

4.1. Makine Öğrenmesi Algoritmaları

Makine Öğrenimi, kurala dayalı çıkarımlar yapar. İlişkilendirme kuralı öğrenme, büyük veritabanlarındaki değişkenler arasındaki ilişkileri keşfetmek için kural tabanlı bir makine öğrenme yöntemidir. Makine öğrenmede veritabanlarında keşfedilen güçlü kuralların bir miktar "ilginçlik" ölçüsü kullanılarak belirlenmesi amaçlanmıştır.

Kural tabanlı öğrenme, bilgiyi depolamak, manipüle etmek veya uygulamak için "kuralları" tanımlayan, öğrenen veya geliştiren herhangi bir makine öğrenme yöntemi için genel bir terimdir. Kural tabanlı öğrenme algoritmasının belirleyici özelliği, sistem tarafından yakalanan bilgileri topluca temsil eden bir dizi ilişkisel kuralın tanımlanması ve kullanılmasıdır. Bu, bir tahminde bulunmak için herhangi bir örneğe evrensel olarak uygulanabilen tekil bir modeli yaygın olarak tanımlayan diğer makine öğrenme algoritmalarının aksine, kural tabanlı makine öğrenme yaklaşımları, öğrenme sınıflandırıcı sistemleri, ilişkilendirme kuralı öğrenme ve yapay bağımsızlık sistemlerini içerir.

Makine öğrenmesi algoritmalarının temel bileşenleri:

- ☐ Hesaplamalar için çeşitli kütüphaneler kullanıldığından gerekli kütüphaneler aktarılır.
- ☐ Veritabanı dosyaları okunur.
- ☐ Etiketli verilerin ağırlıklı ortalama, standart sapma gibi istatistiksel analiz özellikleri belirlenir ve yorumlanır.
- ☐ Verilerin grafiği çizilir.
- ☐ Değerler tahmin edilirken dikkate almak istenilen özellikler ve katsayılar seçilir.
- ☐ Bir modelin doğruluğunu kontrol etmek için, veriler eğitim ve test veri setlerine ayrılır.
- ☐ Model eğitilir
- ☐ Test veri seti için bir tahmin fonksiyonu bulunur.
- ☐ Test verilerinin doğruluğu kontrol edilir: Gerçek değerleri veri setindeki tahmin edilen değerlerle karşılaştırılarak bir modelin doğruluğu kontrol edilebilir.

Makine Öğreniminin beş popüler algoritması:

- ☐ Karar ağaçları
- ☐ Sinir Ağları (geri yayılım)
- ☐ Olasılık ağları
- ☐ En yakın komşu
- ☐ Vektör makineleri desteklemek

Algoritmadan bağımsız makine öğrenimi nedir?

Matematiksel temellerin belirli bir sınıflandırıcıdan bağımsız olduğu veya öğrenme algoritmasının algoritmadan bağımsız makine öğrenimi olarak adlandırıldığı yerlerde makine öğrenimi?

Makine Öğreniminde Algılayıcı Nedir?

Makine Öğreniminde, Perceptron, girdinin birkaç olası ikili olmayan çıktıdan birine denetimli sınıflandırılması için bir algoritmadır.

İlişkisel değerlendirme tekniklerinin önemli bileşenleri şunlardır:

- ☐ Veri toplama
- ☐ Ground Truth Edinimi
- ☐ Çapraz Doğrulama Tekniği
- ☐ Sorgu Türü
- ☐ Puanlama Metriği
- ☐ Önem Testi

Robotik ve bilgi işlemede sıralı tahmin probleminin ortaya çıktığı alanlar:

- ☐ Taklit Öğrenme
- ☐ Yapılandırılmış tahmin
- ☐ Model tabanlı pekiştirmeli öğrenme

Makine Öğrenmesi Algoritmaları:

1- Regresyon (Tahmin)

Sürekli değerleri tahmin etmek için regresyon algoritmaları kullanılır.

