

Akademik Ağ Analizi

...ve az biraz R ile “Yapay Öğrenme”

Soner Aydın
İlker Birbil

Sabancı Üniversitesi

Bölüm Bilim

İçerik

Yapay Öğrenmenin Matematiksel Temelleri

20-23 Nisan 2017

Nesin Matematik Köyü, Şirince

(Yaz Okulu Sayfası)

Yapay Öğrenme

R Yapay Öğrenme Araçları

MAG: Akademik Ağ Veri Seti

İlk İzlenimler

Eş Yazarlık Ağları

Bağlantı Tahmini

Sonuç ve Açık Sorular



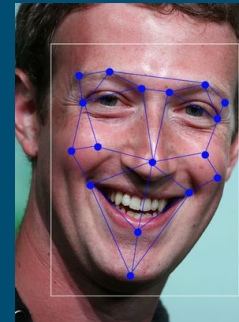
(Kaynak)

Yapay Öğrenme

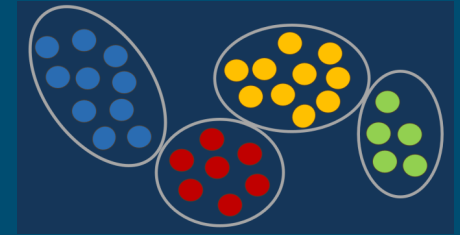
- Yapay Zekâ konusunun altında yer alan bir çalışma alanı
- Bilgisayarların kendi başlarına öğrenmeleri ve daha *akıllıca* karar vermeleri
- İki tip yaklaşım:
 - Güdümlü Öğrenme (supervised learning)
 - Güdümsüz Öğrenme (unsupervised learning)



(Kaynak)



(Kaynak)



(Kaynak)

Güdümlü Öğrenme

Etiketli veri
Bağımsız değişkenler
Bağımlı değişken tahmini
Regresyon
Sınıflandırma

Örnekler: Yüz tanıma, müşteri terk, kanserli hücre tespiti

Güdümsüz Öğrenme

Etiketsiz veri
Verinin içsel yapısı
Bağımlı değişken tahmini
Veri indirgeme
Ayrıştırma

Örnekler: Kümeleme, değişken azaltma, tavsiye sistemi

Eğitim-Test-Doğrulama

- Eldeki veri iki ya da üç parçaya ayrılır:
 1. **Eğitim**: İlk parça ile model eğitilir ve parametreler belirlenir
 2. **Test**: İkinci parça ile model test edilir ve parametreler güncellenir
 3. **Doğrulama**: Üçüncü parça ile modelin başarımı ölçülür
- R programlama dilinde verinin ayrılmasıyla birlikte modelin eğitilmesi ve test edilmesi birkaç satır kod ile yapılabilir

R Yapay Öğrenme Araçları

- Regresyon: glmnet, base
- Sınıflandırma: e1071, caret
- Kümeleme: mclust, fastcluster
- Ağ Analizi: sna, igraph
- Genel: H2O

