 0.6271 0.6250 Iris-versicolor değeri bulunmaması muhtemeldir. Diğer indisteki veriler için de bu geçerlidir.

Bunun sebebi, verilerin program her çalıştığında rastgele yer değiştirmesidir. Aşağıdaki ve dokümandaki diğer veriler, programın bir defa çalışması üzerine alınan bilgiler üzerindeki çalışmalardır.



Yukarıda yapılan işlemler bütün boş kutucuklar için de yapılır ve fark dizisi doldurulur.

Önceki sayfalarda fark dizisini oluşturduk. Fark dizisi sayesinde verilerin birbirine yakınlığını bulacağız.

Tahminler geçici bir fark dizisi oluşturulup içine aktarılır. Bu aktarma işlemi verilerin güncel halini listelemek istediğimizde orijinal fark dizisinde herhangi bir değişikliğin kullanıcıya gözükmemesi içindir.

Farkların konumları tutularak en yakın komşuların bulunması:

Yeni bir doğrulama seti adeti kadar satır sayısına , k tane sütuna sahip dizi oluşturulur. Konum tutucu dizi.

Doğrulama setindeki verilerin(satırlar 0,1,2,...,37) , eğitim setindeki en yakın(farkların en az olduğu) k tane komşusunun konumu konum tutan yeni diziye atanır.

Sonra da konumlarındaki iris adları eğitim setinden bakılarak, adlardan baskın olan doğrulama setindeki verilere tahmin olarak atanır.

Önce her doğrulama verisi için, veriye en yakın(özellik farkları toplamı en az olanlardan) k tane komşunun indis(konum) değerlerini, konum tutucu dizisine atılır.

| | | k tane sütun | | | | | | | | | | | |
|----|--|--------------|----|---|---|---|---|---|---|-------|-----|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | k-1 | k | |
| 0 | | 55 | 23 | | | | | | | | | | |
| 1 | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | |
| 6 | | . | . | . | . | . | . | . | . | . | . | . | . |
| . | | . | . | . | . | . | . | . | . | . | . | . | . |
| . | | . | . | . | . | . | . | . | . | . | . | . | . |
| . | | . | . | . | . | . | . | . | . | . | . | . | . |
| 36 | | | | | | | | | | | | | |
| 37 | | | | | | | | | | | | | |

↑ Doğrulama seti adeti kadar satır

<-- Bu ilk satırdaki değerler şunlardır :

1.doğrulama verisinin özellikleriyle, eğitim setindeki farkının en az olduğu verilerin konumları

Örnek : 55 ve 23 değeri uydurma değerlerdir.

55 yazan yer 1.doğrulama

verisi için en yakın veri özellikleri içeren verinin indisinin, 55 olduğunu bildiriyor.

23 yazan yer en yakın 2. benzer veri

özellikleri içeren verinin indisinin 22 olduğunu belirtir.

<-- Bu satırdaki veriler ise 38.(son) doğrulama verisinin özellikleriyle, eğitim setindeki verilerin farkının en az olduğu konumları tutuyor.

En yakın(benzer) veri = verinin kıyas yapıldığı diğer verilerle farkları toplamının en az olduğu veri

Artık en yakın verilerin konumlarını da bildiğimize göre o konumlara gidip iris ad tahminlerini alabiliriz.

0 kodu iris-setosa , 1 kodu iris-versicolor , 2 kodu iris-virginica türünü gösterir.

Önceki sayfadaki tablodan, konumları belli olan verilerden kaç tanesi hangi türe sahiptir?

0.satırdaki konum verilerinin, konumlarındaki iris adı 0 konumuna sahip olan her bir değer için aşağıdaki

0.satır 0.sütündeki değer +1 artırılır.

Yukarıdaki 0. satırdaki konum verilerinin konumlarına gidilip , hangi iris adı olduğuna bakılır. Her Iris 1 koduna sahip değer için aşağıdaki 0.satır 1.sütündeki değer +1 artırılır.

$k = 5$

| | | Iris adlarının kodları | | |
|----|--|------------------------|---|---|
| | | 0 | 1 | 2 |
| 0 | | 2 | 3 | 0 |
| 1 | | 4 | 1 | 0 |
| 2 | | 3 | 2 | 0 |
| 3 | | 5 | 0 | 0 |
| 4 | | 1 | 1 | 3 |
| 5 | | 2 | 2 | 1 |
| 6 | | . | . | . |
| . | | . | . | . |
| . | | . | . | . |
| . | | . | . | . |
| 36 | | 1 | 2 | 2 |
| 37 | | 0 | 3 | 2 |

Doğrulama seti adeti kadar satır

$$2+3+0=k$$

$$4+1+0=k$$

$$3+2+0=k$$

$$5+0+0=k$$

$$1+1+3=k$$

$$2+2+1=k$$

$$1+2+2=k$$

$$0+2+3=k$$

Satırdaki değerlerin toplamı k ya eşit olmalıdır.

