

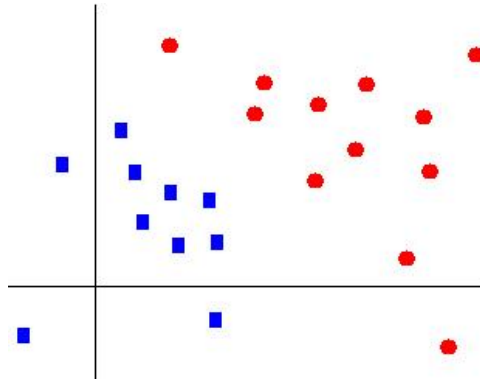


K-En yakın komşu(K-Nearest Neighbor)

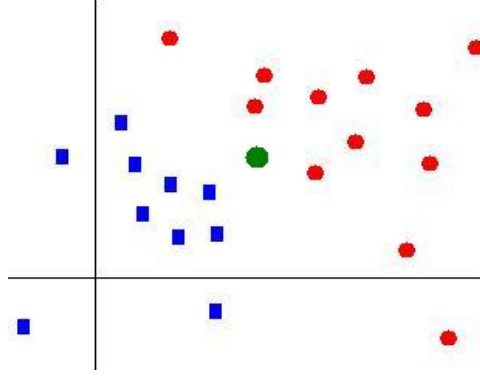
KNN ilk olarak 1951'de Evelyn Fix ve Joseph Hodges tarafından geliştirilen ve daha sonra Thomas Cover tarafından genişletilen parametrik olmayan bir sınıflandırma yöntemidir.

K en yakın komşu algoritması bir verinin ona en yakın olan veri sınıfı ile sınıflandırılmasıyla oluşturulan sınıflandırma algoritmasıdır.

Benzerlik fonksiyonlarının (similarity functions) kullanıldığı bir algoritma türüdür.



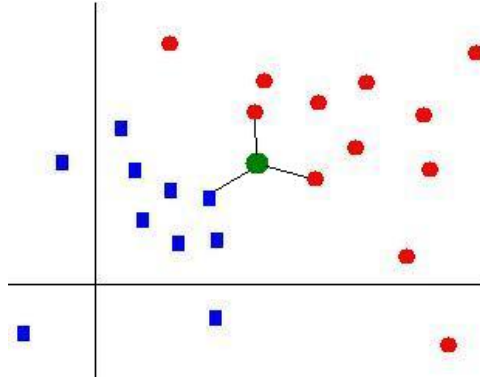
Elimizde şekildeki gibi bir veri seti olduğunu düşünelim.



Eğer şekildeki yeşil verinin KNN ile hangi sınıfa sınıflandırılacağını düşünüyorsak burada K sayısı büyük önem taşımaktadır.

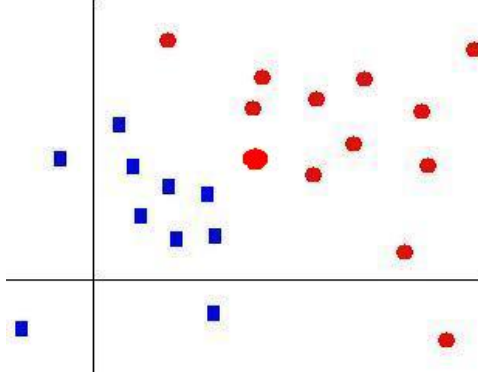
K en yakın komşu sayısını belirten parametredir.

K tek bir sayı olmalıdır. Çift olması eşitlik durumu yaratabileceğinden algoritma için sıkıntılıdır.



K=3 alınan bir KNN algoritmasında , en yakın 3 komşuya bakıldığında yeşil verinin 3 komşusundan 2

si (burada mesafe oklid distance ile hesaplanır) kırmızı veri setindedir. Bu durumda yeşil veri kırmızı veri setine dahil edilir.



Pythonda K-NN

```
#Öncelikle gerekli kütüphane dahil edilir
from sklearn.neighbors import KNeighborsClassifier
#Ardından KNeighborsClassifier sınıfından bir nesne oluşturulur
knn = KNeighborsClassifier(n_neighbors=5,penalty = "l2" , c = 10.0)
"""
n_neighbor paramatersi kaç adet komşu olacağını belirttiğimiz
parametredir ve tek sayı olması gerekir.
Çift sayı olduğunda program hata vermez ancak tek sayı olması eşitliği önlemek
için daha iyi olacaktır. Bir değer girilmediği sürece varsayılan olarak 5 kabul edilir.
"""

#Eğitim işlemi yapılır
knn.fit(xtrain,ytrain)
#Eğitim yapılan veri seti üzerinden tahmin işlemi yapılır
tahmin = knn.predict(xtest)
#Karşılaştırma için karmaşıklık matrisi kullanılabilir
from sklearn.metrics import confusion_matrix
print(confusion_matrix(tahmin,ytest))
```

Mert Aydoğan - Gazi ME