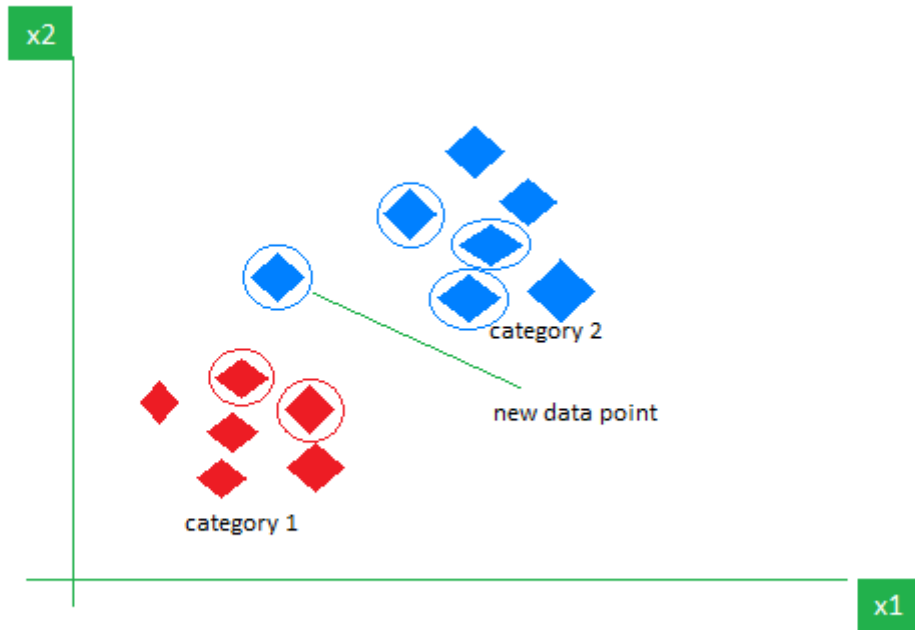


K-EN YAKIN KOMŞU

Denetimli Öğrenme algoritmalarından biridir. Sınıflandırma ve regresyon için kullanılır. Parametrik olmayan tembel bir öğrenme algoritmasıdır. Eğitim verilerini ezberler.

KNN sınıflandırma algoritmasıdır. KNN algoritması sınıflandırılmak istenen veriyi daha önceki verilerle olan yakınlık ilişkisine göre sınıflandırır. Burada K algoritmaya dahil edilecek veri kümesindeki veri sayısıdır. Yani algoritmada K adet en yakın komşu aranır. Bir tahmin yapılacağı zaman , tüm veri setinde en yakın komşuları arar.



K değeri büyüdükçe tahmine duyulan güveni artırır. Ancak K çok büyük bir değer seçilirse kararlar yanlış olabilir

KNN, bir tahmin yapmak için tüm veri kümesini depolar. Bir girdi gözlemi ile veri kümesindeki farklı gözlemler arasındaki benzerliği hesaplayarak tam anında tahminler yapar.

KNN algoritmasında en yakın komşu sınıfı bularak hedef etiketini tahmin etmektir. Sınıflandırılacak noktaya en yakın sınıf, Öklid mesafesi kullanılarak ölçülür. Bunun yanında Manhattan Distance, Euclidean Distance , Hamming Distance metodları da kullanılır.

MESAFE BENZERLİĞİ ÖLÇÜM METODLARI:

Euclidean Distance:

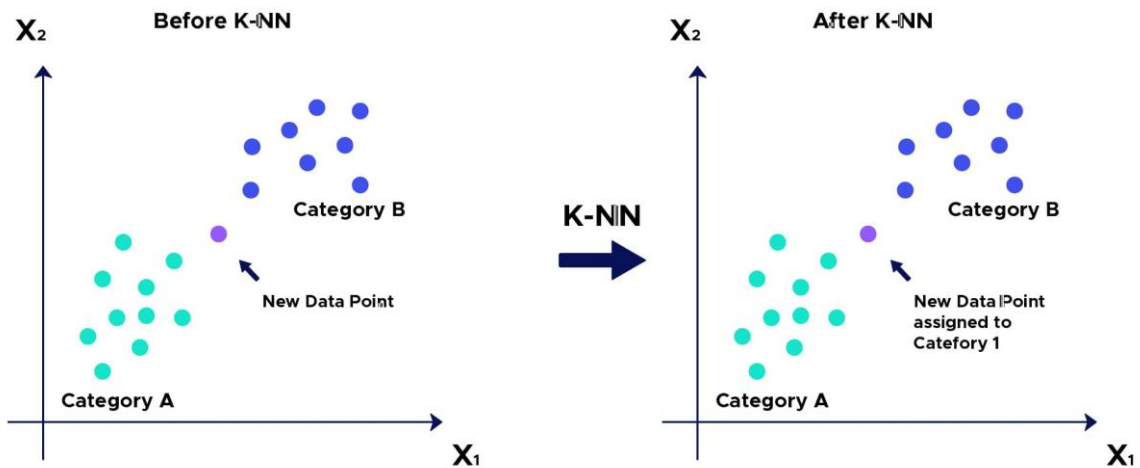
$$\text{distance}(x, X_i) = \sqrt{\sum_{j=1}^d (x_j - X_{ij})^2}$$

Manhattan Distance:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

Minkowski Distance:

$$d(x, y) = (\sum_{i=1}^n (x_i - y_i)^p)^{\frac{1}{p}}$$



KNN'nin nasıl çalıştığına ilişkin adım adım açıklama aşağıda tartışılmaktadır:

Adım 1: K'nin optimal değerinin seçilmesi ,

Burada K tahmin yaparken dikkate alınması gereken en yakın komşuların sayısını temsil eder.

Adım 2: Mesafenin hesaplanması

Hedef ve eğitim veri noktaları arasındaki benzerliği ölçmek için Öklid mesafesi kullanılır. Veri kümesindeki veri noktalarının her biri ile hedef nokta arasındaki mesafe hesaplanır.

Adım 3: En Yakın Komşuları Bulma

Hedef noktaya en küçük mesafelere sahip k veri noktası en yakın komşulardır.

Adım 4: Sınıflandırma İçin Oy Verme veya Regresyon İçin Ortalama Alma

Sınıflandırma probleminde çoğunluk oylaması yapılarak sınıf etiketleri belirlenir. Komşular arasında en çok tekrarlanan sınıf, hedef veri noktası için tahmin edilen sınıf olur. Regresyon probleminde sınıf etiketi, K en yakın komşunun hedef değerlerinin ortalaması alınarak hesaplanır. Hesaplanan ortalama değer, hedef veri noktası için tahmin edilen çıktı haline gelir.

KNN Algoritmasının Artıları ve Eksiler

Artıları:

- Anlaması Kolay
- Veriler Hakkında Varsayım Yok
- Sınıflandırma ve regresyon için kullanılır
- Çok sınıflı problemlerde kolayca çalışır

Eksiler:

- Yoğun bellek kullanımı
- Veri ölçeğine duyarlı
- Çok sayıda bağımsız değişken
- K değerinin doğru seçmediğinde oluşacak risk