

R İLE PROGRAMLAMA VE MAKİNE ÖĞRENMESİ

Yazar

Muhlis Özdemir

Editör

Yakup Çelikkbilek



R İLE PROGRAMLAMA VE MAKİNE ÖĞRENMESİ

Yazar: Muhlis Özdemir - Editör: Yakup Çelikkbilek

Yayın No. : 2615
Mühendislik-Teknik No.: 279
ISBN : 978-605-033-228-5
Basım Sayısı : 1. Basım, Şubat 2020

© Copyright 2020, NOBEL AKADEMİK YAYINCILIK EĞİTİM DANIŞMANLIK TİC. LTD. ŞTİ. SERTİFİKA NO.: 40340
Bu baskının bütün hakları Nobel Akademik Yayıncılık Eğitim Danışmanlık Tic. Ltd. Şti.ne aittir. Yayınevinin yazılı izni olmaksızın, kitabın tümünün veya bir kısmının elektronik, mekanik ya da fotokopi yoluyla basımı, yayımı, çoğaltımı ve dağıtımı yapılamaz.

Nobel Yayın Grubu, 1984 yılından itibaren ulusal ve 2011 yılından itibaren ise uluslararası düzeyde düzenli olarak faaliyet yürütmekte ve yayınladığı kitaplar, ulusal ve uluslararası düzeydeki yükseköğretim kurumları kataloglarında yer almaktadır.

Genel Yayın Yönetmeni: Nevzat Argun -nargun@nobelyayin.com-
Yayın Koordinatörü : Gülfem Dursun -gulferm@nobelyayin.com-

Redaksiyon : Ahsen Aktitiz -ahsen@nobelyayin.com-
Sayfa Tasarım : Ahmet S. Baydar -ahmet@nobelyayin.com-
Kapak Tasarım : Mehtap Yürümez -mehtap@nobelyayin.com-
Baskı Sorumlusu : Yavuz Şahin -yavuz@nobelyayin.com-
Baskı ve Cilt : Atalay Matbaacılık / Sertifika No.: 15689-
Büyük Sanayi 1 Cad. Elif Sok. No.:7/236-237 İskitler / ANKARA

Kütüphane Bilgi Kartı

Özdemir, Muhlis., Çelikkbilek, Yakup.
R İLE PROGRAMLAMA VE MAKİNE ÖĞRENMESİ / Yazar: Muhlis Özdemir - Editör: Yakup Çelikkbilek
1. Basım. x + 304 s. 19,5x27,5 cm. Kaynakça var, dizin yok.
ISBN: 978-605-033-228-5
1. Kodlama 2. R Programlama 3. Makine Öğrenmesi

Genel Dağıtım

ATLAS AKADEMİK BASIM YAYIN DAĞITIM TİC. LTD. ŞTİ.
Adres: Bahçekapı mh. 2465 sk. Oto Sanayi Sitesi No:7 Bodrum Kat Şaşmaz-ANKARA - siparis@nobelyayin.com-
Telefon: +90 312 278 50 77 - Faks: 0 312 278 21 65
E-Satış: www.nobelkitap.com - www.atlaskitap.com - Bilgi: esatis@nobelkitap.com - info@atlaskitap.com

Dağıtım ve Satış Noktaları: Alfa Basım Dağıtım, Arasta, Arkadaş Kitabevi, D&R Mağazaları, Dost Dağıtım, Ekip Dağıtım, Kida Dağıtım, Kitapsan, Nezh Kitabeveleri, Pandora, Prefix, Remzi Kitabeveleri

ÖNSÖZ

Bu kitap iki bölümden oluşmaktadır. Birinci bölümde, R Programlama Dili'nin temelleri anlatılmaktadır. Daha önce hiç R Programlama Dili'ni bilmeyen okuyucu, ilk bölümü tamamlaması durumunda programlama dili hakkında yeterli düzeyde bilgi sahibi olabilecektir. Daha önce R Programlama Dili'ni kullanan okuyucunun ise bilgi düzeyini artırması muhtemeldir. Ağır istatistiksel ve de matematiksel hesaplamaların veya teoremlerin yer almadığı bu kitap, ilgili analizi gerçekleştirmek için gerekli olan açıklamaları okuyucuya sunmaktadır. Makine öğrenmesi başlığı altında yer alan yöntemler, örnekler ile desteklenerek anlatılmaktadır. Makine öğrenmesi başlığı dört ana bölümden oluşmaktadır. Bu bölümlerden ilkinde, makine öğrenmesine başlamadan önce bilinmesi gereken konulara değinilmiştir. Test ve eğitim setleri oluşturmaya, makine öğrenmesi için model kurarken kullanılacak formüllerin oluşturulmasına ve model başarısının ölçülmesinde kullanılacak temel fonksiyonlara değinilmiştir. Makine öğrenmesi başlığı altında yer alan diğer başlıklar; regresyon ve korelasyon analizi, denetimli öğrenme ve denetimsiz öğrenme yöntemleridir. Her bir konunun sonunda, ilgili konu için yazılan kodlar bir bütün hâlinde paylaşılmaya çalışılmıştır. Teorik bilgiye sahip olup doğrudan konunun uygulanmasını öğrenmek isteyen okuyucular, konu sonlarında bulunan kod blokları ile çalışabilirler.