Satırdaki en yüksek değer o satırın iris ad tahminidir.

Satırdaki değerler eşit olursa eşit olan iris isminden hangisinin tahmin olarak seçildiğinin bir önemi yoktur.

0.satırdaki iris adı tahmini, en fazla iris kodunu tutan 3 değerinin sütun indisidir. 0.satır için tahmin iris kodu = 1

1.satır için 4 taneyle baskın olan iris kodu tahmini = 0

2.satır için 3 taneyle baskın olan iris kodu tahmini = 0

37.satır için 3 taneyle baskın olan iris kodu tahmini 1

[Tahmin değerleri uydurmadır]

Bu tahmin değerleri 38x1 lik matris dizisinde tutulur.

Tahminler artık netleşmiştir.

- doğrulama verisi(75.indisteki) tahmini iris kodu 1 dir.
- doğrulama verisi(76.indisteki) tahmini iris kodu 0 dır.
- doğrulama verisi(77.indisteki) tahmini iris kodu 0 dır.
- doğrulama verisi(78.indisteki) tahmini iris kodu 0 dır.
- .
- .
- .
- .
- .
- .

37. doğrulama verisi(112.indisteki) tahmini iris kodu 1 dir.

| | 0 |
|----|---|
| 0 | 1 |
| 1 | 0 |
| 2 | 0 |
| 3 | 0 |
| 4 | 2 |
| 5 | 1 |
| 6 | . |
| . | . |
| . | . |
| . | . |
| 36 | 2 |
| 37 | 1 |

Doğrulama seti adeti kadar satır

Tahmin yüzdesi hesaplamak

Tahminlerimiz bellidir. Tahmin oranımız şuan % 0 dır.

Tahmin oranımızı doğrulama setindeki satırların iris kodlarını, yukarıdaki 38x1lik matristeki iris kodlarıyla kıyaslarız. Her doğru kıyas için tahmin yüzdesi +1 artırılır.

Görsel açıklama :

| | 0 | 1 | 2 | 3 | 4 | 5 | | 0 | |
|----|---|---|---|---|---|---|----|---|---|
| 0 | | | | | | 1 | 0 | 1 | 1 eşit mi 1'e eşit ise tahmin yüzde ++ |
| 1 | | | | | | 2 | 1 | 0 | . |
| 2 | | | | | | 0 | 2 | 2 | . |
| 3 | | | | | | 2 | 3 | 1 | . |
| 4 | | | | | | 1 | 4 | 0 | . |
| 5 | | | | | | 0 | 5 | 1 | . |
| 6 | . | . | . | . | . | . | 6 | . | . |
| . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . |
| 36 | | | | | | 0 | 36 | 1 | 0 eşit ise 1'e tahmin yüzdesini 1 artır |
| 37 | | | | | | 1 | 37 | 0 | 1 eşit ise 0'a tahmin yüzdesini 1 artır |

Doğrulama seti adeti kadar satır

Yukarıdaki değerler kontrol edilen 0 ve 1 ler tamamen uydurma değerlerdir.

Program içinde bu değerler çıktı olarak kullanıcıya sunulur.

$$\text{tahminYuzdesi} = ((100.0 / \text{dogrulamaSeti}) * \text{tahminYuzdesi})$$

doğrulama seti = bizim veri setimiz için 38

Mesala 38 doğru tahminin 25 i doğruysa $2.63157 * 25 = \% 65.7$ lik doğru tahmin yapılmıştır.

Tahmin adetini yüzdelik olarak hesapladık. Çıktı ekranında yanlış tahmin edilen verilerin yanlarında

X işareti bulunur.

Neden tahminler hatalı olabiliyor?

Cevabı hatalı olan tahminleri inceleyerek bulalım.

(77.indisteki) Doğrulama verisinin iris adi : Iris-versicolor Tahmin edilen iris adi : Iris-virginica X

Doğruluk oranı % 94.7368 ama hatalı bir tahmin yapılmış

Örnek:

| | | | | | |
|-----------|-----------|-----------|-----------|-----------------|----------------------------|
| 1.özellik | 2.özellik | 3.özellik | 4.özellik | 5.özellik | |
| 0.4722 | 0.2917 | 0.6949 | 0.6250 | Iris-versicolor | = HATALI TAHMİN EDİLEN TÜR |

Hatalı tahmin verisi türünün 4.özellığı olan ve genel olarak 0.6000 nın altında olan sınırın dışına çıkmış.

Hatalı tahmin verisi türünün 3.özellığı olan ve genel olarak 0.6200 üzerinde olan sınırın dışına çıkmış.

Bu nedenlerden dolayı türler tam anlamda doğru tahmin edilemiyorlar.

