

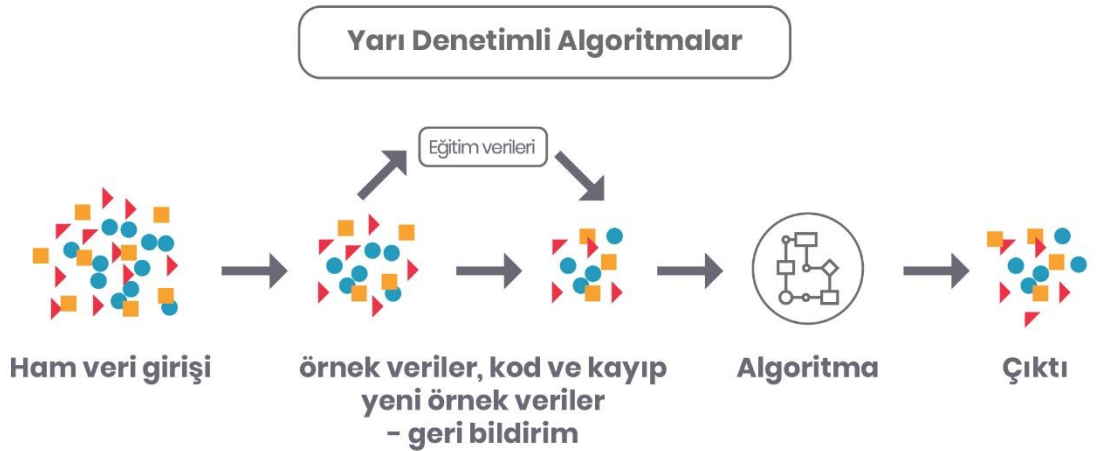
Yarı Denetimli Öğrenme

Yarı denetimli öğrenme (semi-supervised learning), hem etiketlenmiş hem de etiketlenmemiş veri örneklerini kullanarak öğrenme yapmayı amaçlayan bir makine öğrenmesi paradigmasıdır. Bu yöntem, sınırlı sayıda etiketlenmiş veriye sahip olduğumuz durumlarda genellikle kullanılır ve etiketlenmemiş veri örneklerinden de yararlanarak modelin performansını artırmayı hedefler.

Yarı Denetimli Öğrenme Süreci

Yarı denetimli öğrenme süreci, genellikle aşağıdaki adımları içerir:

- Etiketlenmiş ve Etiketlenmemiş Veri Toplama:** İlk adım, problem alanına uygun şekilde hem etiketlenmiş hem de etiketlenmemiş veri örneklerini toplamaktır. Etiketlenmiş veri örnekleri, genellikle maliyetli veya zaman alıcı olduğu için sınırlıdır.
- Veri Hazırlama:** Toplanan veri örnekleri, gerekli ön işleme adımlarından geçirilir ve uygun formata dönüştürülür. Bu adım, veri setinin kalitesini artırmak ve modelin daha iyi öğrenme yapmasını sağlamak için önemlidir.
- Model Seçimi:** Hem etiketlenmiş hem de etiketlenmemiş veri örneklerini kullanabilen özel yarı denetimli öğrenme algoritmalarından biri seçilir. Bu algoritmalar, genellikle sınıflandırma veya regresyon gibi belirli görevler için optimize edilmiştir.
- Model Eğitimi:** Seçilen yarı denetimli öğrenme algoritması, hem etiketlenmiş veri örneklerini hem de etiketlenmemiş veri örneklerini kullanarak eğitilir. Etiketlenmiş veri örnekleri, modelin doğru sonuçlar elde etmesi için kullanılırken; etiketlenmemiş veri örnekleri, modelin genel veri setindeki desenleri anlamasına yardımcı olur.
- Model Değerlendirme:** Eğitilen model, genellikle ayrılmış bir test veri seti üzerinde değerlendirilir. Modelin performansı, doğruluk, hassasiyet, geri çağırma gibi metrikler kullanılarak ölçülür. Bu adım, modelin genelleme yeteneğini ve etkinliğini değerlendirmek için önemlidir.
- Model İyileştirme ve Ayar:** Modelin performansını artırmak için hiperparametre ayarlaması veya özellik seçimi gibi iyileştirme teknikleri uygulanabilir. Bu adımlar, modelin daha iyi sonuçlar vermesini sağlamak için gereklidir.



Yarı Denetimli Öğrenme Algoritmaları ve Uygulama Alanları

Yarı denetimli öğrenme için çeşitli algoritmalar bulunmaktadır, bunlar arasında şunlar yer alır:

- **Etiketleme İle İyileştirme (Label Propagation):** Etiketlenmiş veri örneklerinden başlayarak etiketlenmemiş veri örneklerine etiket atar.
- **Co-Öğrenme (Co-Training):** Birden fazla özellik setiyle çalışan ve her bir setin diğerini etiketlemesine yardımcı olan bir öğrenme yöntemidir.
- **Öz-Yönetimli Öğrenme (Self-Training):** İlk başta yalnızca etiketlenmiş veri kullanır, ardından bu verileri kullanarak etiketlenmemiş veri örneklerine etiket atar.
- **Transdüktif Öğrenme (Transductive Learning):** Model, hem eğitim sırasında hem de tahmin sırasında etiketlenmemiş veri örneklerini kullanır.

Yarı Denetimli Öğrenmenin Avantajları ve Dezavantajları

Avantajlar

- **Sınırlı Etiketli Veri Kullanımı:** Etiketlenmiş veri örnekleri sınırlı olduğunda bile model performansını artırabilir.
- **Genelleme Yeteneği:** Etiketlenmemiş veri örneklerinden elde edilen ek bilgi, modelin genelleme yeteneğini artırabilir.
- **Daha Geniş Uygulama Alanı:** Etiketlenmiş veri setlerinin maliyetli veya zaman alıcı olduğu alanlarda etkili bir çözüm sunar.

