

Makine Öğrenmesi (Machine Learning)

Erdem Korhan Akçay

2022-10-11

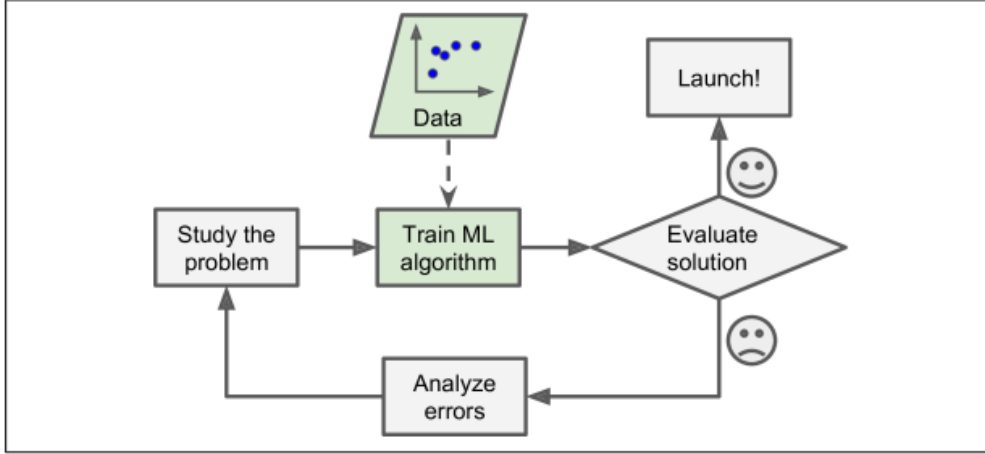
1) Makine Öğrenmesi Nedir?

Makine öğrenmesi, bilgisayarların verilerden çıkarım yaparak, öğrenim sağlayan bir bilim veya sanattır.

Arthur Samuel'in de dediği gibi makine öğrenmesi bilgisayarlara açıkça programlanmadan öğrenebilme yeteneği verildiği bir çalışma alanıdır.

Aynı zamanda mühendislik tanımını yapacak olursak, bilgisayar amacına (T) göre deneyim kazanarak (E) öğrenir ve performans ölçüsüyle (P) değerlendirir.

Örnek verecek olursak, spam filtreleme işlemi bir makine öğrenmesi uygulamasıdır, spam olarak işaretlenmiş e-postalardan ve normal e-postalardan neyin spam olup olmadığını öğrenir. Bu örnekte sistemin öğrenmesi için kullandığı veri seti "Eğitim seti" olarak adlandırılır. Burada, görevimiz -amacımız- (T) yeni gelecek e-postaların spam kararını vermek, deneyimimiz (E) eğitim verisi ve performans ölçüsü (P)'dir. Verilen örnekte performans ölçüsü doğru sınıflandırılmış e-postaların oranı olabilir. Aynı zamanda performans ölçüsüne **doğruluk** (accuracy) denir ve genel olarak sınıflandırma görevlerinde kullanılır.



Şekil 1. Makine Öğrenmesi yaklaşımı

1.1) Kaça Ayrılır ve Yöntemleri Nelerdir?

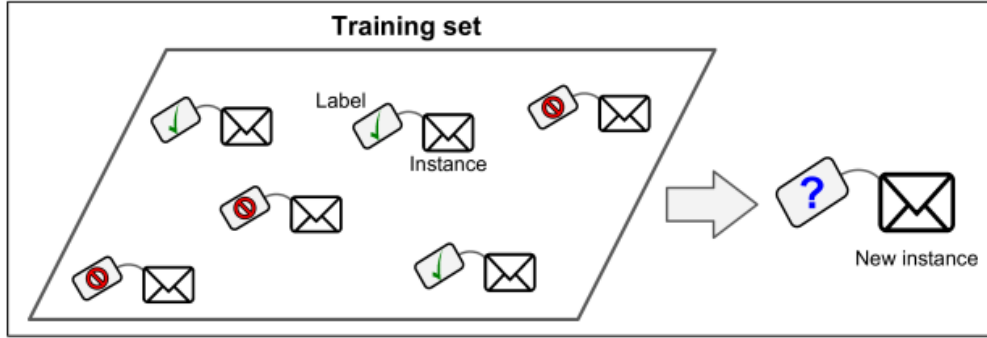
Makine öğrenmesinde uygulanabilecek bir çok model vardır. Fakat bunları geniş kategorilere ayıracak olursak:

- İnsanlar denetiminde eğitilip eğitilmeyen (Denetimli Öğrenme, Denetimsiz Öğrenme, Yarı-denetimli Öğrenme, Pekiştirmeli Öğrenme)
- Kademeli olarak anında öğrenip öğrenemeyen (Batch Öğrenme)
- Bilinen veri noktalarını, yeni veri noktalarıyla eski veri noktalarını karşılaştırarak öğrenen veya bunun yerine eğitim setindeki verilerin yapılarını tanımlayıp tahmine dayalı bir model oluşturan. (Örnek Tabanlı ve Model Tabanlı Öğrenme)

Bu ders kapsamında Denetimli, Denetimsiz ve Pekiştirmeli Öğrenme incelenecektir.

