

Makine Öğreniminin Türleri

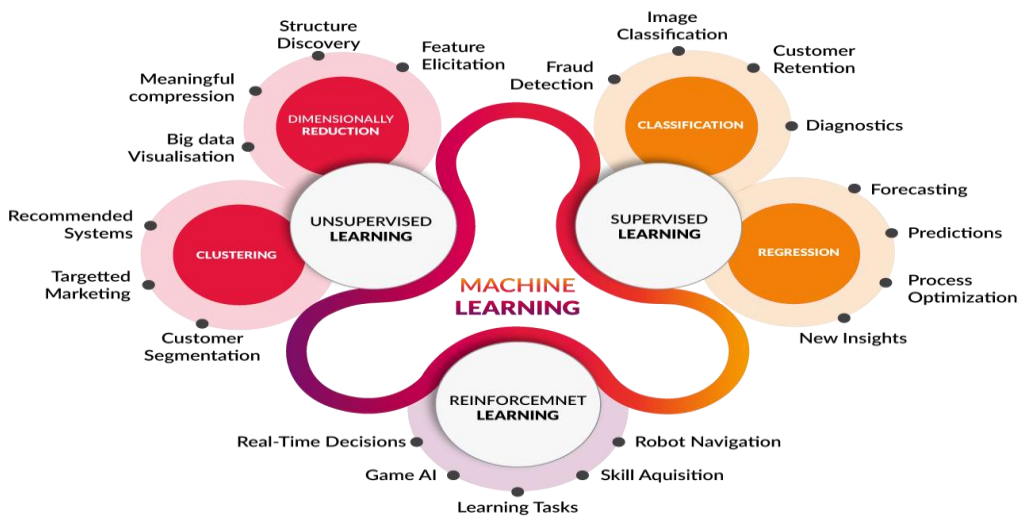
Makine öğrenmesi yaklaşımları, öğrenme sistemi için mevcut olan "sinyal" veya "geri bildirim" in doğasına bağlı olarak geleneksel olarak üç geniş kategoriye ayrılır:

Makine öğrenmesi, bir öğrenme algoritmasının, veri yığını içinde öğrenme veri setini deneyimledikten sonra görevler üzerinde doğru bir şekilde performans gösterebilmesidir. Öğrenme örnekleri, genel olarak bilinmeyen bazı olasılık dağılımından (olayların alanını temsil eden kabul edilir) gelir ve makine öğrenmesi, bu alan hakkında yeni durumlarda yeterince doğru tahminler üretmesini sağlayan genel bir model oluşturmaktadır.

Makine öğrenmesi algoritmalarının hesaplamalı analizi ve performansları, hesaplamalı öğrenme teorisi olarak bilinen teorik bilgisayar biliminin bir dalıdır. Öğrenme setleri sonlu ve gelecek belirsiz olduğundan, öğrenme teorisi genellikle algoritmaların performansının garantisini vermez. Bunun yerine, performans üzerindeki olasılık sınırları oldukça yaygındır. Eğilim-varyans ayrışması, genelleme hatasını ölçmenin bir yoludur.

Genelleme bağlamında en iyi performans için, hipotezin karmaşıklığı, verilerin altında yatan fonksiyonun karmaşıklığıyla eşleşmelidir. Hipotez işlevden daha az karmaşıksa, model verileri altına yerleştirmiştir. Yanıt olarak modelin karmaşıklığı artarsa, öğrenme hatası azalır. Ancak hipotez çok karmaşıksa, model aşırı uyuma tabidir ve genelleme daha zayıf olacaktır.

Performans sınırlarına ek olarak, öğrenme kuramcıları öğrenmenin zaman karmaşıklığını ve fizibilitesini inceler. Hesaplamalı öğrenme teorisinde, bir hesaplama polinom zamanında yapılabilirse uygulanabilir olarak kabul edilir. İki tür zaman karmaşıklığı sonucu vardır. Olumlu sonuçlar, polinom zamanında belirli bir fonksiyon sınıfının öğrenilebileceğini göstermektedir. Olumsuz sonuçlar, bazı sınıfların polinom zamanında öğrenilemeyeceğini göstermektedir.



