

Yapay Zeka Nedir?

A. Talha Yalta

TOBB Ekonomi ve Teknoloji Üniversitesi

İKT-457 Ekonomi ve Finans İçin Yapay Zeka 1
Sürüm 0,92 (Güz 2020)



Bu belge “Creative Commons Attribution-ShareAlike 3.0 Unported” (CC BY-SA 3.0) lisansı altında bir açık ders malzemesi olarak genel kullanıma sunulmuştur. Bazı şekiller “An Introduction to Statistical Learning, with applications in R” (Springer, 2017) kitabından yazarların izniyle alınmıştır. Tüm belge eserin ilk sahibinin belirtilmesi ve geçerli lisansın korunması koşuluyla özgürce kullanılabilir, çoğaltılabilir, ve değiştirilebilir. Creative Commons örgütü ve CC-BY-SA 3.0 lisansı ile ilgili ayrıntılı bilgi <https://creativecommons.org> Internet adresinde yer almaktadır. Ders notlarımın güncel sürümlerine <http://yalta.etu.edu.tr> adresinden ulaşabilirsiniz.

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
2020 



- 1 Yapay Zeka Nedir?
 - Ekonomi ve finansta yapay zeka
- 2 Yapay Zekanın Uygulanması
 - Makine öğrenmesi
 - İstatistiksel öğrenme
- 3 Örnek Uygulamalar
 - Ücret verileri
 - Borsa verileri
 - Gen anlatım verileri



Ders Planı

- 1 Yapay Zeka Nedir?
 - Ekonomi ve finansta yapay zeka
- 2 Yapay Zekanın Uygulanması
 - Makine öğrenmesi
 - İstatistiksel öğrenme
- 3 Örnek Uygulamalar
 - Ücret verileri
 - Borsa verileri
 - Gen anlatım verileri



Yapay Zeka Nedir?

Yapay zeka

Yapay zeka (artificial intelligence); insan gibi düşünebilen, davranan, bilinç sahibi olabilen insan yapımı makine ya da bilgisayarlara denir.

- Kökeni oldukça eskiye giden bu kavram bilgisayar biliminin yanı sıra popüler kültür, edebiyat, felsefe ve hatta antik çağ mitolojisinin de yer aldığı çok geniş bir alanı içine alır.



Yapay Zekanın Kısa Tarihi

- Mitolojide insan gibi davranan mekanik araçlara sıkça rastlanır.
- Yüzyıllar içinde akıl ya da farkındalığa sahip insan yapımı varlık ve robotlar birçok sanat ve edebiyat eserine konu olmuştur.
- Modern anlamda yapay zekanın fikir babasının ise İngiliz matematikçi Alan Turing (1912-1954) olduğu kabul edilmektedir.
- Turing'in 1950 yılında yayımlanan "Computing Machinery and Intelligence" başlıklı çalışması önemli bir dönüm noktası olmuştur.
- "Yapay zeka" terimi ilk kez 1956 yılında İngiltere Dartmouth Koleji'nde düzenlenen bir bilimsel toplantı sonrasında kullanılmıştır.
- 1960'larda yapay zekayı uygulamanın zor olduğu farkedilmiştir.
- 1980'lerde makine öğrenmesi algoritmaları geliştirilmiştir.
- 1990'larda Bayesçi istatistiksel yaklaşımlar önem kazanmıştır.
- 1995'te yapay zeka bir bilim dalı olarak görülmeye başlamıştır.
- Yapay zeka 1997'de satranç, 2011'de Riziko yarışması ve 2016'da da go oyununda insan şampiyonları yenmeyi başarmıştır.



Ekonomi ve Finans Alanlarında Yapay Zeka

Günümüzde yapay zekanın ekonomi ve finans alanında birçok uygulaması bulunmaktadır. Bunlara verilebilecek birkaç örnek şunlardır:

- Hisse senedi fiyatları ya da döviz kurları gibi finansal verilerin gelecekteki değerlerinin tahmin edilmesi
- Çok sayıda gösterge kullanılarak çeşitli makroekonomik değişkenlerin izleyeceği yönün kestirilmesi
- Karmaşık verilere dayanarak tarımsal üretim, otomobil satışları, sağlık harcamaları gibi ekonomik büyüklüklerin öngörülmesi
- Kredi ya da sigorta başvurularında farklı müşterilerin kredi notu ya da risk priminin belirlenmesi
- Demografik verilere bağlı olarak potansiyel müşterilerin bulunması ve bunların gruplandırılması
- Müşteriler için kişiselleştirilmiş ürün tavsiyeleri oluşturulması
- Çeşitli müşteri grupları için farklı reklam bütçeleri ve stratejileri uygulanması



Ders Planı

- 1 Yapay Zeka Nedir?
 - Ekonomi ve finansta yapay zeka
- 2 Yapay Zekanın Uygulanması
 - Makine öğrenmesi
 - İstatistiksel öğrenme
- 3 Örnek Uygulamalar
 - Ücret verileri
 - Borsa verileri
 - Gen anlatım verileri



Makine Öğrenmesi Nedir?

