Makine öğrenmesi, bilgisayarların verilerden öğrenmesini ve bu öğrenmeyi kullanarak gelecekteki verilere dayalı tahminler yapmasını sağlayan bir yapay zeka dalıdır. Burada amaç, bilgisayarların herhangi bir programlama müdahalesine gerek kalmadan, verilerdeki örüntüleri tespit edip bunlardan yararlanarak karar vermesidir.

**Yapay Zeka ve Makine Öğrenmesi Arasındaki Farklar**

Yapay zeka, insanın düşünme süreçlerini taklit eden geniş bir kavramdır ve problem çözme, dil anlama gibi zeka gerektiren görevleri üstlenir. Makine öğrenmesi ise yapay zekanın bir alt dalı olup, bilgisayarların verilerden öğrenmesini sağlayarak, belirli görevlerde daha iyi performans göstermelerini amaçlar. Yani, her makine öğrenmesi uygulaması yapay zeka kullanır, fakat yapay zekanın tüm uygulamaları makine öğrenmesiyle ilgili değildir.

**Makine öğrenmesi türleri**

1. **Denetimli Öğrenme**: Bu türde model, etiketli veriler ile eğitilir. Her veri parçası doğru sonuçla etiketlenmiştir ve model bu veriler üzerinden öğrenir.
2. **Denetimsiz Öğrenme**: Burada etiketli veri yoktur ve model, verilerdeki gizli yapıları veya kümeleri bulmaya çalışır.
3. **Pekiştirmeli Öğrenme**: Bir ajan çevresiyle etkileşime girer ve aldığı ödüller veya cezalarla nasıl hareket etmesi gerektiğini öğrenir.

**Makine Öğrenmesinin Uygulama Alanları**

* **Finans**: Kredi risk analizi, algoritmik ticaret
* **Sağlık**: Hastalık teşhisi, hasta takibi
* **Otomotiv**: Otonom sürüş sistemleri
* **Pazarlama**: Müşteri Segmentasyonu

**Sayısal Veri**: Sayısal bir değere sahip olan ondalıklı veya tamsayı içeren verilerdir.

**Kategorik Veri**: Sıralama içeren veya içermeyen belirli kategorilere ayrılan verilerdir.

**İstatistiksel Temeller**

**Ortalama**: Verideki sayısal değerlerin toplanıp verilerin adetine bölünmesiyle bulunur.

**Medyan**: Sıralı bir veride ortanca değerdir.

**Varyans**: Her bir değerin ortalamadan farklarının karelerinin ortalamasıdır.

**Standart Sapma**: Varyansın kareköküdür.

**Doğrusal Regresyon**

Doğrusal regresyon, bağımlı bir değişkenle bağımsız değişkenler arasındaki doğrusal ilişkiyi modellemeye yönelik bir yöntemdir. Temelde, modelin amacı en iyi doğrusal denklemi bulmak ve bu denklem sayesinde bağımsız değişkenlerdeki değişimlerin bağımlı değişken üzerinde nasıl bir etkisi olacağını tahmin etmektir.

* **Bağımlı Değişken**: Tahmin edilmek istenen değişkendir.
* **Bağımsız Değişkenler**: Bağımlı değişkeni etkileyen faktörlerdir.

**Sınıflandırma Nedir?**

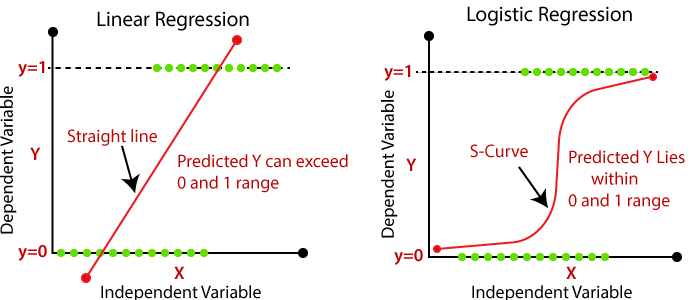
Sınıflandırma, verilerin önceden belirlenmiş kategorilere ayrılması işlemidir. Bir model, eğitim sırasında öğrendiği örüntüler sayesinde, yeni gelen veriyi doğru sınıfa yerleştirir.

**Lojistik Regresyonun Temel Mantığı**

Lojistik regresyon, bağımlı değişkenin yalnızca iki sınıftan birine ait olduğu durumlarda kullanılan bir sınıflandırma tekniğidir. Temelde, doğrusal bir kombinasyonun sonucunu bir sigmoid (lojistik) fonksiyona sokarak, 0 ile 1 arasında bir olasılık değeri üretir. Bu olasılık değeri, genellikle 0.5 ile karşılaştırılır ve sınıf kararı verilmiş olur.

**Modelin Doğruluğunu Ölçme Kavramları**

* **Accuracy**: Modelin doğru sınıflandırdığı örneklerin toplam örnek sayısına oranıdır. Formül:
* **Precision**: Modelin pozitif olarak etiketlediği örneklerin, gerçekten pozitif olma oranını ölçer. Formül:



**Karar Ağaçlarının Çalışma Prensibi**

Karar ağaçları, veriyi özelliklerine göre dallara ayırarak karar verme sürecini görselleştirir ve basitleştirir. Her bir düğümde, veriler belirli bir özellik üzerinden iki ya da daha fazla dala ayrılır. Bu ayrımlar, veri setindeki en anlamlı farkları ortaya çıkaracak özelliklere göre yapılır. Sonuçta, ağacın yapraklarında sınıflandırma ya da tahmin sonuçları elde edilir.

**Sınıflandırma Ağaçları**: Veriyi belirli kategorilere ayırır ve her yaprak düğümünde bir sınıf etiketi verir.

**Regresyon Ağaçları**: Sürekli değerleri tahmin eder. Her yaprak düğümünde, örneklerin ortalaması gibi sürekli bir değer hesaplanır.

diyagram, ekran görüntüsü, çizgi, plan içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**K-NN**

K-NN, bir örneğin sınıfını veya değerini belirlemek için eğitim verisindeki en yakın "k" komşusunu inceler. Komşuluk, genellikle Öklidyen mesafe gibi metrikler kullanılarak hesaplanır.

K değeri, en yakın komşu sayısını belirler. K değeri modelin performansını çok büyük derecede etkiler.

metin, kalıp, desen, düzen içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Model Başarısını Ölçmek İçin Kullanılan Metrikler**

**Accuracy**: Modelin doğru tahmin ettiği örneklerin toplam örnek sayısına oranıdır.

**Hata Oranı**: Yanlış tahmin edilen örneklerin toplam örnek sayısına oranıdır. Doğruluk ile ters orantılıdır.

**F1 Skoru**: Precision ve recall değerleri ile hesaplanır. Dengesiz sınıf dağılımlarında iyi bir değerlendirme ölçütüdür.

Formülü:

**Aşırı Öğrenme (Overfitting) Nedir?**

Aşırı öğrenme, modelin eğitim verisindeki her detayı aşırı derecede öğrenip, eğitim verisinin gürültü ve özel durumlarını da ezberlemesidir. Bu, modelin test verisi üzerinde düşük performans göstermesine neden olur. Örneğin, bir karar ağacı modeli, eğitim verisindeki her küçük farkı ayırıcı kriter olarak kullanırsa, eğitimde yüksek doğruluk gösterse de yeni verilere uyum sağlayamaz.