Makine öğreniminde "Aşırı Uyum"

Makine öğreniminde, istatistiksel bir model, temelde yatan ilişki yerine rastgele hata veya gürültüyü tanımladığında "aşırı uyum" ortaya çıkar. Bir model aşırı derecede karmaşık olduğunda, eğitim veri türlerinin sayısına göre çok fazla parametrenin olması nedeniyle normal olarak aşırı uyum gözlemlenir. Model, aşırı uyumlu olan zayıf performans sergiliyor. Modeli eğitmek için kullanılan kriterler, bir modelin etkinliğini değerlendirmek için kullanılan kriterlerle aynı olmadığından, aşırı uyum olasılığı oluşur. Çok fazla veri kullanarak aşırı uydurma önlenemez, küçük bir veri kümeniz olduğundan ve siz ondan öğrenmeye çalıştığınız için aşırı uydurma nispeten gerçekleşir. Ancak küçük bir veritabanınız varsa ve buna dayalı bir modelle gelmek zorunda kalırsanız. Böyle bir durumda çapraz doğrulama olarak bilinen bir teknik kullanabilirsiniz. Bu yöntemde veri kümesi, test ve eğitim veri kümeleri olmak üzere iki bölüme ayrılır, test veri kümesi yalnızca modeli test ederken, eğitim veri kümesinde veri noktaları modelle birlikte gelir. Bu teknikte, bir modele genellikle eğitimin (eğitim veri seti) çalıştırıldığı bilinen bir veri kümesi ve modelin test edildiği bilinmeyen verilerden oluşan bir veri kümesi verilir. Çapraz doğrulama fikri, eğitim aşamasında modeli "test etmek" için bir veri kümesi tanımlamaktır.

Makine öğrenmesinin yapay zeka ile ilişkisi:

Bilimsel bir çaba olarak, makine öğrenmesi yapay zeka arayışından doğdu. Akademik bir disiplin olarak yapay zekanın (Artificial intelligence – AI) ilk günlerinde, bazı araştırmacılar makinelerin verilerden öğrenmesini istiyorlardı. Soruna çeşitli sembolik yöntemlerle ve daha sonra "sinir ağları" olarak adlandırılan yöntemle yaklaşılmaya çalışıldı; bunlar çoğunlukla algılayıcılar ve daha sonra geliştirilmiş doğrusal istatistik modellerinin yeniden icadı olarak bulunan diğer modellerdi. Olasılıksal akıl yürütme, özellikle otomatik tıbbi tanıda da kullanılmıştır.

Yapay öğrenme ile makine öğrenimi arasındaki fark nedir?

Yapay Öğrenme

Yapay öğrenme, daha geniş bir kavram olan yapay zekanın (AI) bir alt dalıdır ve makinelerin insan benzeri öğrenme yeteneklerine sahip olmasını hedefler. Bu terim, makinelerin bilgi edinme, anlamlandırma ve problem çözme yeteneklerini genel anlamda ifade eder. Yapay öğrenme, genellikle insan zekasına benzer bir şekilde öğrenme ve adaptasyon süreçlerini içerir ve kapsamı oldukça geniştir. Bu kavram, makine öğrenimi, derin öğrenme, doğal dil işleme (NLP) ve daha birçok yöntemi kapsar.

Makine Öğrenimi

Makine öğrenimi, yapay öğrenmenin bir alt kümesidir ve daha spesifik bir odaklanma alanına sahiptir. Makine öğrenimi, bilgisayar sistemlerinin doğrudan programlanmadan öğrenmelerini sağlayan algoritma ve teknikleri geliştirmeye yönelik bir disiplindir. Makine öğreniminde, algoritmalar büyük veri setleri üzerinde eğitilir ve bu verilerden kalıplar çıkararak gelecekte benzer verilerle ilgili tahminler veya kararlar verirler. Makine öğrenimi, genellikle üç ana kategoriye ayrılır: denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme.

Farklılıklar

1. Kapsam:

- **Yapay Öğrenme:** Daha geniş bir kavramdır ve makine öğrenimi dahil olmak üzere birçok farklı öğrenme yöntemi ve algoritmayı kapsar.
- **Makine Öğrenimi:** Yapay öğrenmenin bir alt kümesidir ve belirli algoritmalar ve tekniklerle sınırlıdır.