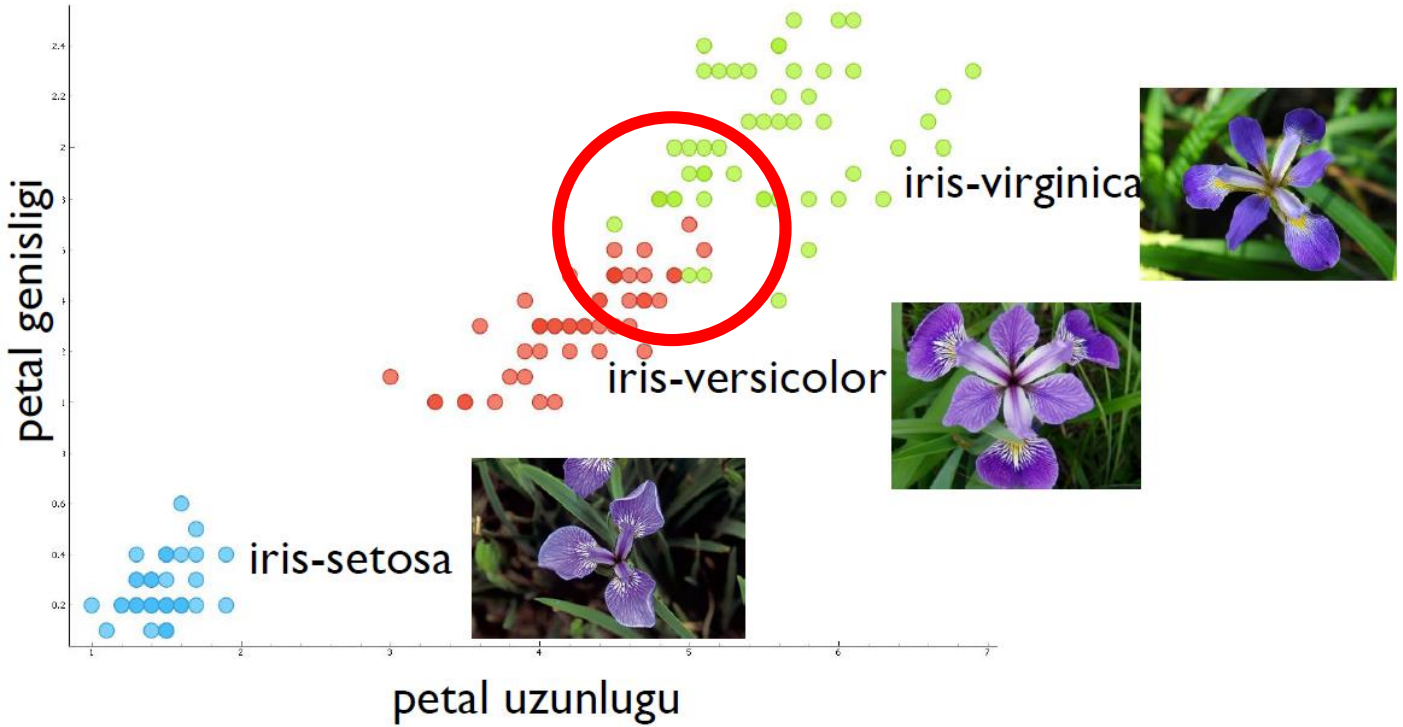
1.özellik 2.özellik 3.özellik 4.özellik 5.özellik

1.özellik 2.özellik 3.özellik 4.özellik 5.özellik

| | | | | | | | | | |
|--------|--------|--------|--------|-----------------|--------|--------|--------|--------|-----------------|
| 0.7500 | 0.5000 | 0.6271 | 0.5417 | Iris-versicolor | 0.5833 | 0.5000 | 0.5932 | 0.5833 | Iris-versicolor |
| 0.7222 | 0.4583 | 0.6610 | 0.5833 | Iris-versicolor | 0.3333 | 0.1250 | 0.5085 | 0.5000 | Iris-versicolor |
| 0.6111 | 0.3333 | 0.6102 | 0.5833 | Iris-versicolor | 0.3889 | 0.3333 | 0.5932 | 0.5000 | Iris-versicolor |
| 0.5556 | 0.5417 | 0.6271 | 0.6250 | Iris-versicolor | 0.1667 | 0.1667 | 0.3898 | 0.3750 | Iris-versicolor |
| 0.6389 | 0.3750 | 0.6102 | 0.5000 | Iris-versicolor | 0.2500 | 0.2917 | 0.4915 | 0.5417 | Iris-versicolor |
| 0.1944 | 0.0000 | 0.4237 | 0.3750 | Iris-versicolor | 0.4444 | 0.4167 | 0.5424 | 0.5833 | Iris-versicolor |
| 0.4722 | 0.0833 | 0.5085 | 0.3750 | Iris-versicolor | 0.5000 | 0.3750 | 0.6271 | 0.5417 | Iris-versicolor |
| 0.3611 | 0.3750 | 0.4407 | 0.5000 | Iris-versicolor | 0.6667 | 0.4583 | 0.5763 | 0.5417 | Iris-versicolor |
| 0.3611 | 0.4167 | 0.5932 | 0.5833 | Iris-versicolor | 0.4167 | 0.2917 | 0.5254 | 0.3750 | Iris-versicolor |
| 0.5278 | 0.0833 | 0.5932 | 0.5833 | Iris-versicolor | 0.3611 | 0.2083 | 0.4915 | 0.4167 | Iris-versicolor |
| 0.6944 | 0.3333 | 0.6441 | 0.5417 | Iris-versicolor | 0.6667 | 0.4167 | 0.6780 | 0.6667 | Iris-versicolor |
| 0.3333 | 0.1667 | 0.4746 | 0.4167 | Iris-versicolor | 0.3333 | 0.1667 | 0.4576 | 0.3750 | Iris-versicolor |
| 0.5278 | 0.3750 | 0.5593 | 0.5000 | Iris-versicolor | 0.2222 | 0.2083 | 0.3390 | 0.4167 | Iris-versicolor |