Dezavantajlar

- **Etiketlenmemiş Verilerin Kalitesi:** Etiketlenmemiş veri örneklerinin kalitesi veya doğruluğu problem olabilir.
- **Modelin Karmaşıklığı:** Hem etiketlenmiş hem de etiketlenmemiş veri örneklerini kullanmak, modelin karmaşıklığını artırabilir.
- **Optimal Parametre Ayarı:** Algoritmalar genellikle iyi bir performans için dikkatli bir şekilde ayarlanmalıdır.

Sonuç

Yarı denetimli öğrenme, sınırlı etiketli veri örnekleriyle çalışan ve genellikle etiketlenmemiş veri örneklerinden de yararlanarak modelin performansını artırmayı hedefleyen bir makine öğrenmesi yaklaşımıdır. Bu yöntem, özellikle büyük veri setlerinde etiketleme maliyetlerini azaltmak ve daha geniş bir veri anlayışı sağlamak için önemlidir. Her problem için en uygun algoritma seçimi ve modelin doğru şekilde değerlendirilmesi, başarılı yarı denetimli öğrenme uygulamaları için kritik öneme sahiptir.

Örnek

Yarı denetimli öğrenme (semi-supervised learning), sınırlı sayıda etiketli veri ile çok sayıda etiketlenmemiş veri kullanarak bir model eğitmeye yönelik bir yaklaşımdır. Bu yöntem, tamamen denetimli öğrenme için yeterli etiketli veriye sahip olmadığımız durumlarda kullanılır.

Scikit-learn kütüphanesi, yarı denetimli öğrenme için "Label Spreading" ve "Label Propagation" algoritmalarını sağlar. Aşağıda, "Label Spreading" algoritmasını kullanarak yarı denetimli bir öğrenme uygulaması örneğini bulabilirsiniz.

```
# Gerekli kütüphanelerin import edilmesi
import numpy as np
import matplotlib.pyplot as plt
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.semi_supervised import LabelSpreading
from sklearn.metrics import accuracy_score

# Iris veri setinin yüklenmesi
iris = datasets.load_iris()
X, y = iris.data, iris.target

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

# 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)

# Eğitim setindeki etiketlerin bir kısmını gizleyerek yarı denetimli öğrenme için veri oluşturma
n_labeled_points = int(len(y_train) * 0.1) # %10 etiketli veri
unlabeled_indices = np.arange(len(y_train))
np.random.shuffle(unlabeled_indices)
unlabeled_indices = unlabeled_indices[n_labeled_points:]
```

```
# Etiketleri gizlemek için -1 kullanıyoruz
y_train[unlabeled_indices] = -1

# Label Spreading modelinin oluşturulması ve eğitilmesi
label_spread = LabelSpreading(kernel='knn', alpha=0.8)
label_spread.fit(X_train, y_train)

# Test verisi üzerinde tahminlerin yapılması
y_pred = label_spread.predict(X_test)

# Modelin doğruluğunun hesaplanması
accuracy = accuracy_score(y_test, y_pred)
print(f"Test seti doğruluğu: {accuracy * 100:.2f}%")

# Sonuçların görselleştirilmesi
plt.figure(figsize=(8, 6))
for i, label in enumerate(np.unique(y)):
    plt.scatter(X_test[y_test == label][:, 0], X_test[y_test == label][:, 1], label=iris.target_names[label], c='blue', marker='o')
    plt.scatter(X_test[y_pred != y_test][:, 0], X_test[y_pred != y_test][:, 1], c='red', marker='x')
plt.title('Yarı Denetimli Öğrenme: Label Spreading')
plt.xlabel('Özellik 1')
plt.ylabel('Özellik 2')
plt.legend()
plt.show()
```

Açıklamalar

1. Veri Setinin Yüklenmesi:

- o `datasets.load_iris()` fonksiyonu ile Iris veri seti yüklenir.
- o `x` değişkeni özellikleri (features) ve `y` değişkeni sınıf etiketlerini (labels) içerir.

2. Verilerin Standartlaştırılması:

- o `StandardScaler` kullanılarak veriler standartlaştırılır (her bir özellik, ortalaması 0 ve standart sapması 1 olacak şekilde ölçeklenir).

3. Veri Setinin Eğitim ve Test Setlerine Ayrılması:

- o `train_test_split()` fonksiyonu ile veri seti eğitim ve test setlerine ayrılır. Bu örnekte, veri setinin %30'u test seti olarak kullanılır.

4. Etiketlerin Gizlenmesi:

- o Eğitim setindeki etiketlerin %90'ı gizlenir. Gizlenen etiketler -1 değeriyle değiştirilir.

5. Label Spreading Modelinin Oluşturulması ve Eğitilmesi:

- o `LabelSpreading` algoritması kullanılarak model oluşturulur ve `fit()` fonksiyonu ile eğitilir.

6. Test Verisi Üzerinde Tahminlerin Yapılması:

- o `predict()` fonksiyonu ile test seti üzerinde tahminler yapılır.

7. Modelin Doğruluğunun Hesaplanması:

- o `accuracy_score` fonksiyonu ile modelin doğruluğu hesaplanır ve ekrana yazdırılır.

8. Sonuçların Görselleştirilmesi:

- o `matplotlib` kullanılarak sonuçlar görselleştirilir. Hatalı tahminler kırmızı 'x' ile işaretlenir.

Bu örnek, yarı denetimli öğrenmenin nasıl uygulanacağını ve etiketlenmemiş verilerin nasıl kullanılabileceğini göstermektedir.