2. Uygulama Alanı:

- **Yapay Öğrenme:** Genel olarak, makinelerin insan benzeri öğrenme ve adaptasyon yeteneklerini ifade eder ve çeşitli AI teknolojilerini içerir.
- **Makine Öğrenimi:** Veri analizi ve modelleme üzerine odaklanır, verilerden öğrenme ve tahmin yapma yeteneği sağlar.

3. Metodoloji:

- **Yapay Öğrenme:** Daha genel metodolojilere dayanır, yapay sinir ağları, genetik algoritmalar, bulanık mantık gibi çeşitli yöntemler kullanabilir.
- **Makine Öğrenimi:** Spesifik algoritmalar ve modeller kullanır, örneğin, regresyon, karar ağaçları, destek vektör makineleri ve derin öğrenme algoritmaları.

4. Amaç:

- **Yapay Öğrenme:** Makinelerin genel anlamda öğrenme kapasitesini geliştirmeyi hedefler.
- **Makine Öğrenimi:** Özellikle veri üzerinde öğrenme ve tahmin yapma yeteneklerini geliştirmeyi hedefler.

Sonuç olarak, yapay öğrenme ve makine öğrenimi birbirleriyle ilişkilidir, ancak farklı odak noktaları ve uygulama alanları vardır. Yapay öğrenme, daha geniş bir kavram olarak, makine öğrenimini ve diğer AI tekniklerini kapsar; makine öğrenimi ise veri analizi ve modelleme üzerine yoğunlaşır ve bu alanda spesifik teknikler kullanır.

Makine öğrenmesinin veri madenciliği ile ilişkisi:

Makine öğrenmesi ve veri madenciliği genellikle aynı yöntemleri kullanır ve önemli ölçüde örtüşür, ancak makine öğrenmesi öğrenme verilerinden öğrenilen öngörüye odaklanırken, veri madenciliği verilerde (önceden) bilinmeyen özelliklerin keşfine odaklanır (bu veritabanlarında bilgi keşfinin analiz basamağı). Veri madenciliği birçok makine öğrenme yöntemi kullanır, ancak farklı hedefleri vardır; Öte yandan, makine öğrenmesi aynı zamanda veri denetim yöntemlerini "denetimsiz öğrenme" veya öğrenme doğruluğunu geliştirmek için bir ön işleme adımı olarak kullanır. Makine öğrenmesinde, performans genellikle Bilinen bilgiyi yeniden üretirken, bilgi keşfi ve veri madenciliği için temel görev önceden bilinmeyen bilgilerin keşfidir. Bilinen bilgilere göre değerlendirildiğinde, bilgisiz (denetimsiz) bir yöntem diğer denetimli yöntemlerle kolayca daha iyi performans gösterirken, tipik bir veri madenciliği görevinde denetimli yöntemler öğrenme verilerinin bulunamaması nedeniyle kullanılamaz.

Makine öğrenmesinin optimizasyon ile ilişkisi:

Makine öğrenmesinin de optimizasyon ile yakın bağları vardır: birçok öğrenme problemi, bir öğrenme seti örneğindeki bazı kayıp işlevlerinin en aza indirilmesi olarak formüle edilmiştir. Kayıp fonksiyonları, eğitilmekte olan modelin tahminleri ile gerçek problem örnekleri arasındaki tutarsızlığı ifade eder (örneğin, sınıflandırmada, örneklere bir etiket atamak ister ve modeller, bir dizi grubun önceden atanmış etiketlerini doğru şekilde tahmin etmek için eğitilir örnekler). İki alan arasındaki fark genelleme amacından kaynaklanmaktadır: optimizasyon algoritmaları bir öğrenme setindeki kaybı en aza indirirken, makine öğrenmesi görünmeyen numunelerdeki kaybı en aza indirmekle ilgilidir. Parametrelerin hassasiyetinin belirlenmesinde optimizasyon algoritmaları, makine öğrenmesinde kritik rol oynamaktadır.