R Yapay Öğrenme Örneği

```
library(caret)
library(ellipse)
library(e1071)
library(kernlab)
library(MASS)
library(randomForest)

data(iris)
dataset <- iris

# %80'i eğitim, %20'si doğrulama
validation_index <- createDataPartition(dataset$Species, p=0.80, list=FALSE)
validation <- dataset[-validation_index,]
dataset <- dataset[validation_index,]
head(dataset)

x <- dataset[,1:4] # Bağımlı değişkenler (girdi)
y <- dataset[,5] # Bağımsız değişkenler (çıktı)

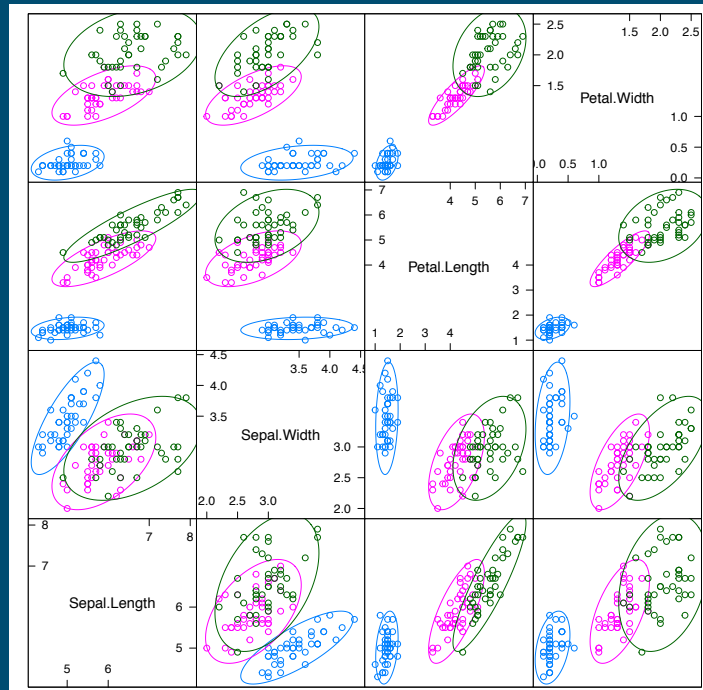
# Özniteliklerin gösterimi
featurePlot(x=x, y=y, plot="ellipse")

# Algoritmaları 10'ar kez çalıştır
control <- trainControl(method="cv", number=10)
metric <- "Accuracy"

# Metod 1: LDA
set.seed(7)
fit.lda <- train(Species~., data=dataset, method="lda",
                 metric=metric, trControl=control)

# Metod 2: kNN
set.seed(7)
fit.knn <- train(Species~., data=dataset, method="knn",
                 metric=metric, trControl=control)

# Metod 3: SVM
set.seed(7)
fit.svm <- train(Species~., data=dataset, method="svmRadial",
```