- Günümüzde yapay zekayı geliştirmeye yönelik çalışmalar bilgisayar biliminin alt dalı olan makine öğrenmesinin konusudur:

Makine öğrenmesi

Makine öğrenmesi (machine learning), yapay zekayı uygulamaya yönelik bilgisayar programları ve algoritmaları konu alan bilim dalıdır.

- Üç tür makine öğrenmesi vardır:
 - **Denetimli öğrenme** (supervised learning)
 - **Denetimsiz öğrenme** (unsupervised learning)
 - **Yarı-denetimli öğrenme** (semi-supervised learning)
- Denetimli öğrenmede çeşitli eğitim verileri temel alınarak belirli bir hedef değişken tahmin edilir ya da sınıflandırılır.
- Denetimsiz öğrenmede hedef değişken yoktur. Verilerin içindeki bilinmeyen örüntü ve benzerlikler bulunmaya çalışılır.
- Yarı-denetimli öğrenmede ise hedef değişken kısmen vardır.



İstatistik ve Veri Bilimi

- Görece yeni bir alan olan makine öğrenmesinin kullandığı araçların önemli bir bölümü istatistik biliminden gelmektedir:

İstatistik

İstatistik (statistics); verileri toplamayı, çözümlmeyi, yorumlamayı ve anlatmayı konu alan bilim dalıdır.

- Günümüzde özellikle büyük veri setlerini bilgisayar yardımıyla çözümlmeye yönelik disiplinler-arası bir bilim dalı olarak **veri bilimi** (data science) kavramı da sıkça kullanılmaktadır.
- Ancak birçok bilim insanına göre veri bilimi ile istatistik temelde aynı şeydir. Zaman içinde istatistiğin kullandığı araçların gelişmesiyle birlikte veri bilimi moda bir terim olarak ortaya çıkmıştır.



İstatistiksel Öğrenme

- Günümüzde çeşitli makine öğrenmesi algoritmaları istatistiksel öğrenme kuramını temel almaktadır:

İstatistiksel öğrenme

İstatistiksel öğrenme (statistical learning); veriler içindeki ilişki ve örüntüleri anlamaya yönelik geniş kuram, yöntem ve araçlar bütünüdür.

- Çeşitli istatistiksel araçları modern bilgisayarların yüksek kapasitesiyle birleştiren istatistiksel öğrenme, istatistik bilimi ile makine öğrenmesi arasında bir köprü görevi üstlenmektedir.
- İstatistiksel öğrenme, 1960'larda ortaya çıkmış bir terim olmakla birlikte kullandığı yöntemlerin birçoğu daha eskiye gitmektedir.



İstatistiksel Öğrenmenin Kısa Tarihi

- 1800'lerin başında Legendre ve Gauss **sıradan enküçük kareler** (ordinary least squares), kısaca **SEK** (OLS) yöntemini buldular.
- Nitel verilere ilişkin olarak ise 1936 yılında Fisher **doğrusal diskriminant** (linear discriminant) çözümlemesini geliştirdi.
- Yine, nitel veri alanında **lojistik regresyon** (logistic regression) yöntemi 1940'larda farklı araştırmacılar tarafından bulundu.
- 1972'de Nelder ve Wedderburn SEK ve lojistik regresyonu da içine alan geniş bir araç setini **genellemeli doğrusal modeller** (generalized linear models) olarak ortaya koydular.
- Bilgisayarların güçlenmesiyle birlikte 1980'lerde Breiman, Friedman, Olshen ve Stone birlikte **sınıflandırma** (classification) ve **regresyon ağaçları** (regression trees) araçlarını geliştirdiler.
- 1996'da Saffran vd. bebeklerin dil öğrenimindeki anahtar sürecin istatistiksel öğrenme olduğunu ortaya koydular.
- İstatistik ve bilgisayar bilimlerindeki gelişmelere bağlı olarak istatistiksel öğrenme 2000 sonrasında hızla gelişmeyi sürdürdü.