Çok şey öğrendiğim ve bu alana yönelmemde son derece etkili olan değerli hocam Prof. Dr. Tunçhan Cura'ya teşekkürler.

Kitabın editörlüğünü üstlenen meslektaşım Dr. Yakup Çelikbilek'e teşekkürler.

Uzunca zaman alan kitap yazma sürecinde bana elinden gelen desteği sunan eşim Behice Özdemir'e teşekkürler.

Okuyucusuna faydalı olması dileğiyle...

İÇİNDEKİLER

1. BÖLÜM: GENEL BİLGİLER.....	1
1.1. Arayüz.....	1
1.2. Atama Operatörü	8
1.3. Karşılaştırma Operatörleri.....	9
1.3.1. Küçüktür Operatörü.....	9
1.3.2. Büyüktür Operatörü.....	10
1.3.3. Küçük Eşittir Operatörü	10
1.3.4. Büyük Eşittir Operatörü	10
1.3.5. Ve Operatörü	11
1.3.6. Veya Operatörü	12
1.3.7. Eşit mi Operatörü.....	12
1.3.8. Değil Eşit Operatörü.....	13
1.3.9. Değil Operatörü	13
1.4. Temel Değer Tipleri ve Veri Yapıları.....	13
1.4.1. Temel Değer Tipleri	13
1.4.1.1. ls() ve rm() fonksiyonları	15
1.4.2. Temel Veri Yapıları.....	16
1.4.2.1. Vektörler	17
1.4.2.1.1. rep() Fonksiyonu	18
1.4.2.1.2. Bir Vektörün İndislerinin Adlandırılması	19
1.4.2.1.3. Vektörlerde Aritmetik İşlemler	19
1.4.2.1.4. Vektörlerde İndislere Erişim.....	23
1.4.2.1.5. which.max() ve which.min() Fonksiyonları.....	24
1.4.2.1.6. order() Fonksiyonu	25
1.4.2.1.7. %in% Operatörü	26
1.4.2.2. Matrisler.....	27
1.4.2.2.1. set.seed() Fonksiyonu	30
1.4.2.2.2. Matrisin Satır ve Sütunlarının İsimlendirilmesi	31
1.4.2.2.3. Matrisin Satır ve Sütun Sayısı	34
1.4.2.2.4. Matris İndislerine Erişim	34
1.4.2.2.5. Matrislerde Karşılaştırma Operatörlerinin Kullanılması.....	36
1.4.2.3. Faktörler.....	36
1.4.2.3.1. Faktör İndislerine Erişim.....	38
1.4.2.3.2. Faktörlerde Dikkat Edilmesi Gereken Hususlar	39
1.4.2.4. NA ve NaN Değerleri	41
1.4.2.5. Data.Frame'ler	42
1.4.2.5.1. head() ve tail() Fonksiyonları.....	42
1.4.2.5.2. Data.frame'lerde Sütun İsimlendirme.....	44
1.4.2.5.3. Data.frame'lerde İndislere Erişim	45
1.4.2.5.4. Data.frame'e Yeni Bir Sütun Ekleme	46
1.4.2.5.5. subset() Fonksiyonu	46
1.4.2.6. Listeler.....	48

