

Makine Öğrenmesi, Derin Öğrenme ve Doğal Dil İşleme

Giriş

21. yüzyılda teknoloji, verinin merkezde olduğu sistemlerle evrim geçirmektedir. Bu dönüşümde, **Makine Öğrenmesi (ML)**, **Derin Öğrenme (DL)** ve **Doğal Dil İşleme (NLP)** alanları ön plana çıkmaktadır. Büyük veri, artan işlem gücü ve yeni algoritmalar sayesinde, bu alanlar sadece akademik araştırmalarda değil, günlük hayatımızda da kendini göstermektedir: sesli asistanlar, öneri sistemleri, çeviri uygulamaları ve daha fazlası...

Bu yazı; makine öğrenmesini, derin öğrenmeyi ve doğal dil işlemi anlamak isteyen birinin temelden başlayarak konuyu öğrenmesini sağlamak üzere yazılmıştır. Anlatım sadeleştirilmiş, örneklerle zenginleştirilmiş ve teorik bilgilerle donatılmıştır.

1. Makine Öğrenmesi (ML): Temel Kavramlar

1.1 Tanım

Makine öğrenmesi, bilgisayarların **verilerden öğrenmesini sağlayan** algoritmalarıdır. Geleneksel programlamada bir problemi çözmek için açık talimatlar yazmamız gerekirken, makine öğrenmesinde “veriyi ver, çözümü bulsun” yaklaşımı vardır.

Arthur Samuel’in tanımına göre:

“Makine öğrenmesi, bilgisayarların açıkça programlanmaksızın öğrenmesini sağlayan çalışma alanıdır.”

1.2 Öğrenme Türleri

• **Denetimli Öğrenme (Supervised Learning):**

Etiketli veriyle çalışır. Örnek: Elimizde 1000 meyve resmi ve her birinin ne meyve olduğu bilgisi var. Model bu örnekleri öğrenerek yeni bir meyvenin ne olduğunu tahmin etmeyi öğrenir.

- **Denetimsiz Öğrenme (Unsupervised Learning):**

Etiket yok. Algoritma, kendi kendine veri içindeki grupları bulur.

Örnek: Süpermarkette alışveriş alışkanlıklarına göre müşteri segmentasyonu.

- **Pekiştirmeli Öğrenme (Reinforcement Learning):**

Bir ortamda hareket eden bir ajan (örneğin oyun oynayan bir robot), her hareketinin sonucunda ödül veya ceza alır. Amaç, maksimum ödülü kazanmaktır.

Örnek: Satranç oynayan yapay zekâ.

- **Yarı Denetimli Öğrenme:**

Verinin az bir kısmı etiketlidir. Hem etiketli hem etiketsiz veriyi birlikte kullanır.

Örnek: Elimizde 10 bin görüntü var ama sadece 100'ü etiketli. Yarı denetimli algoritmalarla maliyeti azaltarak öğrenme sağlanır.

1.3 Temel Algoritmalar

K-En Yakın Komşu (K-NN)

Bu algoritma "yeni gelen bir şeyi, tanıdık olanlarla kıyaslayarak sınıflandırmak" mantığıyla çalışır.

Adım Adım:

1. Yeni gelen veriyi al.
2. Diğer verilerle uzaklığını ölç (genellikle Öklid uzaklığı).
3. En yakın "k" tane veri noktasını seç.
4. Bu komşuların sınıfına bak → çoğunluk hangi sınıftaysa, yeni veri de odur.

Örnek:

Sen sınıfta yenisin. Saç stilin ve giyimin 3 kişiye benziyor. Onlar bilgisayar mühendisliği öğrencisi. Muhtemelen sen de öylesin!

Avantajları:

- Açıklaması ve uygulanması çok kolay.

- Eğitim süreci yok: veri geldiği gibi saklanır.

Dezavantajları:

- Büyük veri setlerinde yavaş çalışır.
- Farklı ölçekteki özelliklerde yanlış sonuç verebilir (örn. cm ile kg birlikteyse → normalize edilmeli).

Gerçek Hayat Kullanımı:

- Netflix / Spotify → “Sana benzeyen kullanıcılar bunu da sevdi.”
- Hastane sistemleri → “Bu hastalığı olanlar bu semptomlara sahip, sen de öyleysen bu olabilir.”

Karar Ağaçları (Decision Trees)

Soru-cevap sistemine benzer. Her düğümde bir soru sorar, cevaplara göre dallanır.

Örnek:

Renk kırmızı mı?
└ Evet → Şekil yuvarlak mı?
 └ Evet → Elma
 └ Hayır → Biber

Avantaj: Yorumu kolay.

Dezavantaj: Ağaç fazla büyürse karmaşıklaşır.

Lojistik Regresyon

Bir şeyin olup olmayacağını (0 mı 1 mi) tahmin eder.

