

KNN (K-En Yakın Komşu) Algoritması

Denetimli Makine Öğrenmesi

Giriş

- KNN (K-Nearest Neighbors), denetimli öğrenme algoritmasıdır.
- Veri noktalarını mesafeye göre sınıflandırır.
- Eğitim aşamasında model oluşturmaz, tahmin yaparken hesaplama yapar.

KNN Algoritmasının Temelleri



- Bir veri noktasının en yakın K komşusuna göre sınıflandırılması



- Mesafe Ölçütleri:
Öklid, Manhattan, Minkowski



- K değerinin seçilmesi, modelin başarısını etkiler.

Mesafe Ölçütleri

-

Öklid Uzaklığı (Euclidean Distance)

Öklidyen mesafe, iki nokta arasındaki doğrudan (kuş uçuşu) mesafeyi hesaplayan en yaygın metriklerden biridir.

$A(x_1, x_2)$ ve $B(y_1, y_2)$ arasındaki Öklid uzaklığı:

$$d(A, B) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2}$$

Avantajları:

- ✓ Kolay anlaşılır ve hesaplanabilir.
- ✓ Sürekli değişkenlerle iyi çalışır.

Dezavantajları:

- ✗ Büyük ölçekli verilerde hesaplama maliyeti yüksektir.
- ✗ Değişkenler farklı ölçeklerdeyse doğru çalışmaya bilir (normalizasyon gerekebilir).

Mesafe Ölçütleri

-

Manhattan Mesafesi (Manhattan Distance)

Manhattan mesafesi, iki nokta arasındaki mesafeyi yalnızca yatay ve dikey hareketleri kullanarak ölçer.

$A(x_1, x_2)$ ve $B(y_1, y_2)$ arasındaki Manhattan uzaklığı:

$$d(A, B) = |y_1 - x_1| + |y_2 - x_2|$$

Avantajları:

- ✓ Daha az hesaplama gerektirir, büyük veri kümeleri için uygundur.
- ✓ Izgara tabanlı sistemlerde (örneğin şehir blokları, lojistik) daha doğru sonuç verir.

Dezavantajları:

- ✗ Eğik mesafeleri doğru ölçemez.
- ✗ Sürekli değişkenlerde bazen yanıltıcı olabilir.

Mesafe Ölçütleri

-

Minkowski Mesafesi (Minkowski Distance)

Minkowski mesafesi, Öklidyen ve Manhattan mesafelerini kapsayan genelleştirilmiş bir mesafe metriğidir.

$A(x_1, x_2)$ ve $B(y_1, y_2)$ arasındaki Manhattan uzaklığı:

$$d(A, B) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}$$

Avantajları:

- ✓ Esnektir, farklı değerler seçerek Öklidyen veya Manhattan mesafesine dönüştürülebilir.
- ✓ Farklı veri türleri için özelleştirilebilir.

Dezavantajları:

- ✗ Parametre (p) seçimi zor olabilir.
- ✗ Yüksek boyutlu verilerde yorumlamak zorlaşabilir.

KNN Algoritmasının Adımları

Eğitim ve test veri seti ayrılır.



Yeni bir veri noktasının mesafeleri hesaplanır.



En yakın K komşu belirlenir.



Çoğunluk sınıfına göre tahmin yapılır.

KNN'nin Kullanım Alanları

Görüntü Tanıma: El yazısı, yüz tanıma ve obje sınıflandırma

Sağlık ve Tıp: Hastalık teşhisleri, genetik analiz, tıbbi karar destek sistemleri

Finans ve Bankacılık: Kredi riski analizi, dolandırıcılık tespiti, sigorta değerlendirmesi

Doğal Dil İşleme (NLP): Spam filtreleme, duygusal analizi, metin kategorilendirme

Tavsiye Sistemleri: Film, kitap ve ürün önerileri

Lojistik ve Harita Sistemleri: Rota optimizasyonu, teslimat tahmini

Performans Değerlendirme ve Optimizasyon



- Doğruluk, kesinlik ve hata oranı değerlendirilir.



- K değerinin optimize edilmesi gereklidir.



- Ağırlıklı KNN modeli kullanılabilir.

Örnek Senaryo: Müşteri Kredi Riski Tahmini



Bir bankanın kredi verme kararını desteklemek için müşterinin kredi riskini "Düşük Risk" veya "Yüksek Risk" olarak sınıflandırmak istiyoruz.