Aşağıdaki görsele bakacak olursak yanlış tahmin edilen verilerin kırmızı ile işaretlenmiş olan, türlerin özellik ayrımının zor yapıldığı bölgeden olması beklenir. Türlerin belirleyici özelliklerinin normalizasyon sırasında ağırlıklandırılması %100 ' e yakın doğru tahmin yapmamızı sağlayabilir.

iris veri seti



Hata yapılan tür genellikle virginica ve versicolor arasında olması programımızca da tespit edilir.

| | | | | |
|---------------------|--------------------------------|-----------------|--------------------------|-------------------|
| 2. (76.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 3. (77.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 4. (78.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 5. (79.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 6. (80.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor X |
| 7. (81.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 8. (82.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 9. (83.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 10. (84.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 11. (85.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |

Hatanın, program tarafından versicolor ve virginica tespiti sırasında olduğu belirtiliyor.

Tesadüf mü? Tabi ki değil.

Dogruluk orani % 97.3684
Yanlis tahmin(ler)in yaninda X bulunur.

%97.3684 lük doğru tahmin yapan program tek hatayı yukarıdaki fotoğraftaki kırmızı ile daire içine alınmış veriler arasında yaptığı görüyor. Bu tesadüf değil.

En iyi K tespiti

K=3 iken de aynı sonucu verebiliyor k=11 olunca da neden böyle?

Çünkü k değeri türlerin eleman sayısına yaklaştıkça veya bu değeri geçtikçe hata oranı artar.

Veri setinde 50 şer taneden üç farklı tür oluşuyorsa, hatanın k= 40 üstü olduğunda hızla artması beklenir.

Özellikle 50 şer taneden, üç farklı türden oluşan veri seti için k = 50 üzeri girildiğinde hata yapma oranı hızla artar.

En yakın komşular aynı türle dolmuş farklı türlerden benzer olanlar komşu olarak tahmin edilmeye başlanmış olur.

Bu durum hatalı tahmin yapmayı kesinleştirir.

K değerini programı kapatmadan önce değiştirip tahmin doğruluk oranlarını her bir k için karşılaştıralım en yüksek tahmin oranına sahip k değeri iris veri seti için en iyi k değeridir.

Programı kapatsanız bile ortalama olarak en iyi k değeri değişmeyeceği umulur.

UYARI : Aynı K değeri için performans testi yaparken her k değeri girilmesinden önce, programın verileri karıştırma seçeneğine kullanılmalıdır. Farklı k değerleri için bu işlemin yapılmasına gerek yoktur ama aynı k değeri için verilerin yerleri karıştırılırsa farklı rastgele dizilimlerde aynı k değeri için farklı doğruluk oranları bulunur.

Yani:

k=3 için dizi bir kez karıştırılırsa tahmin oranı

Dogruluk oranı % 92.1053

k=3 için, dizi karıştırılmadan alınan tahmin yüzdesi hep aynı olacaktır.

örneğin: k=3 için sonuç:

Dogruluk oranı % 92.1053

tekrar aynı tahmin oranı verildi.

dizi yerlerini karıştırıp k=3 girilirse :

Verilerin yerleri 3 kez karıştırıldı. Güncel veri setini listeleyebilirsiniz.

Dizi yerleri karıştırıldı ve k=3 girildi. Sonuç:

Dogruluk oranı % 94.7368

farklı bir sonuç verdi. Ortalama olarak aynı sonuçları verecektir.