1.2) Denetimli Öğrenme(Supervised Learning):

Denetimli öğrenmede algoritmayı beslediğimiz eğitim verileri(training data), etiketlenen(labels) verileri ve istenen çözümleri içerir.

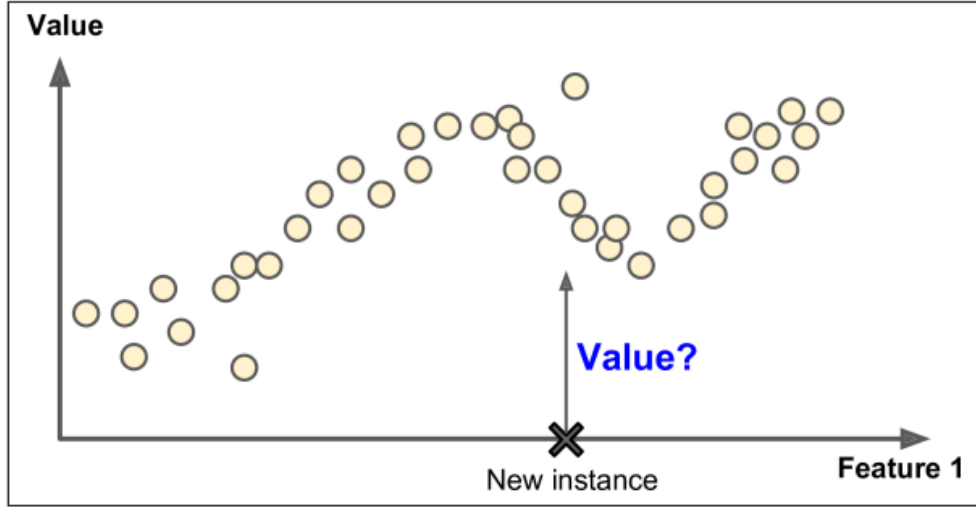


Şekil 1.1 Denetimli öğrenmede etiketlenmiş veri seti (Spam sınıflandırma örneği)

İki çeşit denetimli öğrenme vardır. Bunlardan biri sınıflandırma, spam sınıflandırması bunun güzel bir örneğidir. Bir çok örnek e-postalardan (spam olan, olmayan) öğrenir ve onları nasıl sınıflandırılacağını öğrenir.

Bir diğer çeşit ise sayıların tahmini, araba fiyatlandırması buna bir örnek olabilir. Verilen özellikler (km, yaş, marka vb.) değişken olarak adlandırılır. Buna benzer algoritmalar regresyon olarak adlandırılır. Modeli eğitmek için yeterince araba verisine ihtiyaç vardır.

#fotoğrafları açıkla



Aşağıda bir kaç önemli ve çok kullanılan denetimli öğrenme algoritmaları gösterilmektedir:

- K-En Yakın Komşu (K-Nearest Neighbors)
- Doğrusal Regresyon (Linear Regression)
- Lojistik Regresyon (Logistic Regression)
- Destek Vektör Makineleri (Support Vector Machines)
- Karar Ağaçları ve Rassal Ormanlar (Decision Trees and Random Forests)

1.3) Denetimsiz Öğrenme(Unsupervised Learning):

Denetimsiz öğrenmede, tahmin edileceği üzere, eğitim seti etiketlenmemiştir. Sistem bir eğitmen olmadan öğrenmeye çalışır.

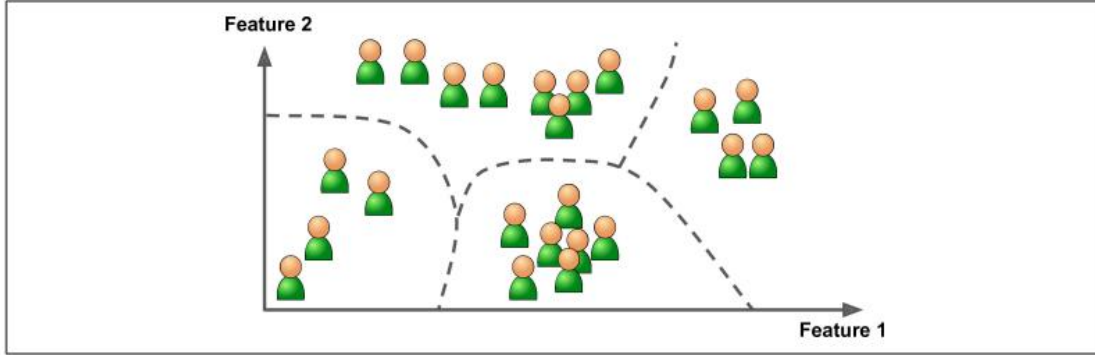
Aşağıda bir kaç önemli denetimsiz öğrenme algoritmaları gösterilmektedir:

- **Kümeleme(Clustering)**
 - K Ortalamalar(K Means)
 - Hiyerarşik Kümeleme (Hierarchical Cluster)
- **Anomali Tespiti(Anomaly Detection)**
 - Tek sınıflı Destek Vektör Makinesi (One-class SVM)
 - İzolasyon Ormanlar (Isolation Forest)
- **Görselleştirme ve Boyut Azaltma (Visualization and Dimensionality Reduction)**
 - Temel Bileşen Analizi (Principal Component Analysis)
 - Kernel PCA

• Birliktelik Kuralı(Association Rule)

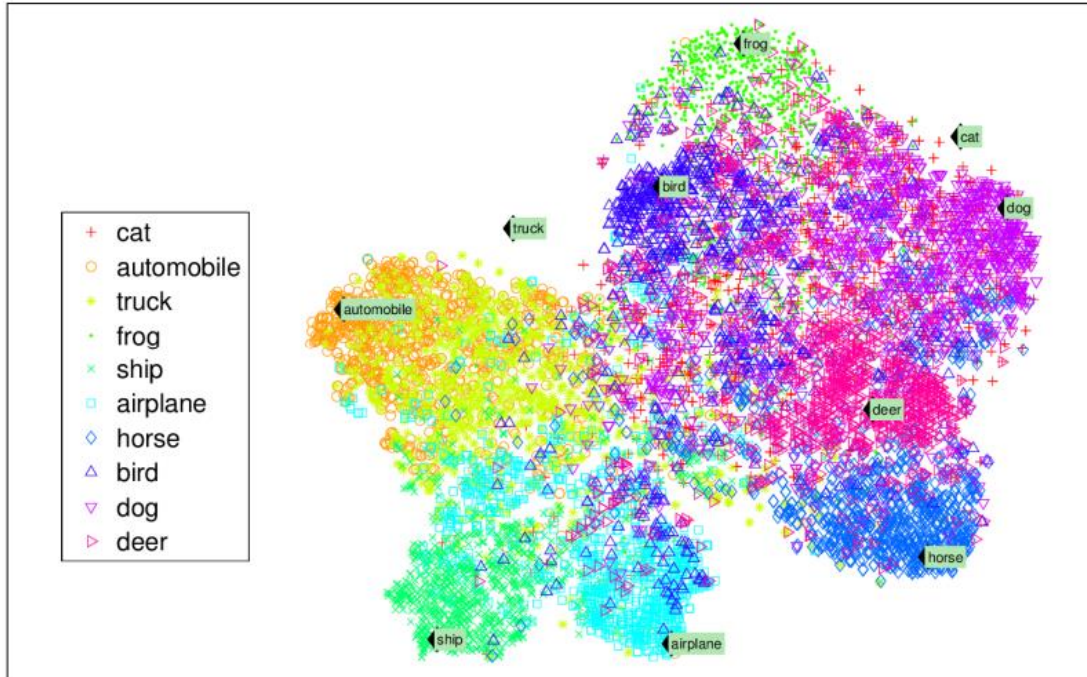
- Apriori
- Eclat

Örneklerini verecek olursak, bir blog sitesine giren ziyaretçilerin hangi gruba ait olduğunun tespitinde kümeleme algoritmasını kullanabiliriz.



Şekil 1.3 Kümeleme örneği

Aynı zamanda görselleştirme(visualization) denetimsiz öğrenme algoritmalar için iyi bir örnektir. Algoritmayı karmaşık(complex) ve etiketlenmemiş(unlabeled) veriyle besledikten sonra bizlere 2D veya 3D çizimlerini(plot) gösterir. Bu algoritmalar mümkün olduğu kadar yapıyı korumaya çalışır.