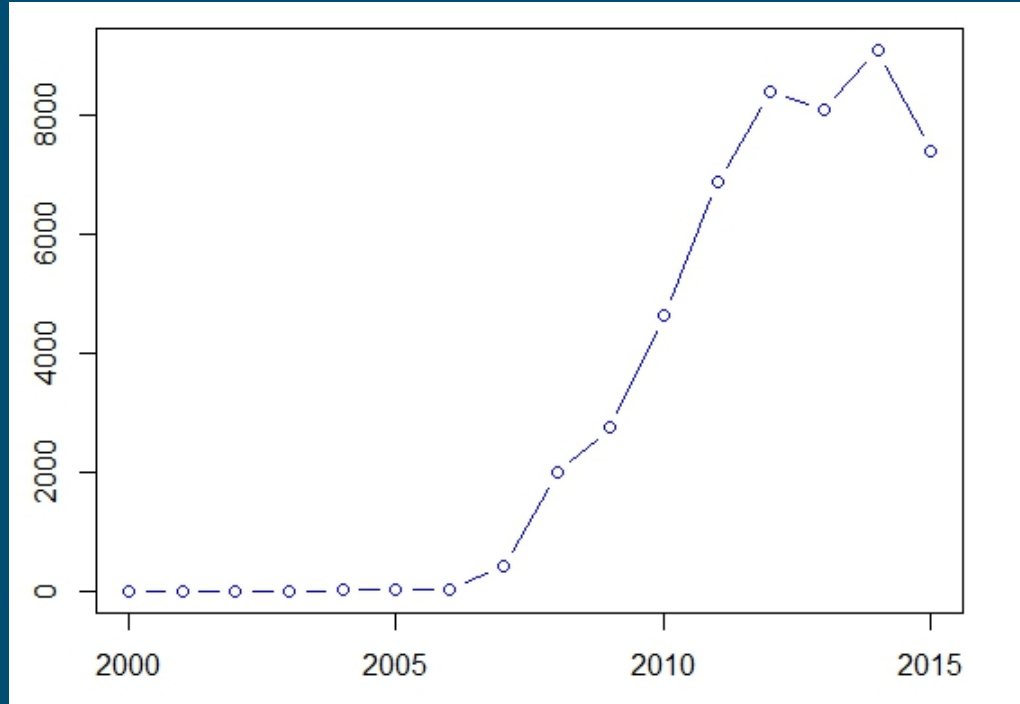


Akademik Ağ Analizi

- Veri: [Microsoft Academic Graph](#)
- Tüm verinin büyüklüğü ~100 GB
- Türkiye alt kümesi ~250 MB
- Toplam 12 CSV formatında dosya
- İlişkisel veri tabanı yapısı

Affiliations					
Affiliation ID	Affiliation name				
Authors					
Author ID	Author name				
ConferenceSeries					
Conference series ID	Short name (abbreviation)	Full name			
Conferences					
Conference series ID	Conference instance ID	Short name (abbreviation)	Full name	Location	Official conference URL
FieldsOfStudy					
Field of study ID	Field of study name				
Journals					
Journal ID	Journal name				
Papers					
Paper ID	Original paper title	Normalized paper title	Paper publish year	Paper publish date	Paper Document Object Identifier (DOI)
PaperAuthorAffiliations					
Paper ID	Author ID	Affiliation ID	Original affiliation name	Normalized affiliation name	Author sequence number
PaperKeywords					
Paper ID	Keyword name	Field of study ID mapped to keyword			
PaperReferences					
Paper ID	Paper reference ID				
PaperUrls					
Paper ID	URL				