Regresyon algoritmaları:

- Doğrusal Regresyon
- Polinom Regresyon
- Üstel Regresyon
- Lojistik regresyon
- Logaritmik Regresyon

2-Sınıflandırma

Bir dizi öğenin sınıfını veya kategorisini tahmin etmek için sınıflandırma algoritmaları kullanılır.

Sınıflandırma algoritmaları:

- K-En Yakın Komşular
- Karar ağaçları
- Rastgele Orman
- Destek Vektör Makinesi
- Naive Bayes

3- Kümeleme

Özetlemek veya verileri yapılandırmak için kümeleme algoritmaları kullanılır.

Kümeleme algoritmaları:

- ☐ K-means
- ☐ DBSCAN
- ☐ Mean Shift
- ☐ Hierarchical

4- İlişkilendirme

Birlikte meydana gelen öğeleri veya olayları ilişkilendirmek için ilişkilendirme algoritmaları kullanıyoruz.

İlişkilendirme algoritmaları:

- Apriori

5- Anomali (Sapma) Algılama:

Anormal etkinlikleri ve dolandırıcılık tespiti gibi olağandışı durumları keşfetmek için anormallik algılama kullanılır.

Veri madenciliğinde, aykırı tespit olarak da bilinen anomali tespiti, verilerin çoğundan önemli ölçüde farklılık göstererek şüphe uyandıran nadir maddelerin, olayların veya gözlemlerin tanımlanmasıdır. Tipik olarak, anormal kelimeler banka sahtekarlığı, yapısal bir kusur, tıbbi sorunlar veya bir metindeki hatalar gibi bir konuyu temsil eder. Anomalilere aykırı değerler, yenilikler, gürültü, sapmalar ve istisnalar denir.

Özellikle, kötüye kullanım ve ağ izinsiz giriş tespiti bağlamında, ilginç nesneler genellikle nadir nesneler değil, faaliyette beklenmedik patlamalardır. Bu model, bir aykırı değer nadir bir nesne olarak ortak istatistiksel tanımına uymaz ve uygun şekilde toplanmadığı sürece birçok aykırı algılama yöntemi (özellikle denetimsiz algoritmalar) bu verilerde başarısız olur. Bunun yerine, bir küme analiz algoritması bu örüntüler tarafından oluşturulan mikro kümeleri tespit edebilir.

Üç geniş anomali tespit tekniği kategorisi bulunmaktadır. Gözetimsiz anomali tespit teknikleri, veri kümesindeki örneklerin çoğunun normal olduğu varsayımıyla etiketlenmemiş bir test veri kümesindeki anormallikleri, veri kümesinin geri kalan kısmına en az uyan görünen örnekleri arayarak tespit eder. Denetimli anomali algılama teknikleri, "normal" ve "anormal" olarak etiketlenmiş ve bir sınıflandırıcıyı (diğer birçok istatistiksel sınıflandırma problemi için temel fark, aykırı algılamanın doğasında dengesiz doğasıdır) içeren bir veri seti gerektirir. Yarı denetimli anomali tespit teknikleri, belirli bir normal eğitim veri setinden normal davranışı temsil eden bir model oluşturur ve daha sonra model tarafından bir test örneğinin üretilme olasılığını test eder.

6- Sıra Desen Madenciliği

Örüntü - Pattern: Bir nesnenin ya da olayın iki veya üç boyutlu, uzaysal ve geometriksel davranışının matematiksel ifadesidir. Diğer bir ifadeyle , nesnenin davranışı ile ilgili uzayda gözlenebilir veya ölçülebilir geometrik bilgilerdir. Tersinden desen madenciliği. Örüntü hazırlanır. Veri yığını içerisinde gezinir. Hedef geldiğinde uyanır, kendisini kabul ettirir. Ayrıca veri yığı içerisinde dolaşan ve arayan desenler geliştirilmektedir.

Örüntü Tanımanın kullanıldığı alanlar:

- ☐ Örüntü Tanıma şu alanlarda kullanılabilir:
- ☐ Bilgisayar görüşü
- ☐ Konuşma tanıma
- ☐ Veri madenciliği
- ☐ İstatistik
- ☐ Gayri Resmi Erişim
- ☐ Biyo-Bilişim

Bir dizideki veri örnekleri arasındaki sonraki veri olaylarını tahmin etmek için sıralı model madenciliği kullanıyoruz.