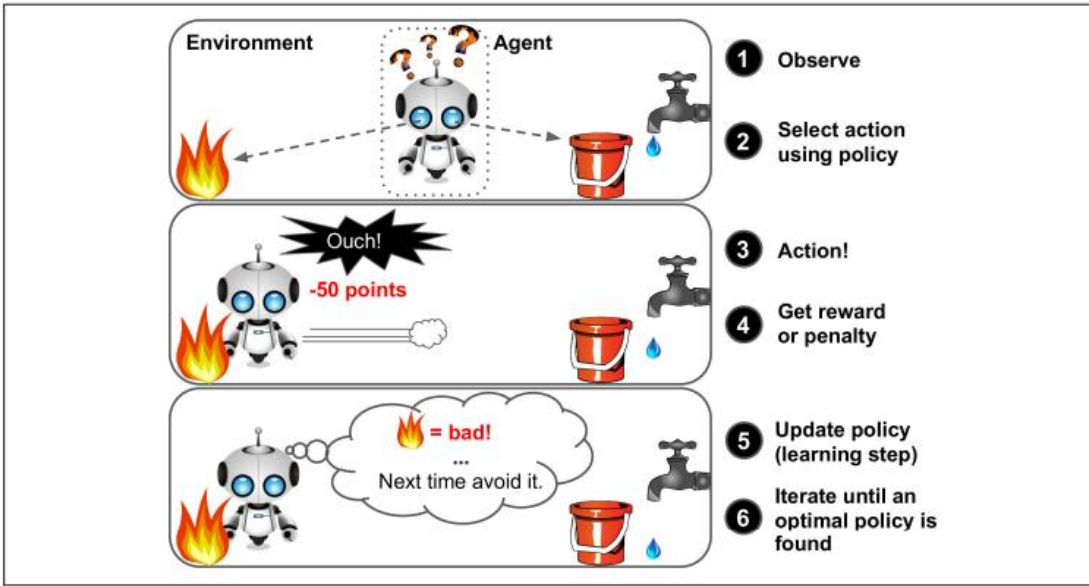
Şekil 1.4 Görselleştirme örneği

Boyut azaltmada ise, amacımız veri hakkında fazla bilgi kaybetmeden olabildiğince sadeleştirmektir. Bunu yapmanın yollarından biri bir kaç ilişkili özelliği bir arada tutmaktır. Örnek verecek olursak arabanın yaptığı kilometresi, yaşıyla doğru orantılı olabilir, bu şekilde boyut azaltma algoritmasında bu özellikleri bir araya getirir.

Başka bir makine algoritmasını kullanmadan önce boyut azaltma algoritmasıyla eğitim setinin boyutunu azaltmak, daha hızlı, daha az ram kullanarak sonuca varırız ve bazı durumlarda daha iyi performans gösterebilir.

1.4)Pekiştirmeli Öğrenme (Reinforcement Learning):

Pekiştirmeli öğrenme diğerlerine göre farklıdır. Öğrenme yapısını ajan(agent) diye kodluyoruz. Ajan çevreyi ve seçenekleri gözlemler ve aksiyona(action) geçer. Bu aksiyondan sonra ödül(reward) kazanır veya ceza(penalty) puanı yer. Bu mekanizmayla en iyi stratejiyi geliştirene kadar öğrenir, buna da zeka kıvraklığı (policy) diyebiliriz, mümkün olan en kısa zamanda olabildiğince çok ödül kazanmaya çalışır.



Şekil 1.5 Pekiştirmeli Öğrenme

Örnek olarak, bir çok robot yürümeyi öğrenmek için pekiştirmeli öğrenme algoritmaları kullanıyor. Aynı zamanda DeepMind's AlphaGo programı pekiştirmeli öğrenme için iyi bir örnektir. AlphaGo Mayıs 2017'de Go oyununda dünya şampiyonu Ke Jie'ye karşı kazandı. Bunu milyonlarca oyunu analiz ederek ve kendi kendine oynayarak stratejisini geliştirerek başardı.

İstatistikle İlişkisi Nedir?

İstatistiksel öğrenme, makine öğrenmesinin amaçları çok benzese de aralarındaki en büyük fark çıkarımdan(inference) gelmektedir. İstatistiksel öğrenme verideki değişkenler arasındaki ilişki ve açıklayıcılığı ile ilgilenirken, makine öğrenmesi tahminlemenin doğruluk oranı(accuracy) ile ilgilenmektedir. Ayrıca makine öğrenmesinde ki algoritma çeşitliliği, yan ve varyans dengesinin esnek olması ve çok yüksek boyutlu verilerde tahminleme gücünün

yüksek olması gibi avantajları ile büyük veri dünyasında çok hızlı şekilde yerini almasını sağladı.

REFERENCES:

Aurélien Géron (2019), Hands-On Machine Learning with Scikit-Learn and TensorFlow, syf.(8-15)