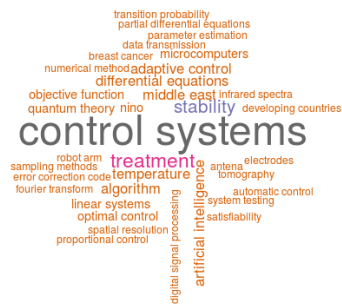
Veri Kurcalama



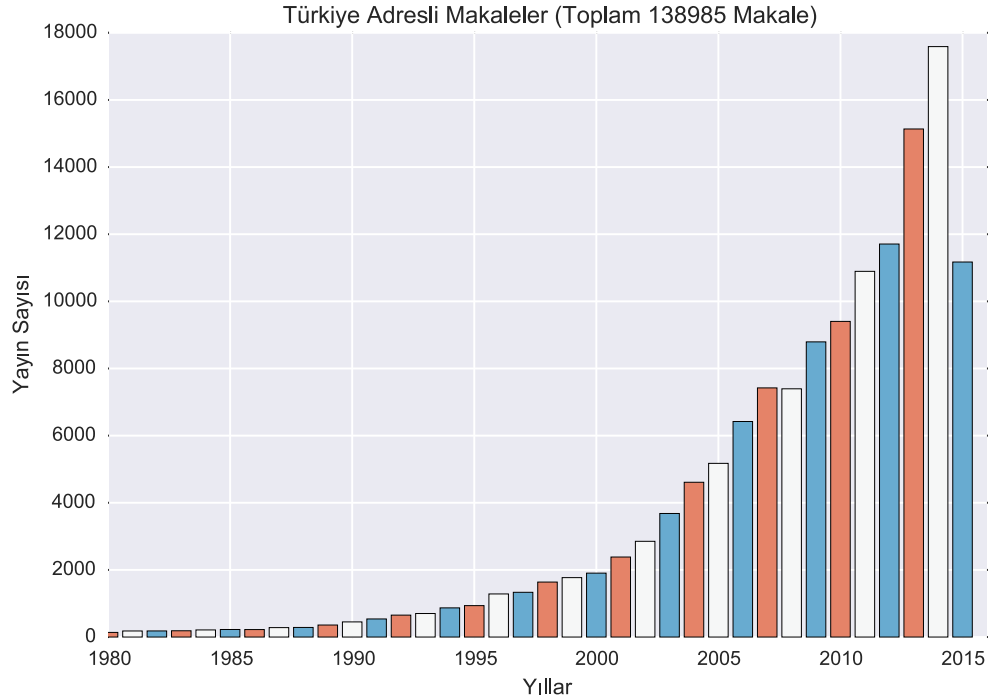
Yıllara Göre Konferans Sayıları

Veri Kurcalama