Fakat farklı k değerleri için dizi karıştırma yapmadan tahmin yüzdesi hesaplanırken, aynı k değeri için analiz yapılmak istenirse, veri dizisi karıştırma işlemine tabi tutulmalıdır. Aksi halde dizi ve k değeri aynı olduğu için hep aynı tahmin oranı bulunur.

SUDOKU esprisi neden programda var ?

Gün içinde satranç veya sudoku gibi zihinsel faaliyetleri geliştiren etkinliklerin artırılması amacıyla programa eklendi.

İlk defa 21 üzeri değer girildiğinde çalışacaktır. İkinci 21 üzeri değer girilmesinde aynı espriyi ikinci defa yapılmaz.

Sonuç

Programın çalışması için DevC++ yazılımının yüklü olması gereklidir.

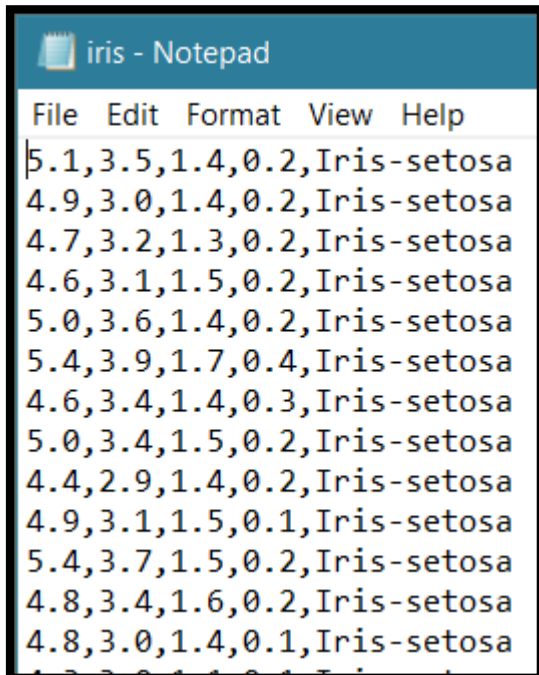
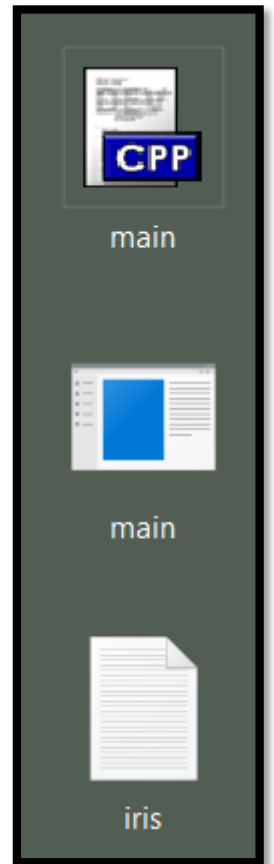
Veya main yazan cpp uzantılı dosya yerine main.exe dosyası bulunmalıdır.

Sağ taraftaki dosyaların aynı dizinde olması gerekir.

Dosyalar sağ taraftaki gibi isimlere sahip olmalıdır.

Virüs engelleyici programlar, main programı çalışınca exe haline dönüşen programı silebilir veya iris.txt dosyasına erişme isteğini engelleyebilir.

Program çalışmasında sıkıntı yaşarsanız antivirüs programınızı duraklatın.



iris.txt içindeki veriler soldaki gibi olmalıdır.

Alt alta yazılı ve virgül ile ayrılmış olmalıdır.

Programın kod kısmından genel bir görüntü :

```

Recep KARADEMİR
IRIS : Süsengiller (Iris), Asparagales takımına ait bir çiçekli bitki familyasıdır.

Iris familyasındaki üç türün temel özelliklerinden oluşan veri seti yardımıyla,
en iyi tür tahmini yapabilen K (KNN) değerini bulan C++ programı.
*/
#include <iostream>
#include <iomanip> // formatlı setw çıktısı alabilmek için.
#include <conio.h> // getch() için.
#include <ctime> // time gerçek rastgele sayıların üretimini zamana bağlama. */
#include <cstdlib> /* srand() ile veri setini rastgele karıştırmak için*/
#include <fstream> // txt içerisindeki verileri okuyabilmek için.
#include "windows.h" // sleep fonk için.
// #include <algorithm> // kısa sort kodları için test amaçlı.
using namespace std;
// Global değişkenler bütün fonksiyonlardan erişilebilir.
void espri();
void veriOkuFonk();
void getRandom();
void standartizasyon();
void verileriYazdir();
void enYakinKomsuAlgoritmasi();
void komşulukFarkHesapla(int);
int veriAdet = 0, veriNitelik = 5, eğitimSeti, doğrulamaSeti, testSeti; // setlerin indislerini tutmak için
// veri adeti txt deki boş olmayan satırların adetini ifade eder. 5 ise verinin özelliklerini(sepalL,.., IrisName vs)
bool espriBitir = false;

float veriSeti[150][5];
// veriSeti dizisinin 1.sütunun da sepalL 2.sütununda sepalW 3.sütununda petalL 4.sütununda petalW 5.sütununda irisName bulunacak.
// veriSeti dizisinin 1.satırında txtdeki ilk verinin bilgileri, n.satırda txt içindeki n.satırdaki verinin bilgileri bulunur.
// irisName 0 ise iris-setosa, 1 ise iris-versicolor, 2 ise iris-virginica verilerini ifade edecek şekilde kodlanmıştır.

int main(int argc, char** argv)
{

void standartizasyon()
{

void veriOkuFonk()
{

void verileriYazdir()
{

void espri()// knn eğer 9 üzerinde bir değer girilirse bu fonksiyon çalıştırılır.
{

void getRandom()
{

void enYakinKomsuAlgoritmasi()
{

void komşulukFarkHesapla(int k)// komşuluk farkları hesaplanacak.
{