1.4.2.6.1. Liste Özelliklerinin İsimlendirilmesi.....	51
1.4.2.6.2. Listelerde İndislere Erişim	52
1.4.2.6.3. Listeye Yeni Özellik Ekleme	53
1.5. Kolon Operatörü	54
1.6. Kontrol İfadeleri.....	56
1.6.1. if Kontrol İfadesi.....	56
1.6.2. else Kontrol İfadesi	56
1.6.3. else if Kontrol İfadesi	57
1.6.4. İç içe Kontrol İfadeleri.....	57
1.7. Döngüler	58
1.7.1. for Döngüsü	58
1.7.1.1. Vektörlerde for Döngüsü Kullanımı	59
1.7.1.2. Listelerde for Döngüsü Kullanımı	61
1.7.1.3. Matrislerde for Döngüsü Kullanımı	63
1.7.1.4. Data.frame'lerde for Döngüsü Kullanımı	65
1.7.1.5. ifelse() Fonksiyonu	66
1.7.2. while Döngüsü	68
1.7.3. Döngülerde break ve next Kullanımı	70
1.8. Fonksiyonlar.....	72
1.8.1. Bir Fonksiyonun Kapsamı.....	86
1.8.2. mean(), median(), sd() ve rnorm() Fonksiyonları.....	87
1.8.3. Fonksiyon İçerisinde Fonksiyon Kullanımı	91
1.8.4. Ön Tanımlı Parametre Değerlerine Sahip Fonksiyonlar.....	93
1.8.5. Fonksiyon Dökümantasyonu	95
1.9. Apply Fonksiyonları.....	95
1.9.1. apply() Fonksiyonu.....	95
1.9.2. lapply() Fonksiyonu.....	100
1.9.2.1. Köşeli Parantez Kullanımı.....	109
1.9.2.2. Kendi Fonksiyonlarımızın lapply ile Kullanımı	110
1.9.3. sapply() Fonksiyonu	113
1.9.3.1. Kendi Fonksiyonlarımızın sapply() ile Kullanımı	115
1.9.4. İç içe apply() Fonksiyonlarının Kullanımı.....	118
1.9.5. which.min() ve which.max() Fonksiyonlarının Apply Ailesinde Kullanımı	120
1.10. R Programlama Dili'nde Tarih ve Saat.....	121
1.10.1. Kısaltmalar	124
1.10.2. Tarih Saat ile Çalışırken Paket Kullanımı	127
1.10.3. Tarih Saat Yuvarlama	129
1.10.4. Tarih Saat Objeleri ile Aritmetik İşlemler.....	130
1.10.5. Tarih Saat Dizileri Oluşturma	134
1.11. Dosya İşlemleri.....	136
1.11.1. Dosyadan Veri Okuma	136
1.11.1.1. read.table() Fonksiyonu	137
1.11.2. Eksik Verileri Tamamlama	139
1.11.3. Dosyaya Veri Yazma.....	145
1.12. dplyr Paketi ile Veri Süzme ve Manipülasyonu	147

1.12.1. Pipe Operatörü	148
1.12.2. select() Fonksiyonu	149
1.12.3. filter() Fonksiyonu	153
1.12.4. arrange() Fonksiyonu	156
1.12.5. mutate() Fonksiyonu	157
1.12.6. summarize() Fonksiyonu	158
2. BÖLÜM: MAKİNE ÖĞRENMESİ	161
2.1. <i>Başlamadan Önce</i>	162
2.1.1. Test ve Eğitim Seti Oluşturma	162
2.1.2. Formül Oluşturma	166
2.1.3. Model Başarısı	168
2.2. <i>Regresyon ve Korelasyon Analizi</i>	170
2.2.1. Korelasyon Analizi	170
2.2.1.1. Ayrıntılı Korelasyon Hesabı	173
2.2.2. Basit Doğrusal Regresyon Analizi	175
2.2.3. Çok Değişkenli Doğrusal Regresyon Analizi	182
2.2.4. Polinom Doğrusal Regresyon Analizi	187
2.2.5. Doğrusal Olmayan Regresyon Analizi	197
2.2.6. Destek Vektör Makinesi Regresyon Analizi	207
2.2.7. Regresyon Ağaçları	210
2.2.7.1. Ağaç Regresyonu	211
2.2.7.2. Rastsal Orman Regresyonu	216
2.3. <i>Denetimli Öğrenmede Sınıflandırma Yöntemleri</i>	220
2.3.1. K - En Yakın Komşu Algoritması	220
2.3.2. Bayes Sınıflandırma	227
2.3.3. Lojistik Regresyon Analizi	233
2.3.4. Destek Vektör Makineleri	242
2.3.5. Sınıflandırma Ağaçları	251
2.3.5.1. Ağaç Sınıflandırması	251
2.3.5.2. Rastsal Orman Sınıflandırması	258
2.4. <i>Denetimsiz Öğrenmede Kümeleme Yöntemleri</i>	264
2.4.1. K - Ortalamalar Kümeleme Analizi	265
2.4.1.1. Küme Sayısının Belirlenmesi	274
2.4.1.1.1. Elbow Yöntemi	275
2.4.1.1.2. Silhouette Yöntemi	276
2.4.1.1.3. Gap Statistic Metodu	277
2.4.2. Hiyerarşik Kümeleme Analizi	280
2.4.2.1. Küme Sayısının Belirlenmesi	297
2.4.2.1.1. Elbow Yöntemi	297
2.4.2.1.2. Silhouette Yöntemi	299
2.4.2.1.3. Gap Statistic Metodu	300
KAYNAKÇA	303

