

Denetimsiz Öğrenme (Unsupervised Learning)

Denetimsiz öğrenme, makine öğrenmesi alanında kullanılan bir tekniktir ve denetimli öğrenmenin aksine, etiketlenmemiş veri setleri üzerinde çalışır. Bu teknikte, algoritmalar veri setindeki desenleri veya yapıları keşfetmeye çalışır ve veri içindeki doğal gruplamaları ortaya çıkarmayı hedefler. Temel amacı, veri setindeki yapıları anlamak ve bilinmeyen ilişkileri bulmaktır.

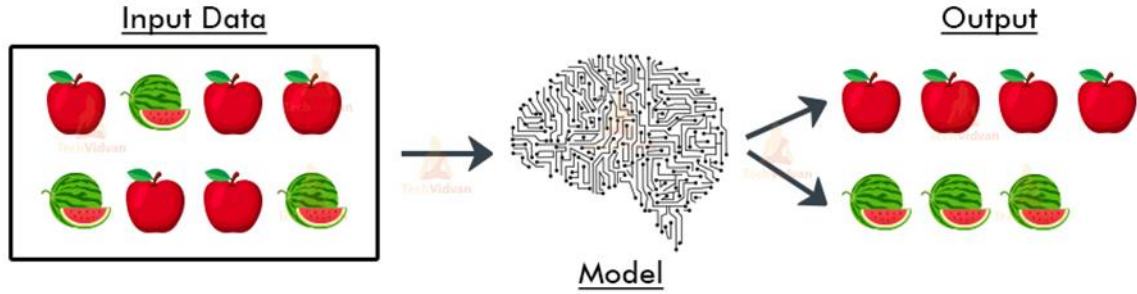
Denetimsiz Öğrenme Süreci

- Veri Toplama:** İlk adım, problem alanına uygun verilerin toplanmasıdır. Bu veriler genellikle etiketlenmemiş veya doğrudan gözlemlenmiş veriler olabilir.
- Veri Hazırlama:** Toplanan veriler, algoritmalar tarafından işlenebilir formata dönüştürülür. Özellikle büyük veri setleri için ön işleme ve boyut azaltma teknikleri kullanılabilir.
- Model Seçimi:** Veri setinin özelliklerine ve problem türüne göre uygun bir denetimsiz öğrenme algoritması seçilir. Örneğin, kümeleme (clustering) veya boyut azaltma (dimensionality reduction) gibi.
- Model Eğitimi:** Seçilen algoritma, veri seti üzerinde uygulanır ve veri içindeki yapıları veya desenleri belirlemeye çalışır. Bu süreçte, algoritma genellikle veri setindeki benzerlikleri veya farklılıkları ölçer.
- Sonuçların Değerlendirilmesi:** Denetimsiz öğrenme algoritmaları, genellikle kalitatif olarak değerlendirilir. Örneğin, kümeleme algoritmalarında oluşturulan grupların anlamlılığı veya boyut azaltma algoritmalarında veri setinin yapısını koruma kabiliyeti göz önünde bulundurulur.
- Model İyileştirme ve Yorumlama:** Denetimsiz öğrenme sonuçları, genellikle insanlar tarafından yorumlanır ve işlenir. Özellikle kümeleme sonuçlarının anlamlandırılması ve boyut azaltma sonuçlarının yorumlanması önemlidir.

Denetimsiz Öğrenme Algoritmaları ve Uygulama Alanları

- Kümeleme (Clustering):** Veri setindeki benzer özelliklere sahip veri noktalarını gruplamak için kullanılır. Örneğin, pazar segmentasyonu veya sosyal ağ analizi gibi.
- Boyut Azaltma (Dimensionality Reduction):** Yüksek boyutlu veri setlerindeki gereksiz veya gürültülü özellikleri ortadan kaldırarak veri setinin boyutunu azaltır. Özellikle görüntü ve video analizi gibi alanlarda kullanılır.
- İşaretçi Analizi (Anomaly Detection):** Veri setindeki anormal veya beklenmeyen davranışları tespit etmek için kullanılır. Örneğin, kredi kartı sahtekarlığı tespiti veya ağ güvenliği analizi gibi.
- Birliktelik Kuralları (Association Rule Learning):** Veri setindeki öğeler arasındaki ilişkileri keşfetmek için kullanılır. Örneğin, alışveriş sepet analizi veya tavsiye sistemleri gibi.
- Doğrusal Olmayan Yapıları Keşfetme:** Denetimsiz öğrenme, veri setindeki doğrusal olmayan yapıları (örneğin, veri setindeki gruplamalar veya alt yapışıklıklar) belirlemek için kullanılabilir.