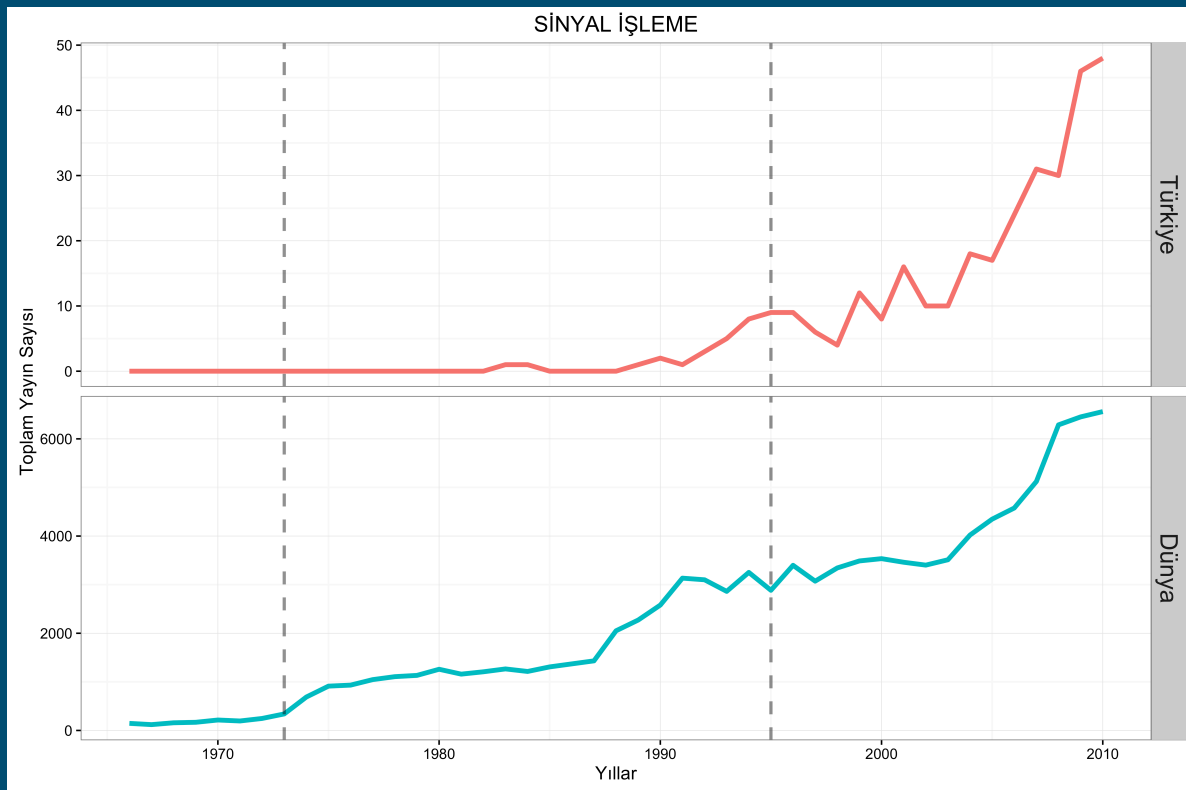
1990



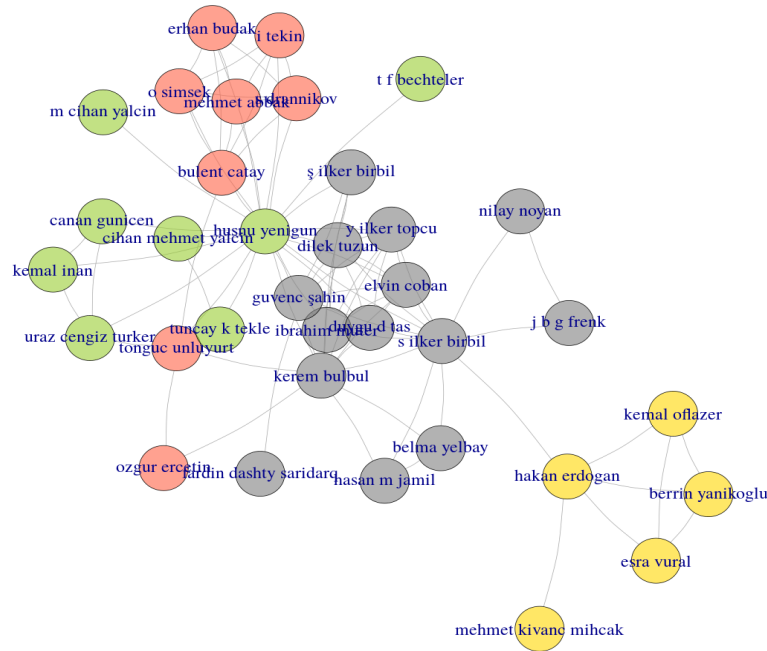
Veri Kurcalama



Veri Kurcalama



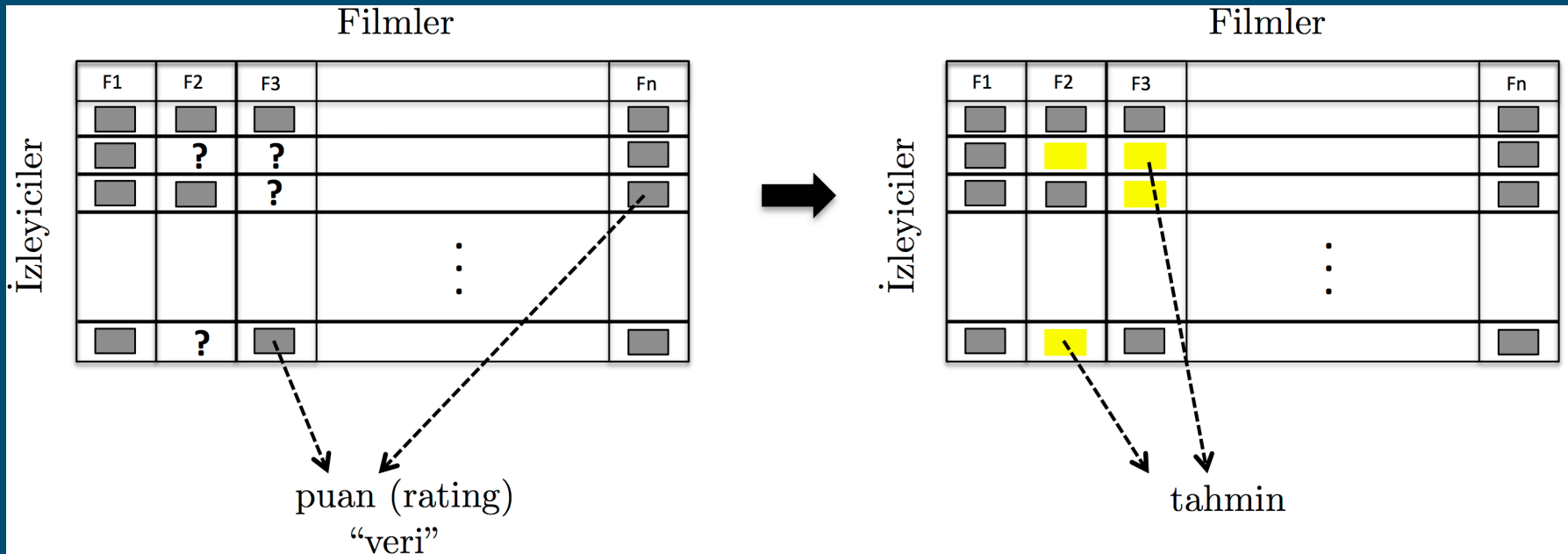
Veri Kurcalama