Makine öğrenmesinin istatistiklerle ilişkisi:

Makine öğrenmesi ve istatistikler, yöntemler açısından yakından ilişkili alanlardır, ancak temel amaçlarında farklıdır: İstatistik, bir örneklemden analiz etmeye ya da yorumlamaya yönelik çıkarımlar elde ederken, makine öğrenmesi genelleştirilebilir tahmin modelleri bulur. Metodolojik prensiplerden teorik araçlara kadar makine öğrenmesi fikirleri istatistiklerde uzun bir geçmişe sahiptir. Ayrıca, veri bilimi terimi, genel alanı çağırarak için bir yer tutucudur. İstatistiksel modelleme paradigmasını ayırt edilirse: veri modeli ve makine öğrenme algoritmaları aşağı yukarı, makine öğrenmesi yöntemlerinin benimsediği istatistiksel öğrenme birleşik bir alana yol açmıştır.

İstatistik ve Makine Öğrenimi verileri tanımlamak, analiz etmek ve modellemek için işlevler ve uygulamalar sağlar. Açıklayıcı veri analizi için tanımlayıcı istatistikleri, görselleştirmeleri ve kümelemeyi kullanabilir, olasılık dağılımlarını verilere uydurabilir, Monte Carlo simülasyonları için rasgele sayılar üretebilir ve hipotez testleri gerçekleştirebilirsiniz. Regresyon ve sınıflandırma algoritmaları, Sınıflandırma ve Regresyon Öğrenici uygulamalarını kullanarak etkileşimli olarak verilerden çıkarımlar yapmanıza ve tahmine dayalı modeller oluşturmanıza olanak tanır.

Çok boyutlu veri analizi ve özellik çıkarma için araç kutusu, değişkenleri en iyi tahmin gücüyle tanımlamanıza izin veren temel bileşen analizi (PCA), düzenlileştirme, boyutsallık azaltma ve özellik seçme yöntemleri sağlar.

Araç kutusu, destek vektör makineleri (SVM'ler), artırılmış karar ağaçları, k-ortalamlar ve diğer kümeleme yöntemleri dahil olmak üzere denetimli, yarı denetimli ve denetimsiz makine öğrenimi algoritmaları sağlar. Kısmi bağımlılık çizimleri ve LIME gibi yorumlanabilirlik tekniklerini uygulayabilir ve gömülü dağıtım için otomatik olarak C/C++ kodu oluşturabilirsiniz. Birçok araç kutusu algoritması, bellekte saklanamayacak kadar büyük veri kümelerinde kullanılabilir.

Makine öğrenmesi ve matematik:

Faydalı sonuçlar elde etmek için, belirli genel makine öğrenimi ilkeleri ve bireysel algoritmaların iç işleyişi hakkında iyi matematiksel sezgilere ihtiyaç var.

İyi bir matematik alt yapısı ile,

- Problem için doğru algoritmalar seçilir.
- Parametre ayarları, doğrulama stratejileri hakkında iyi seçimler yapılır.
- Fazla veya yetersiz uyum tanınır.

- Yetersiz ya da belirsiz sonuçlar giderilir.
- Sonuçlara uygun güven ya da belirsizlik sınırları konulur.
- Kodlama algoritmalarında daha iyi bir iş yapılır veya onlar daha karmaşık hale getirilir.

Makine öğreniminde hipotezler veya model oluşturmanın üç aşaması:

- Model oluşturma
- Model testi
- Modeli uygulamak

Makine Öğrenimindeki farklı yaklaşımlar:

- Konsept ve Sınıflandırma Öğrenimi
- Sembolik vs İstatistiksel Öğrenme
- Endüktif Vs Analitik Öğrenme Makine Öğreniminin iki tekniği şunlardır:
- Genetik Programlama
- Endüktif Öğrenme

Makine Öğrenimindeki farklı teknik türleri:

Makine öğrenmesi algoritmalarının türleri, yaklaşımlarına, girdikleri ve çıktıkları veri türlerine ve çözmeleri amaçlanan görev veya sorun türlerine göre farklılık gösterirler.

- Denetimli Öğrenme (Supervised Learning Algorithms)
- Denetimsiz Öğrenme (Unsupervised Learning Algorithms)
- Yarı denetimli Öğrenme
- Takviye Öğrenme
- Transdüksiyon
- Öğrenmeyi öğrenmek