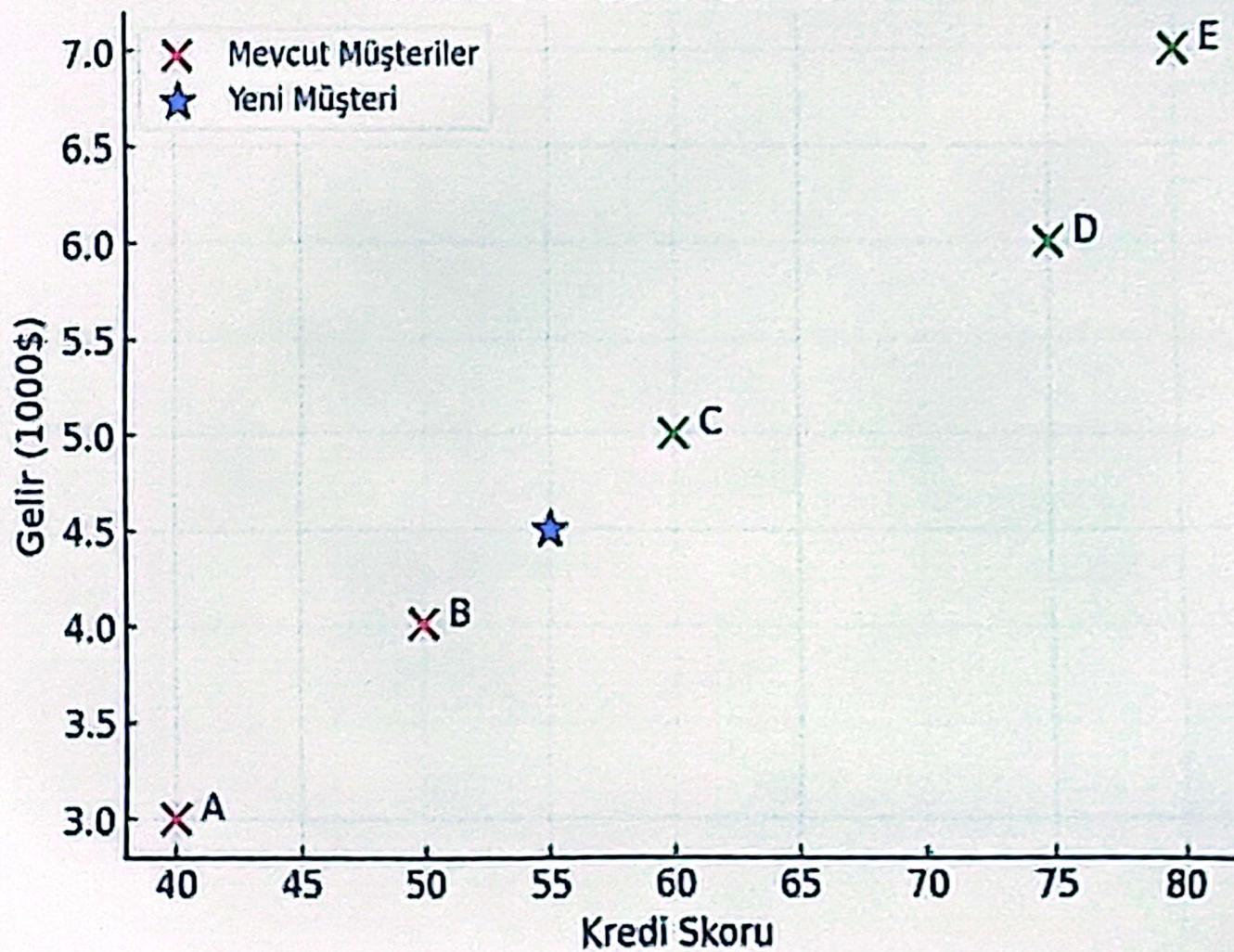
Müşteriler hakkında elimizde gelir (**1000\$ biriminde**) ve kredi geçmişi skoru (**100 üzerinden**) olmak üzere iki özellik var.

Müşteri	Gelir (1000\$)	Kredi Skoru	Kredi Riski
A	3	40	Yüksek Risk
B	4	50	Yüksek Risk
C	5	60	Düşük Risk
D	6	75	Düşük Risk
E	7	80	Düşük Risk

Veri Kümesi

Geliri 4.5K \$ ve kredi skoru 55 olan yeni bir müşterinin kredi riskini tahmin değerini nedir?

KNN ile Kredi Riski Tahmini



Öklid Uzaklık Hesaplama

Müşteri	(Gelir, Kredi Skoru)	Mesafe
A	(3,40)	$\sqrt{(4.5 - 3)^2 + (55 - 40)^2} = 15.13$
B	(4,50)	$\sqrt{(4.5 - 4)^2 + (55 - 50)^2} = 5.02$
C	(5,60)	$\sqrt{(4.5 - 5)^2 + (55 - 60)^2} = 5.02$
D	(6,70)	$\sqrt{(4.5 - 6)^2 + (55 - 75)^2} = 20.06$
E	(7,80)	$\sqrt{(4.5 - 7)^2 + (55 - 80)^2} = 25.5$

K En Yakın Komşunun Seçilmesi (K=3)

En yakın 3 komşu:

B (5.02) → Yüksek Risk

C (5.02) → Düşük Risk

A (15.13) → Yüksek Risk

Oy çoğunluğu:

2 Yüksek Risk – 1 Düşük Risk

*Yeni müşteri "Yüksek Risk" grubuna atanır.