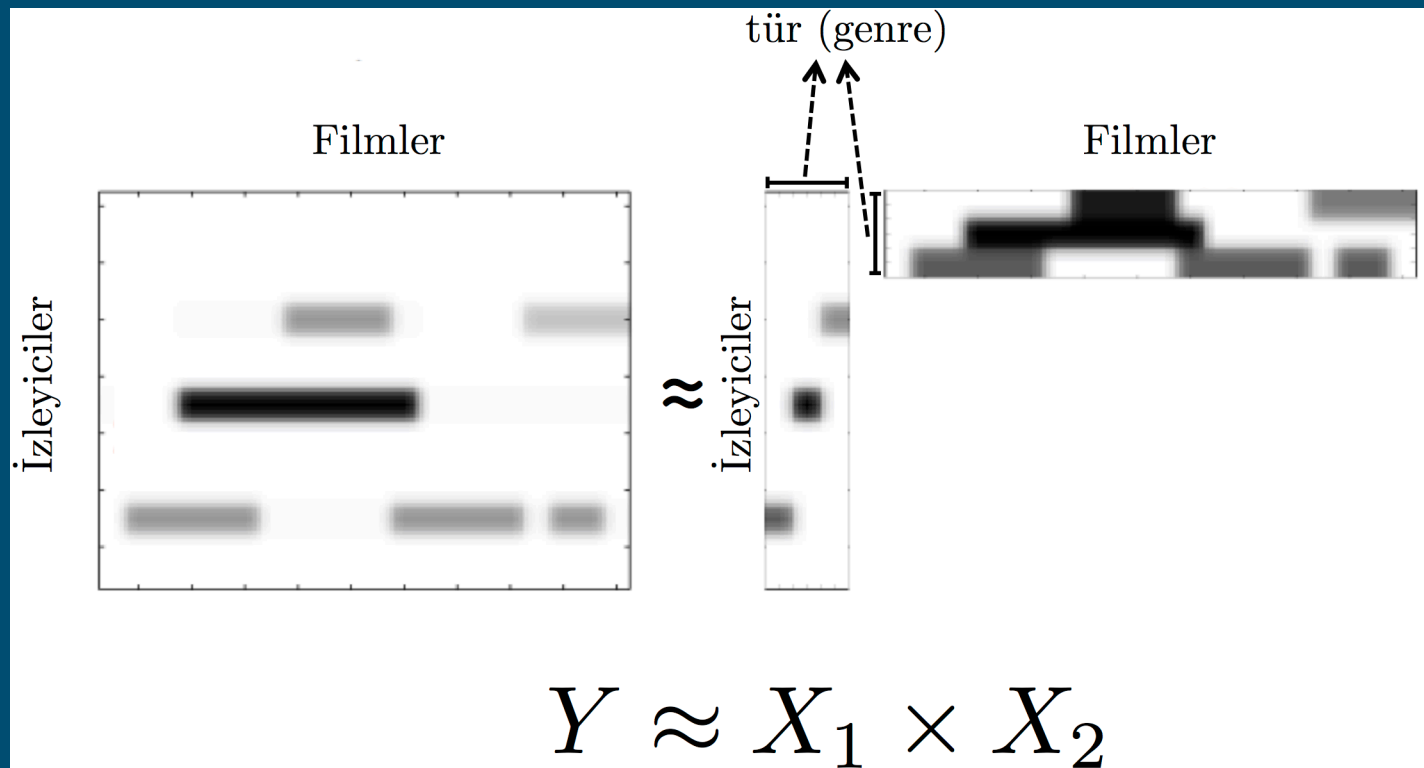
Eş Yazar Tavsiyesi

Yazarların makalelerine ve birlikte çalıştıkları insan sayılarına bakarak, eş yazar tavsiyesi yapılabilir mi?