Makine Öğrenmesi ve İstatistiksel Öğrenme



- Görüldüğü gibi, dersimizin konusu yapay zeka olmakla birlikte biz, yapay zekayı uygulamak amacıyla istatistik ve makine öğrenmesini de içine alan **istatistiksel öğrenme** tekniklerini inceliyoruz.



İstatistiksel Öğrenme

“Hep söylüyorum, önümüzdeki 10 yılda cazip meslek istatistikçiler olacak diye. Ve şaka yapmıyorum!”

HAL VARIAN, Google Başekonomisti



Ekonometrinin Rolü

- Makine öğrenmesi algoritmaları özellikle büyük veri setleri içinde örüntü bulma ve kestirim yapma konularında son derece iyidir.
- İstatistik ise çıkarsama ve çeşitli önsav sınamalarında güçlüdür.
- Ancak bunların her ikisi de temelde bir kuram yattığı zaman daha da başarılı olabilmektedir.
- Örnek olarak, makine öğrenmesini kullanarak farklı şehirlerdeki polis sayısı ile suç oranı arasında güçlü bir örüntü görebiliriz.
- Ancak bu bize hangisinin diğerine neden olduğunu söyleyemez.
- Söz gelimi, polis sayısı artarsa suç oranına ne olacağını iyi tahmin edemez çünkü eldeki verileri oluşturan süreç bunun tersidir.
- Dolayısıyla burada büyük veri seti kullanmak da işe yaramaz.
- Ayrıca birçok veri setinde **özilinti**, (autocorrelation) **sansürlü gözlem**, (censored observation) **kariştirici değişken** (confounding variable) gibi özel durumlar bulunabilmektedir.
- Sonuç olarak, uygulamada alan bilgisi ve kuramsal modelleme önemlidir. Ekonomi ve ekonometri işte bu altyapıyı bize sunar.



Ders Planı

- 1 Yapay Zeka Nedir?
 - Ekonomi ve finansta yapay zeka
- 2 Yapay Zekanın Uygulanması
 - Makine öğrenmesi
 - İstatistiksel öğrenme
- 3 Örnek Uygulamalar
 - Ücret verileri
 - Borsa verileri
 - Gen anlatım verileri

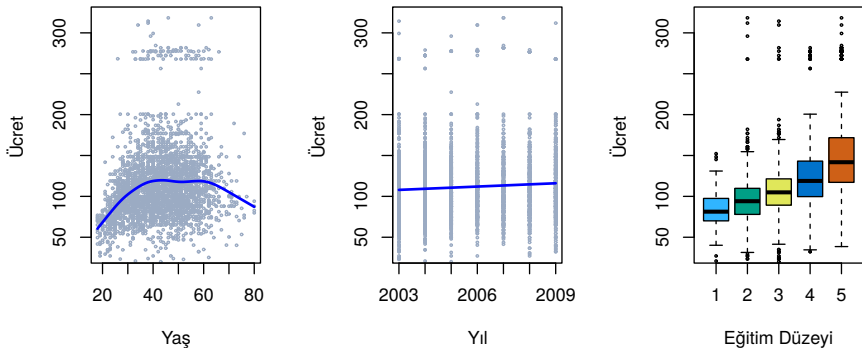


Örnek 1: Ücret Verileri (1)

- İstatistiksel öğrenme, **verileri anlamak** amacıyla kullanılan çok geniş bir yöntem ve araç setini içerir.
- Amacımız farklı durumlarda hangi araçları nasıl kullanacağımızı öğrenmek ve sonuçları doğru yorumlayabilmektir.
- Örnek olarak, önümüzdeki bölümlerde ayrıntılı olarak göreceğimiz birkaç farklı yöntemi burada kısaca tartışalım.
- İlk olarak, ABD'deki bir grup erkek çalışanın ücretlerini ele alalım.
- Çalışanların yaşı ve eğitim düzeyinin yanı sıra takvim yılının ücretler üzerindeki ilişkisini anlamaya çalışalım.
- Bu ilişkiler Şekil 1'de üç farklı görsel araç ile gösterilmiştir.



