

Sınıflandırma Algoritması

Sınıflandırma, makine öğrenimi alanında, veri noktalarını belirli kategorilere veya sınıflara atayan bir öğrenme türüdür. Temel olarak, sınıflandırma algoritmaları, belirli özelliklere dayanarak bir girdi verisini sınıflandırmak için bir model oluşturur. Bu sınıflandırma işlemi, genellikle denetimli öğrenme olarak adlandırılan bir öğrenme paradigması altında gerçekleşir. Denetimli öğrenme, eğitim verileri ile birlikte gelen etiketlenmiş verileri kullanarak bir model oluşturur ve bu modeli test verileri üzerinde değerlendirir.

Sınıflandırma algoritmaları, genellikle aşağıdaki kategorilere ayrılır:

1. **Doğrusal Sınıflandırma Algoritmaları:** Bu algoritmalar, sınıfları ayırmak için bir doğru veya hiperdüzlem gibi doğrusal bir sınır kullanır. Örneğin, lojistik regresyon ve destek vektör makineleri (SVM) doğrusal sınıflandırma algoritmalarına örnektir.
2. **Doğrusal Olmayan Sınıflandırma Algoritmaları:** Doğrusal sınırlarla ayrılamayan veri setleri için kullanılır. Örneğin, karar ağaçları, k-En Yakın Komşu (KNN) ve yapay sinir ağları (NN) doğrusal olmayan sınıflandırma algoritmalarına örnektir.
3. **Ensemble Sınıflandırma Algoritmaları:** Bu algoritmalar, birden çok sınıflandırıcıyı bir araya getirerek daha güçlü bir sınıflandırıcı oluştururlar. Örneğin, rastgele ormanlar ve gradient boosting gibi algoritmalar ensemble sınıflandırma algoritmalarına örnektir.
4. **Derin Öğrenme Tabanlı Sınıflandırma Algoritmaları:** Derin öğrenme, karmaşık yapılar ve büyük veri setleri üzerinde etkili olan yapay sinir ağları kullanır. Görüntü tanıma, doğal dil işleme ve ses tanıma gibi alanlarda sınıflandırma problemlerinde derin öğrenme algoritmaları yaygın olarak kullanılır.

SINIFLANDIRMA ALGORİTMASINA ÖRNEK:

Destek Vektör Makineleri (SVM):

Destek vektör makineleri, sınıflandırma ve regresyon problemleri için kullanılan güçlü bir makine öğrenimi algoritmasıdır. Özellikle doğrusal ve doğrusal olmayan sınıflandırma problemlerinde etkilidir.

Örneğin, bir müşterinin bankacılık ürünlerini satın alma olasılığını tahmin etmek için SVM kullanalım. Elimizde, müşterilere ait özellikler (yaş, gelir, kredi skoru gibi) ve satın alma yapma durumlarına ilişkin etiketler bulunsun. Bu verileri kullanarak bir SVM modeli oluşturabilir ve yeni müşterilerin satın alma yapma olasılığını tahmin edebiliriz.

Örnek veri seti:

YAŞ - GELİR - KREDİ SKORU - SATIN ALMA DURUMU

35 - 50000 - 700 - SATIN ALDI

45 - 70000 - 600 - SATIN ALMADI

30 - 60000 - 750 - SATIN ALDI

50 - 80000 - 800 - SATIN ALDI

... - ... - ... - ...

Bu veri setini kullanarak bir SVM modeli eğitebiliriz. Model, müşterilerin yaş, gelir ve kredi skoru gibi özelliklerini kullanarak satın alma durumunu (sınıfı) tahmin etmeyi öğrenir. Daha sonra, yeni bir müşteri verisi verildiğinde, model bu müşterinin satın alma olasılığını tahmin edebilir. Bu şekilde, banka müşterilerinin davranışlarını anlamak ve satın alma olasılıklarını tahmin etmek için SVM gibi sınıflandırma algoritmaları kullanılabilir.