Örnek:

Kalp krizi riski var mı?

Veriler: yaş, kilo, tansiyon → Çıktı: 0 (yok) / 1 (var)

Destek Vektör Makineleri (SVM)

Verileri en iyi şekilde ayıran çizgiyi (veya düzlemi) bulur.

Örnek:

Spam ve spam olmayan e-postaları ayırmak.

Kümeleme – K-Means

Etiketli veri yok. Algoritma benzer verileri kümelere ayırır.

Örnek:

Müşterileri harcama alışkanlıklarına göre grupla. Pazarlama stratejisi her kümeye göre yapılır.

1.4 Model Değerlendirme

- **Doğruluk (Accuracy):** Ne kadar doğru tahmin ettik?
- **Kesinlik (Precision):** "Pozitif" dediklerimizin ne kadarı doğru?
- **Geri Çağırma (Recall):** Tüm gerçek pozitiflerin ne kadarını yakaladık?
- **F1 Skoru:** Precision ve recall'un dengeli ortalaması.
- **Çapraz Doğrulama:** Modelin farklı veri setleriyle test edilmesi.

2. Derin Öğrenme (Deep Learning)

2.1 Tanım

Derin öğrenme, insan beynindeki nöronları taklit eden yapay sinir ağlarını temel alan bir makine öğrenmesi alanıdır. Geleneksel makine öğrenmesinden farkı, **özellikleri elle çıkarmak yerine kendisinin öğrenmesidir.**

2.2 Sinir Ağı Nedir?

- Girdi Katmanı (input): Ham veriler (örn. piksel değerleri)
- Gizli Katman(lar) (hidden layers): Hesaplamalar burada yapılır.
- Çıktı Katmanı (output): Tahmin edilen sonuç.

2.3 Örnek: Görüntü Tanıma

Örnek:

Kedi – köpek ayırt etme.

Gelen görüntü piksellere ayrılır → sinir ağı katmanlarından geçer → “kedi” ya da “köpek” sonucu döner.

2.4 Popüler Derin Öğrenme Yapıları

• *Yapay Sinir Ağı (ANN):*

Basit yapı. Sayısal verilerde kullanılır.

• *Evreli Sinir Ağı (CNN):*

Görüntü işleme için idealdir. Filtrelerle çalışır.

Örnek:

X-ray görüntüsünde tümör tespiti.

• *Tekrarlayan Sinir Ağı (RNN) / LSTM:*

Zaman serisi verilerde, metinlerde etkilidir.

Örnek:

Cümle tamamlamaları, hava tahmini, borsa analizleri

2.5 Derin Öğrenmenin Zorlukları

- Büyük veri ihtiyacı
- Güçlü donanım (GPU)
- Yorumlanabilirlik zayıf (kara kutu problemi)

3. Doğal Dil İşleme (NLP)

3.1 Tanım

NLP, insanların konuşma ve yazma biçimini bilgisayarların anlayabileceği şekilde işler. Dilin yapısını, anlamını, bağlamını analiz eder.

3.2 Uygulamalar

- Makine çevirisi (Google Translate)
- Sesli asistanlar (Siri, Alexa)
- Sohbet botları (ChatGPT)
- Duygu analizi (mutlu/mutsuz yorum tespiti)
- Otomatik özetleme

3.3 NLP Modelleri

• *Bag of Words:*

Kelime sıklıklarını sayar.

Dezavantajı: anlam kaybı.

• *TF-IDF:*

Sık kullanılan kelimelere daha az, nadir ama önemli kelimelere daha fazla ağırlık verir.

• *Word Embeddings (Word2Vec, GloVe):*

Kelimeleri vektörlere çevirir.

Benzer anlamlı kelimeler uzayda yakın olur. (king – queen ilişkisi gibi)

- **Transformer ve BERT, GPT gibi Modeller:**

BERT: Soruları anlayabilir.

GPT: Cevap üretebilir.

Örnek:

Soru: Ankara Türkiye'nin başkenti midir?

BERT: Evet.

GPT: Evet, Ankara Türkiye'nin başkentidir. 13 Ekim 1923'te başkent ilan edilmiştir...

Sonuç

Makine öğrenmesi, derin öğrenme ve doğal dil işleme; çağımızın en kritik teknolojileri arasındadır. Bu teknolojiler sayesinde sağlık, eğitim, ulaşım ve daha birçok alan dönüşmektedir. Bu yazıda, hem teorik hem de uygulamalı olarak konular örneklerle açıklanmıştır. Bilimsel derinlik ile anlaşılabilirliği bir araya getirmek amaçlanmıştır.

Bu içerik, **BTK Akademi** eğitimlerinden edinilen bilgiler çerçevesinde, **Arda Karadağ** tarafından hazırlanmıştır.