Unsupervised Learning in ML



Denetimsiz Öğrenmenin Avantajları ve Dezavantajları

Avantajlar

- **Etiketleme Gereksinimi Yok:** Etiketlenmemiş veri setleri ile çalışabilir, bu da veri toplama ve hazırlama süreçlerini kolaylaştırır.
- **Yeni Veri Yapılarının Keşfi:** Bilinmeyen yapıları veya desenleri keşfetmek için güçlü bir yöntemdir.
- **Öğrenme Süreci Daha Esnek:** Genellikle denetimli öğrenme kadar sıkı etiketleme veya denetim gerektirmez.

Dezavantajlar

- **Sonuçların Yorumlanması Zordur:** Denetimsiz öğrenme sonuçları genellikle kalitatif olarak değerlendirilir ve yorumlanması zor olabilir.
- **Doğruluk Değerlendirmesi Zordur:** Denetimli öğrenmede olduğu gibi doğruluk veya benzer metriklerle performans değerlendirmesi yapmak zordur.
- **Optimal Parametre Ayarı Zorluğu:** Bazı denetimsiz öğrenme algoritmalarında optimal parametre ayarı yapmak zor olabilir.

Sonuç

Denetimsiz öğrenme, etiketlenmemiş veri setlerinde yapıların ve desenlerin keşfi için güçlü bir tekniktir. Kümeleme, boyut azaltma, anomalite tespiti ve birliktelik kuralları gibi çeşitli algoritmaları içerir ve bu algoritmalar farklı veri analiz problemlerine uygun olarak kullanılabilir. Her bir denetimsiz öğrenme problemi için uygun algoritma seçimi ve sonuçların doğru yorumlanması, başarılı uygulamalar için kritik öneme sahiptir.

Örnek

Denetimsiz öğrenme, etiketlenmemiş veri kümelerini analiz etmek ve gruplandırmak için kullanılır. Denetimsiz öğrenmenin yaygın bir örneği, K-means kümeleme algoritmasıdır. Bu algoritma, verileri belirli sayıda kümeye (clusters) ayırır.

Aşağıda, scikit-learn kütüphanesini kullanarak K-means kümeleme algoritmasını uygulayan bir Python örneği bulacaksınız. Bu örnekte, Iris veri setini kullanacağız ancak tür etiketlerini kullanmadan verileri kümelere ayıracağız.

```
# Gerekli kütüphanelerin import edilmesi
import numpy as np
import pandas as pd
from sklearn.datasets import load_iris
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Iris veri setinin yüklenmesi
iris = load_iris()
X = iris.data

# Verilerin standartlaştırılması
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# K-means modeli oluşturulması
kmeans = KMeans(n_clusters=3, random_state=42)

# Modelin eğitilmesi (kümelere ayırılması)
kmeans.fit(X_scaled)

# Küme etiketlerinin tahmin edilmesi
labels = kmeans.labels_
```

```
# Küme merkezlerinin elde edilmesi
centers = kmeans.cluster_centers_

# Sonuçların görselleştirilmesi
plt.figure(figsize=(8, 6))
plt.scatter(X_scaled[:, 0], X_scaled[:, 1], c=labels, cmap='viridis', marker='o', edgecolor='k')
plt.scatter(centers[:, 0], centers[:, 1], c='red', s=200, alpha=0.75, marker='x')
plt.title('K-means Kümeleme (Iris Veriseti)')
plt.xlabel('Özellik 1')
plt.ylabel('Özellik 2')
plt.show()
```

Açıklamalar

1. Veri Setinin Yüklenmesi:

- `load_iris()` fonksiyonu ile Iris veri seti yüklenir.
 - `x` değişkeni, çiçek yapraklarının uzunluk ve genişlik bilgilerini içerir (features).
2. **Verilerin Standartlaştırılması:**
 - `StandardScaler` kullanılarak veriler standartlaştırılır (her bir özellik, ortalaması 0 ve standart sapması 1 olacak şekilde ölçeklenir).
 3. **K-means Modelinin Oluşturulması:**
 - `KMeans()` fonksiyonu ile K-means modeli oluşturulur.
 - `n_clusters=3`, verilerin 3 kümeye ayrılacağını belirtir.
 4. **Modelin Eğitilmesi:**
 - `fit()` fonksiyonu ile model, verileri kullanarak kümeleri bulur.
 5. **Küme Etiketlerinin Tahmin Edilmesi:**
 - `labels` değişkeni, her bir veri noktasının hangi kümeye ait olduğunu belirtir.
 6. **Küme Merkezlerinin Elde Edilmesi:**
 - `cluster_centers_` özelliği, her bir kümenin merkezini verir.
 7. **Sonuçların Görselleştirilmesi:**
 - `matplotlib` kütüphanesi kullanılarak veriler ve küme merkezleri görselleştirilir.

Bu örnek, denetimsiz öğrenme (unsupervised learning) yöntemlerini kullanarak verilerin nasıl kümelere ayrılacağını gösterir. K-means algoritması dışında, hiyerarşik kümeleme, DBSCAN gibi diğer denetimsiz öğrenme algoritmaları da benzer şekilde kullanılabilir.