ŞEKİLLER VE TABLOLAR LİSTESİ

ŞEKİLLER

Şekil 1 Macintosh işletim sistemine sahip bilgisayarda kabuk R görseli	2
Şekil 2 RStudio geliştirme ortamına ait görsel 2.....	3
Şekil 3 RStudio geliştirme ortamına ait görsel 3.....	3
Şekil 4 Script dosyası oluşturmak için izlenmesi gereken yol.....	4
Şekil 5 Script dosyasına genel bir bakış	4
Şekil 6 İlk uygulama	5
Şekil 7 Ayarlar penceresi	7
Şekil 8 Görünüm ayarları	7
Şekil 9 Paket yükleme.....	8
Şekil 10 Veri yapıları ve tiplerine genel bir bakış.....	16
Şekil 11 Ayrıntılı korelasyon hesabı görseli	175
Şekil 12 Dağılım grafiği	177
Şekil 13 Regresyon doğrusu	181
Şekil 14 Ürün ve fiyat görseli	188
Şekil 15 Doğrusal regresyon denklem doğrusu	190
Şekil 16 Polinom regresyon eğrisine ait görsel.....	192
Şekil 17 Basit ve polinom doğrusal regresyon karşılaştırması görseli	193
Şekil 18 Doğrusal ve doğrusal olmayan model karşılaştırması.....	200
Şekil 19 Destek vektör makinesi grafiği	209
Şekil 20 Ağaç regresyonu grafiği.....	213
Şekil 21 Ağaç regresyonu grafiği 2.....	213
Şekil 22 Karar ağacı regresyon görseli	215
Şekil 23 Rastsal orman regresyon görseli.....	218
Şekil 24 K - en yakın komşu algoritması görseli.....	223
Şekil 25 Yeni değer tahmin görseli	224
Şekil 26 Lojistik regresyon dağılım görseli.....	235
Şekil 27 Destek vektörlerin belirlenmesi görseli	243
Şekil 28 Destek vektör makinesi düzleminin elde edilmesi.....	244
Şekil 29 Destek vektör makinesi sınıflandırma görseli	246
Şekil 30 Destek vektör makinesi sigmoid fonksiyon görseli	249
Şekil 31 İşletme tipi ağaç sınıflandırması görseli	253
Şekil 32 İhracatın tahmin edilmesinde ağaç sınıflandırması görseli.....	256
Şekil 33 Kümeleme görseli	265
Şekil 34 Koordinatlara ait görsel.....	267
Şekil 35 Küme merkezleri eklenmiş görsel	269
Şekil 36 Birinci bölge görseli	272
Şekil 37 İkinci bölge görseli.....	273

Şekil 38 Üçüncü bölge görseli.....	274
Şekil 39 K ortalamalar kümeleme analizinde elbow yöntemi ile küme sayısının belirlenmesi	275
Şekil 40 K ortalamalar kümeleme analizinde silhouette yöntemi ile küme sayısının belirlenmesi.....	276
Şekil 41 K ortalamalar kümeleme analizinde gap statistic yöntemi ile küme sayısının belirlenmesi.....	277
Şekil 42 hclust() fonksiyonunun farklı yöntemler ile kullanımı	283
Şekil 43 Ward'ın minimum varyans yöntemi ile kümeleme görseli	286
Şekil 44 Hiyerarşik kümeleme analizinde pltree() fonksiyonu ile görsel elde edilmesi.....	291
Şekil 45 Hiyerarşik kümeleme analizinde diana() fonksiyonu ile kümeleme görselinin elde edilmesi	294
Şekil 46 Hiyerarşik kümeleme analizi nihai kümeleme görseli.....	295
Şekil 47 Hiyerarşik kümeleme analizinde elbow yöntemi ile küme sayısının belirlenmesi.....	298
Şekil 48 Hiyerarşik kümeleme analizinde silhouette yöntemi ile küme sayısının belirlenmesi	299
Şekil 49 Hiyerarşik kümeleme analizinde gap statistic yöntemi ile küme sayısının belirlenmesi	300

TABLULAR

Tablo 1 Matris ve data.frame'lerde satır-sütun isimlendirme	44
Tablo 2 İlişki grafikleri.....	171
Tablo 3 Hiyerarşik kümeleme analizinde kullanılan kümeleme yöntemleri.....	281