Örnek 1: Ücret Verileri (2)



Şekil 1: Bir grup çalışana ait ücretlerin yaş, yıl ve eğitim ile ilişkisi



Örnek 1: Ücret Verileri (3)

- Şekildeki sol panelde çalışanların yaşları ile kazandıkları ücretler arasındaki ilişki verilmiştir.
- Her gri nokta bir gözlemdir. Farklı çalışanlara ait yaş ve ücret değerlerini göstermektedir.
- Mavi renk çizgi ise yaş ve ücret arasındaki ortalama ilişkiyi gösteren **spline** (single parametric line) çizgisidir.
- Ortadaki panelde ücretlerin yıldan yıla nasıl değiştiği verilmiştir.
- Gri noktalar yine gözlemlerdir. Mavi çizgi ise kullanacağımız en temel araçlardan biri olan regresyon doğrusudur.
- Son olarak, sağdaki panelde ise ücretler ile eğitim düzeyi arasındaki ilişki gösterilmiştir.
- Buradaki renkli kutular da farklı eğitim düzeyleri için ortalama ücretleri ve bunların ne kadar değişken olduğunu anlatmaktadır.



Örnek 1: Ücret Verileri (4)

- Ücret verilerine ilişkin her üç şekil de önemli bilgiler sunmaktadır.
- Soldaki panelde mavi çizgiye bakarak 20 yaş sonrası ortalama ücretin önce arttığını, 35-60 yaşlar arasında yaklaşık sabit kaldığını ve 60 yaştan sonra düşmeye başladığını anlıyoruz.
- Ortadaki panelde, yıldan yıla ücretlerin yavaş da olsa giderek arttığını görüyoruz.
- Sağdaki panelde ise eğitim düzeyi arttıkça ortalama ücretin de arttığını ve yüksek eğitim düzeylerinde ücretlerin daha fazla değişkenlik gösterebildiğini öğreniyoruz.
- Sonuç olarak, bu şekil bize belli bir çalışanın ücretini açıklamak için tek bir faktöre bakmanın yeterli olmayacağını gösteriyor.
- Doğru tahminler için birçok değişkeni birlikte ele almak gerekiyor.

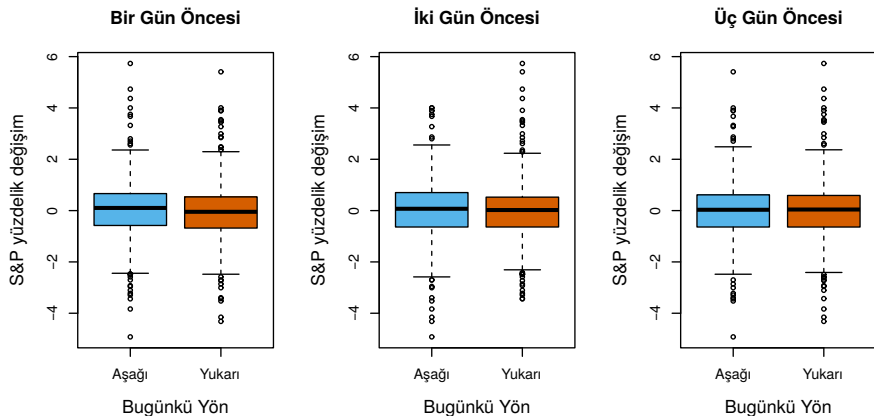


Örnek 2: Borsa Verileri (1)

- Ücret veri setinin konusu ücretleri tahmin etmektir. Ücret, sayısal bir değişkendir. Ancak bazı durumlarda sayısal olmayan, kategorik bir değişkenle de ilgilenebiliriz.
- Örnek olarak, şimdi Standard & Poor's 500 (S&P) endeksine yönelik borsa verilerini ele alalım.
- Burada amacımız önceki günlerde endeksin nasıl değiştiğine bakarak belli bir günde artış mı yoksa azalış mı olacağını kestirmek.
- Bunun için Şekil 2'yi inceleyelim.



Örnek 2: Borsa Verileri (2)



Şekil 2: Belli bir günde borsanın izlediği yönün önceki günlerle ilişkisi

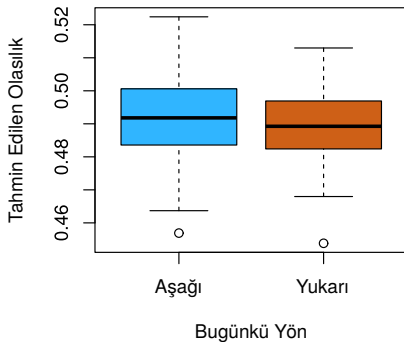


Örnek 2: Borsa Verileri (3)