```

7 harici 1 temel fonksiyon içerir. Global değişkenler vardır.

Programımızın açılış ekranı :

```

C:\Users\RCP\Desktop\veri odev\main.exe

Veriler basariyla okundu .

Iris adlari bellekte 0 ,1 ,2 olacak sekilde kodlanarak tutuluyor.
Fakat listeleme yapilirken iris adinin kodu
0 ise iris-setosa , 1 ise iris-versicolor , 2 ise iris-virginica olarak listelenecektir.

Buradaki secimlerde tavsiye edilen islem sirasi soyledir : 1 , 2 , 3 , 4 , 5

Verilerin guncel halini listelemek icin..... 1 e bas
Verilerin normalize edilmesi icin..... 2 ye bas
Verilerin yerlerini karistirmek(3 defa karistirilir) icin 3 e bas
K-Nearest Neighbor (en yakin komsu) algoritmasi icin..... 4 e bas
Konsol ekranini temizlemek icin ..... 0 a bas
Programi bitirmek icin ..... 5 e bas

```

```

1 : 5.1000 3.5000 1.4000 0.2000 Iris-setosa
2 : 4.9000 3.0000 1.4000 0.2000 Iris-setosa
3 : 4.7000 3.2000 1.3000 0.2000 Iris-setosa
4 : 4.6000 3.1000 1.5000 0.2000 Iris-setosa
5 : 5.0000 3.6000 1.4000 0.2000 Iris-setosa
6 : 5.4000 3.9000 1.7000 0.4000 Iris-setosa
7 : 4.6000 3.4000 1.4000 0.3000 Iris-setosa
8 : 5.0000 3.4000 1.5000 0.2000 Iris-setosa
9 : 4.4000 2.9000 1.4000 0.2000 Iris-setosa
10 : 4.9000 3.1000 1.5000 0.1000 Iris-setosa
11 : 5.4000 3.7000 1.5000 0.2000 Iris-setosa
12 : 4.8000 3.4000 1.6000 0.2000 Iris-setosa
13 : 4.8000 3.0000 1.4000 0.1000 Iris-setosa
14 : 4.3000 3.0000 1.1000 0.1000 Iris-setosa
15 : 5.8000 4.0000 1.2000 0.2000 Iris-setosa

```

Solda ve aşağıda programın verileri listelemesi görülüyor.

```

145 : 6.7000 3.3000 5.7000 2.5000 Iris-virginica
146 : 6.7000 3.0000 5.2000 2.3000 Iris-virginica
147 : 6.3000 2.5000 5.0000 1.9000 Iris-virginica
148 : 6.5000 3.0000 5.2000 2.0000 Iris-virginica
149 : 6.2000 3.4000 5.4000 2.3000 Iris-virginica
150 : 5.9000 3.0000 5.1000 1.8000 Iris-virginica

```

```

Verilerin guncel halini listelemek icin..... 1 e bas
Verilerin normalize edilmesi icin..... 2 ye bas
Verilerin yerlerini karistirmek(3 defa karistirilir) icin 3 e bas
K-Nearest Neighbor (en yakin komsu) algoritmasi icin..... 4 e bas
Konsol ekranini temizlemek icin ..... 0 a bas

```

Verilere normalizasyon uygulaması yapıldığında oluşan ekran.

```
-----
en kucuk sepall : 4.3
en kucuk sepalW : 2
en kucuk petall : 1
en kucuk petalW : 0.1
-----
en buyuk sepall : 7.9
en buyuk sepalW : 4.4
en buyuk petall : 6.9
en buyuk petalW : 2.5
-----
Standartizasyon formulu : normalize olacak veri = (veri - veriMin) / (veriMax - veriMin)

Veriler 0 ile 1 arasinda normalize edildi. Verilerin guncel halini listeleyebilirsiniz.
-----

Verilerin guncel halini listelemek icin..... 1 e bas
Verilerin normalize edilmesi icin..... 2 ye bas
Verilerin yerlerini karistirmek(3 defa karistirilir) icin 3 e bas
K-Nearest Neighbor (en yakin komsu) algoritmasi icin..... 4 e bas
Konsol ekranini temizlemek icin ..... 0 a bas
Programi bitirmek icin ..... 5 e bas
```

Verilerin yerlerini karıştır komutunu seçtiğimizde aşağıdaki ekran oluşur.