Dünya ve ötesinde kainat sürekli veri oluşturan sinyal kaynakları (Elektromanyetik, elektrik, ısı, renk, ses, titreşim, çekimsel kuvvetler, ...) ile doludur. Bu kaynaklarından yayılan sinyal örüntüleri yayıldıkları ortamlar ile etkileşime girerler. O halde bu örüntülerin davranış değişikliklerinden kestirimsel tahminler yapılabilir.

7- Boyut Küçültme (Dimensionality reduction)

Makine Öğrenimi ve istatistikte boyut küçültme, dikkate alınan rastgele değişkenlerin sayısını azaltma işlemidir ve özellik seçimi ve özellik çıkarımı olarak ikiye ayrılabilir.

Bir veri kümesinden yalnızca yararlı özellikleri çıkarmak için verilerin boyutunu küçültmek için boyut azaltma kullanılır. Boyut azaltma, denetimsiz bir öğrenme tekniğidir.

Veri biliminde, boyut indirgeme, bir verinin yüksek boyutlu bir uzaydan, düşük boyutlu bir uzaya, anlamını kaybetmeyecek şekilde dönüştürülmesidir. Yüksek boyutlu bir veriyi işlemek daha fazla işlem yükü gerektirir. Bu yüzden, yüksek sayıda gözlemin ve değişkenin incelendiği sinyal işleme, konuşma tanıma, nöroinformatik, biyoinformatik gibi alanlarda boyut indirmesi sıkça kullanılır.

Boyut indirgeme yaklaşımları doğrusal ve doğrusal olmayan olarak ikiye ayrılır. Boyut indirgeme var olan özneliklerin bir alt kümesini seçerek ya da yeni öznelikler çıkararak yapılabilir. Boyut indirmesi gürültü filtreleme, veri görselleştirme ya da kümeleme analizi amacıyla kullanılabileceği gibi, diğer makine öğrenimi yöntemlerinin ön adımı olarak uygulanabilir.

8- Önerilerden eğilim ya da yön bulma

Öneri motorları oluşturmak için öneri algoritmalarını kullanılır.

Örnekler:

- Netflix öneri sistemi.
- Bir kitap tavsiye sistemi.
- Amazon'da bir ürün öneri sistemi.

Günümüzde yapay zeka, makine öğrenimi, derin öğrenme ve diğerleri gibi pek çok vızıltılı kelime duyuyoruz.

Makine öğrenimi algoritmalarını uygulamak için neden Python?

Python, popüler ve genel amaçlı bir programlama dilidir. Python kullanarak makine öğrenimi algoritmaları yazabiliriz ve bu iyi çalışır. Python'un veri bilimcileri arasında bu kadar popüler olmasının nedeni, Python'un hayatımızı daha rahat hale getiren çok çeşitli modül ve kütüphanelerin halihazırda uygulanmış olmasıdır.

Bazı heyecan verici Python kütüphanelerine kısaca bir göz atalım:

1. Numpy: Python'da n boyutlu dizilerle çalışmak için bir matematik kitaplığıdır. Hesaplamaları etkili ve verimli bir şekilde yapmamızı sağlar.
2. Scipy: Sinyal işleme, optimizasyon, istatistik ve çok daha fazlasını içeren sayısal algoritmalar ve alana özgü araç kutusudur. Scipy, bilimsel ve yüksek performanslı hesaplamalar için işlevsel bir kitaplıktır.
3. Matplotlib: 2D çizimin yanı sıra 3D çizim de sağlayan modaya uygun bir çizim paketidir.
4. Scikit-learn: Python programlama dili için ücretsiz bir makine öğrenimi kitaplığıdır. Sınıflandırma, regresyon ve kümeleme algoritmalarının çoğuna sahiptir ve Numpy, Scipy gibi Python sayısal kütüphaneleri ile çalışır.