- Şekilde yatay eksenlerde yalnızca “Aşağı” ve “Yukarı” şeklinde iki farklı kategori bulunmaktadır.
- Dolayısıyla burada istatistiksel öğrenme problemi sayısal bir tahmin değil, bir **sınıflandırma** (classification) yapmaya yöneliktir.
- Şekli incelediğimizde her üç panelde de “Aşağı” ve “Yukarı” kategorilerine ait çizimlerinin neredeyse aynı olduğunu görüyoruz.
- Buna göre, geçmişteki değişime bakarak borsanın bugünkü yönünü tahmin etmenin basit bir iş olmadığı anlaşılıyor.
- Yine de ileriki bölümlerde göreceğimiz bazı yöntemleri uygulayarak bu konuda da anlamlı sonuçlar elde edebileceğiz.
- Örnek olarak, Şekil 3’te verilerin bir alt kümesi için ikinci derece diskriminant çözümlenmesi sonuçları görülüyor.
- Bu sonuçlar, borsanın düştüğü günler için biraz daha yüksek düşme olasılığı veriyor. Buna dayanarak belli bir gün için endeksin yönünü yüzde 60 olasılıkla doğru tahmin edebiliyoruz.



Örnek 2: Borsa Verileri (4)



Şekil 3: Borsa verilerine yönelik ikinci derece diskriminant çözümlemesi

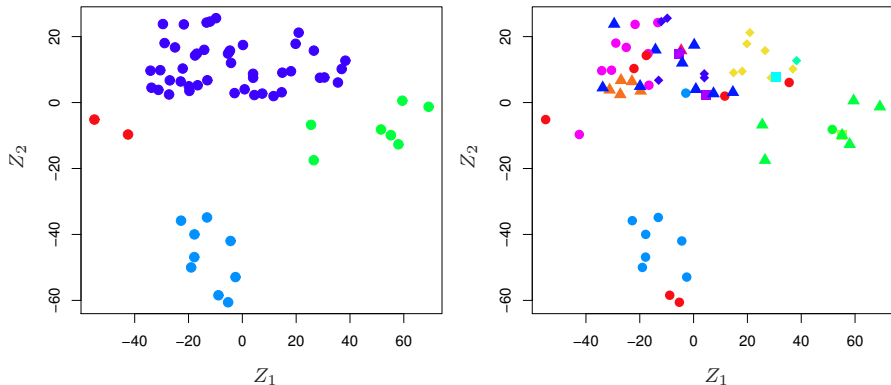


Örnek 3: Gen Anlatım Verileri (1)

- İstatistiksel öğrenmede kimi zaman da çeşitli girdi değişkenlerinin gözleendiği ancak bunlara karşılık gelen bir çıktı değişkeninin olmadığı durumlarla ilgilenebiliriz.
- Söz gelimi bir piyasada potansiyel müşterilere ait demografik verilere bakarak bunları farklı kümelere ayırmak isteyebiliriz.
- Bu yaklaşıma **sınıflandırma** (clustering) çözümlemesi adı verilir.
- Önceki örneklerden farklı olarak burada ücret yada fiyat gibi bir çıktı değişkenini tahmin etmeyiz.
- Bunun yerine, benzer özellikte müşteriler olup olmadığını anlamaya ve bunları ayırt etmeye çalışırız.
- Örnek olarak, Şekil 4'te 64 adet kanser hücresi dizisinin Z_1 ve Z_2 değerlerine göre serpilim çizimi görülmektedir.



Örnek 3: Gen Anlatım Verileri (2)



Şekil 4: Farklı kanser hücresi dizilerine ait Z_1 ve Z_2 ölçümleri



Örnek 3: Gen Anlatım Verileri (3)

- Burada Z_1 ve Z_2 değişkenleri 6.830 gözlemlik ana veri setinden elde edilen ilk iki **temel bileşen** (principal component) değerleridir.
- Verileri bu şekilde özetlemek bir miktar bilgi kaybına yol açmakla birlikte görsel bir inceleme yapabilmemizi sağlar.
- Küme sayısına karar vermek genelde zordur. Sol panelde verilen çözümlene sonuçlarına göre burada en az dört grup bulunuyor.
- Gerçekte ise bu veri seti toplam 14 farklı hücre dizisi içeriyor. Doğru sınıflandırma sağ panelde görülüyor.
- Her iki şekilde de farklı grupların bir miktar örtüşmekle birlikte belirgin kümeler oluşturduğunu anlayabiliyoruz.
- Bu sonuçlar yalnızca iki boyutlu bir çözümlene kullanarak bile gerçeğe yakın bir sınıflandırma yapabileceğimizi teyit ediyor.



Önümüzdeki Dersin Konusu ve Ödev

Ödev

Kitaptan **Bölüm 1** “Giriş” okunacak.

Önümüzdeki Ders

İstatistiksel Öğrenme: Temel Kavramlar