Verilerin yerleri üç kez rastgele olarak karıştırılmış oluyor.

```
Verilerin yerleri 3 kez karistirildi. Guncel veri setini listeleyebilirsiniz.

Verilerin guncel halini listelemek icin..... 1 e bas
Verilerin normalize edilmesi icin..... 2 ye bas
Verilerin yerlerini karistirmek(3 defa karistirilir) icin 3 e bas
K-Nearest Neighbor (en yakin komsu) algoritmasi icin..... 4 e bas
Konsol ekranini temizlemek icin ..... 0 a bas
Programi bitirmek icin ..... 5 e bas
```

Verilerin yerleri karıştırıldıktan ve normalizasyon yapıldıktan sonra listeleme yaparsak veri seti şöyle listeleniyor :

| C:\Users\RCP\Desktop\veri odev\main.exe | | | | |
|---|----------|--------|--------|------------------------|
| 1 | : 0.5000 | 0.4167 | 0.6102 | 0.5417 Iris-versicolor |
| 2 | : 0.0833 | 0.5833 | 0.0678 | 0.0833 Iris-setosa |
| 3 | : 0.9444 | 0.7500 | 0.9661 | 0.8750 Iris-virginica |
| 4 | : 0.8611 | 0.3333 | 0.8644 | 0.7500 Iris-virginica |
| 5 | : 0.1389 | 0.4583 | 0.1017 | 0.0417 Iris-setosa |
| 6 | : 0.1944 | 0.5417 | 0.0678 | 0.0417 Iris-setosa |
| 7 | : 0.5833 | 0.3333 | 0.7797 | 0.8333 Iris-virginica |
| 8 | : 0.6667 | 0.4583 | 0.6271 | 0.5833 Iris-versicolor |
| 9 | : 0.3889 | 0.7500 | 0.1186 | 0.0833 Iris-setosa |
| 10 | : 0.2500 | 0.5833 | 0.0678 | 0.0417 Iris-setosa |
| 11 | : 0.4722 | 0.4167 | 0.6441 | 0.7083 Iris-virginica |
| 12 | : 0.3889 | 0.2083 | 0.6780 | 0.7917 Iris-virginica |
| 13 | : 0.5556 | 0.5417 | 0.8475 | 1.0000 Iris-virginica |
| 14 | : 0.4722 | 0.5833 | 0.5932 | 0.6250 Iris-versicolor |
| 15 | : 0.3611 | 0.3750 | 0.4407 | 0.5000 Iris-versicolor |

K-Nearest Neighbor (en yakın komşu) algoritması için 4 e bastığımızda aşağıdaki gibi bir konsol ekranı çıktısı verilir.

Veri setindeki eleman sayısına bakılarak, verinin setlere ayrılması şöyle yapılmış.

Veri Adedi 150 asagidaki gibi setlere ayrildi :

Egitim seti veri adeti : 75

Dogrulama seti veri adeti : 38

Test seti veri adeti : 37

Setlerin oranli ayrilmasi icin duzeltme yapildi.

Kac yakin komsu icin tahmin yurutulmesini istiyorsun (1-3-5-7-9..-21 vs. girilmesi tavsiye edilir)
Degeri girip entera bas :

K=3 için aldığımız çıktı :

Hatanın virginica ve versicolor arasında olması dikkat çekici. Çünkü bu iki tür arasında hata olma ihtimali yüksekti.

K = 3 En yakın 3 tane komşu incelendi ve tahminler yapıldı.

| | | | |
|--|-----------------|--------------------------|------------------|
| 1. (75.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 2. (76.indisteki) Doğrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 3. (77.indisteki) Doğrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 4. (78.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 5. (79.indisteki) Doğrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 6. (80.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 7. (81.indisteki) Doğrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 8. (82.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 9. (83.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 10. (84.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 11. (85.indisteki) Doğrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 12. (86.indisteki) Doğrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 13. (87.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 14. (88.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 15. (89.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 16. (90.indisteki) Doğrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 17. (91.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 18. (92.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 19. (93.indisteki) Doğrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 20. (94.indisteki) Doğrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 21. (95.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 22. (96.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 23. (97.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 24. (98.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 25. (99.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 26. (100.indisteki) Doğrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 27. (101.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-virginica X |
| 28. (102.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 29. (103.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 30. (104.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 31. (105.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 32. (106.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 33. (107.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 34. (108.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 35. (109.indisteki) Doğrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 36. (110.indisteki) Doğrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 37. (111.indisteki) Doğrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 38. (112.indisteki) Doğrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |

=====

Doğruluk oranı % 97.3684

Yanlış tahmin(ler)in yanında X bulunur.

