

DENETİMLİ ÖĞRENME NEDİR?

Denetimli öğrenme (supervised learning), makine öğrenmesi alanında yaygın olarak kullanılan bir yaklaşımdır. Bu yöntem, etiketlenmiş veri kümeleri üzerinde çalışarak, bir modelin belirli bir girdiye karşılık gelen çıktıyı tahmin etmeyi öğrenmesini sağlar.

Denetimli Öğrenmenin İşleyişi

Denetimli öğrenme süreci genellikle üç ana adımdan oluşur: veri toplama ve etiketleme, model eğitimi ve model değerlendirme.

1. Veri Toplama ve Etiketleme:

İlk adım, modelin eğitilmesi için gerekli olan veri setini toplamaktır. Bu veriler, gerçek dünyadan veya çeşitli veri kaynaklarından toplanabilir. Toplanan verilerin etiketlenmesi, yani her verinin doğru çıktısının belirlenmesi, denetimli öğrenmenin temelidir. Örneğin, bir e-posta spam filtreleme sistemi oluşturmak için, e-postaların "spam" veya "spam değil" olarak etiketlenmesi gereklidir.

2. Model Eğitimi:

Bu adımda, etiketlenmiş veri kümesi kullanılarak bir makine öğrenmesi modeli eğitilir. Model, girdiler (x) ile bunlara karşılık gelen çıktılar (y) arasındaki ilişkileri öğrenir. Modelin öğrenme süreci, belirli bir hata fonksiyonunu minimize etmeye yönelik olarak optimize edilir. Bu süreçte çeşitli algoritmalar kullanılabilir; örneğin, regresyon analizinde doğrusal regresyon, sınıflandırma problemlerinde ise lojistik regresyon, karar ağaçları veya destek vektör makineleri (SVM) gibi yöntemler tercih edilebilir.

(SVM, makine öğrenmesinde hem sınıflandırma hem de regresyon görevleri için kullanılan güçlü ve popüler algoritmalarından biridir. SVM'ler, verileri bir hiperdüzlem (hyperplane) kullanarak ayıran doğrusal sınıflandırıcılardır ve veri setindeki örneklerin en iyi şekilde ayrılmasını hedefler.)

Denetimli Öğrenmenin Avantajları ve Dezavantajları

Avantajları:

Yüksek Doğruluk: Etiketlenmiş veriler kullanılarak eğitildiği için, genellikle yüksek doğruluk ve güvenilirlik sunar.

Kolay Anlaşılabilirlik: Çoğu denetimli öğrenme algoritması, karar sürecini açıkça gösterir, bu da sonuçların yorumlanmasını kolaylaştırır.

Geniş Uygulama Alanı: Çeşitli veri türleri ve problemler için uyarlanabilir.

Dezavantajları:

Büyük Miktarda Etiketlenmiş Veri Gereksinimi: Etiketlenmiş verilerin toplanması zaman alıcı ve maliyetli olabilir.

Öğrenme Süresi: Büyük veri kümeleri ve karmaşık modeller için uzun sürebilir.

Genelleme Problemleri: Eğitim verisine fazla uyum sağlarsa, yeni verilerde performans düşebilir (overfitting).