ÖRNEK 1:

```
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score

# Iris veri setini yükleme
iris = datasets.load_iris()
X = iris.data # Özellikler
y = iris.target # Etiketler

# Veri setini eğitim ve test setlerine ayırma
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)

# Verileri ölçeklendirme
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Destek vektör makinesi modelini oluşturma ve eğitme
svm_classifier = SVC(kernel='linear', random_state=42)
svm_classifier.fit(X_train, y_train)

# Test seti üzerinde modeli değerlendirme
y_pred = svm_classifier.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print("Model Doğruluğu:", accuracy)
```

Bu kod örneği, Scikit-learn kütüphanesini kullanarak **Iris veri seti üzerinde basit bir SVM modeli oluşturur ve değerlendirir. Iris veri seti, bitki türlerinin çeşitli özelliklerini içerir ve bu özelliklere dayanarak bitki türlerini sınıflandırmak için kullanılır. Örnek, veri setini eğitim ve test setlerine böler, özellikleri ölçeklendirir, bir SVM modeli oluşturur ve test seti üzerinde modelin doğruluğunu değerlendirir.**

ÖRNEK 2:

```
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score

# Breast Cancer veri setini yükleme
data = load_breast_cancer()
X = data.data # Özellikler
y = data.target # Etiketler

# Veri setini eğitim ve test setlerine ayırma
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)

# Verileri ölçeklendirme
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Destek vektör makinesi modelini oluşturma ve eğitme
svm_classifier = SVC(kernel='linear', random_state=42)
svm_classifier.fit(X_train_scaled, y_train)

# Test seti üzerinde modeli değerlendirme
y_pred = svm_classifier.predict(X_test_scaled)
accuracy = accuracy_score(y_test, y_pred)
print("Model Doğruluğu:", accuracy)
```

Bu kod örneği, Scikit-learn kütüphanesini kullanarak, meme kanseri veri seti üzerinde bir SVM sınıflandırıcı modeli oluşturur ve test seti üzerinde modelin doğruluğunu değerlendirir. Bu örnekte, özellikleri ölçeklendirme adımı da eklenmiştir, çünkü SVM gibi bazı algoritmalar, özelliklerin ölçeklendirilmesini gerektirebilir.

Regresyon/Sınıflandırma Ne Zaman Kullanılır?

Veri kümesini yanıt değişkeniyle ilişkili farklı sınıflara ayırmaya ihtiyaç duyulduğunda sınıflandırma ağaçları kullanılır. Genellikle bu sınıflar "Evet" veya "Hayır" gibi ikilidir ve birbirini dışlar. İki'den fazla sınıfın olabileceği durumlar olsa da bu senaryolarda sınıflandırma ağacı algoritmasının değiştirilmiş bir versiyonu kullanılır.

Öte yandan sürekli yanıt değişkenleriyle uğraşırken regresyon ağaçlarından yararlanır. Örneğin, yanıt değişkeni bir nesnenin fiyatı veya günün sıcaklığı gibi sürekli değerleri temsil ediyorsa regresyon ağacı uygun seçimdir.

Sınıflandırma	Regresyon
Bu problem ifadesinde hedef değişkenler ayrıktır.	Bu problem tanımında hedef değişkenler sürekli dir.
<u>Spam E-posta Sınıflandırması</u> , <u>Hastalık tahmini</u> gibi sorunlar Sınıflandırma Algoritmaları kullanılarak çözülmektedir.	<u>Konut Fiyat Tahmini</u> , <u>Yağış Tahmini</u> gibi problemler regresyon algoritmaları kullanılarak çözülmektedir.
Bu algoritmada iki sınıfı mümkün olan maksimum ayrımla ayırabilecek mümkün olan en iyi karar sınırını bulmaya çalışıyoruz.	Bu algoritmada verideki genel eğilimi temsil edebilecek en uygun çizgiyi bulmaya çalışıyoruz.
<u>Hassasiyet, Geri Çağırma ve F1-Score</u> gibi değerlendirme metrikleri burada sınıflandırma algoritmalarının performansını değerlendirmek için kullanılır.	<u>Ortalama Karesel Hata, R2 Puanı</u> ve <u>MAPE</u> gibi değerlendirme ölçümleri burada regresyon algoritmalarının performansını değerlendirmek için kullanılır.
<u>Burada ikili Sınıflandırma</u> veya <u>Çok Sınıflı Sınıflandırma</u> problemleri gibi problemlerle karşı karşıyayız .	<u>Burada Doğrusal Regresyon</u> modelleri ve doğrusal olmayan modeller gibi sorunlarla karşı karşıyayız .