=====

Devam etmek için bir tusa bas...Press any key to continue . . .

k değeri 21 den büyük girilirse bir espri yapıyor :

| | | | | | |
|---|---|---|---|---|---|
| 5 | 3 | | 7 | | |
| 6 | | | 1 | 9 | 5 |
| | 9 | 8 | | | 6 |
| 8 | | | 6 | | 3 |
| 4 | | | 8 | 3 | 1 |
| 7 | | | 2 | | 6 |
| 6 | | | | 2 | 8 |
| | | | 4 | 1 | 9 |
| | | | 8 | | 7 |
| | | | | 5 | 9 |

k= 27 için hata oranı artması doğal. Hatalı olan tahminler X ile işaretlendi.

K = 27 En yakın 27 tane komsu incelendi ve tahminler yapıldı.

| | | | | |
|---------------------|--------------------------------|-----------------|--------------------------|-------------------|
| 1. (75.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor X |
| 2. (76.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 3. (77.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 4. (78.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 5. (79.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 6. (80.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 7. (81.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 8. (82.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 9. (83.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 10. (84.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 11. (85.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 12. (86.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 13. (87.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 14. (88.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 15. (89.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 16. (90.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 17. (91.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 18. (92.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 19. (93.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 20. (94.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 21. (95.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 22. (96.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 23. (97.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 24. (98.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 25. (99.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 26. (100.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 27. (101.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 28. (102.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 29. (103.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor X |
| 30. (104.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 31. (105.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 32. (106.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 33. (107.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-virginica |
| 34. (108.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |
| 35. (109.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 36. (110.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor X |
| 37. (111.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-setosa |
| 38. (112.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor |

=====

Dogruluk orani % 92.1053

Yanlis tahmin(ler)in yaninda X bulunur.

=====

Devam etmek icin bir tusa bas...Press any key to continue . . .

Hata artıyor çünkü en yakın komşuların başka türden olma ihtimali arttı.

k=57 için hata artıyor: nedenleri önceki sayfalarda açıklandı.

K = 57 En yakın 57 tane komsu incelendi ve tahminler yapıldı.

| | | | | | |
|---------------------|--------------------------------|-----------------|--------------------------|-----------------|---|
| 1. (75.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-versicolor | X |
| 2. (76.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor | |
| 3. (77.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 4. (78.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 5. (79.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-versicolor | X |
| 6. (80.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 7. (81.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 8. (82.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 9. (83.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-versicolor | X |
| 10. (84.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-versicolor | X |
| 11. (85.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-versicolor | X |
| 12. (86.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 13. (87.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 14. (88.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 15. (89.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor | |
| 16. (90.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 17. (91.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor | |
| 18. (92.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor | |
| 19. (93.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 20. (94.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor | |
| 21. (95.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor | |
| 22. (96.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-versicolor | X |
| 23. (97.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-versicolor | X |
| 24. (98.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 25. (99.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 26. (100.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-versicolor | X |
| 27. (101.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-versicolor | X |
| 28. (102.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-versicolor | X |
| 29. (103.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor | |
| 30. (104.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 31. (105.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor | |
| 32. (106.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 33. (107.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-versicolor | X |
| 34. (108.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor | |
| 35. (109.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor | |
| 36. (110.indisteki) | Dogrulama verisinin iris adi : | Iris-virginica | Tahmin edilen iris adi : | Iris-versicolor | X |
| 37. (111.indisteki) | Dogrulama verisinin iris adi : | Iris-setosa | Tahmin edilen iris adi : | Iris-versicolor | X |
| 38. (112.indisteki) | Dogrulama verisinin iris adi : | Iris-versicolor | Tahmin edilen iris adi : | Iris-versicolor | |

```
=====
Dogruluk orani % 28.9474
Yanlis tahmin(ler)in yaninda X bulunur.
=====
Devam etmek icin bir tusa bas...Press any key to continue . . . █
```

Yaptığım testler sonucu en iyi k değerlerinin 1,2,3,4,5 civarı değerler olduğunu buldum.

en kötü değerler ise 35 ve üzeri k değerleri için oldu.

Kod parçası , algoritma işleyişi ve bu doküman hakkındaki bütün sorularınızı

recepkarademir0018@gmail.com adresinden bana iletebilirsiniz.

KAYNAKÇA:(normalizasyon yardımı için)

<https://veribilimcisi.com/2017/07/20/k-en-yakin-komsu-k-nearest-neighborsknn/>