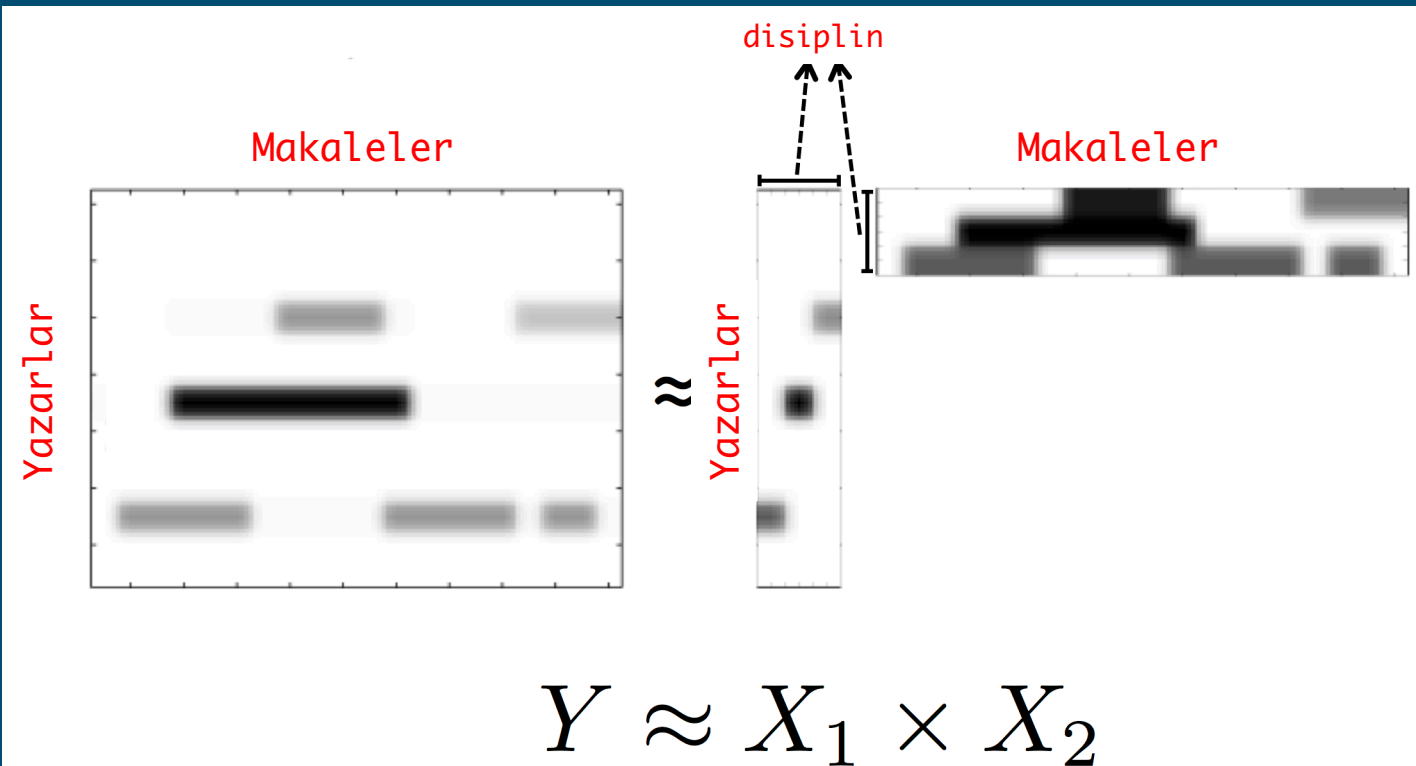
Eş Yazar Tavsiyesi



Eş Yazar Tavsiyesi



Eş Yazar Tavsiyesi



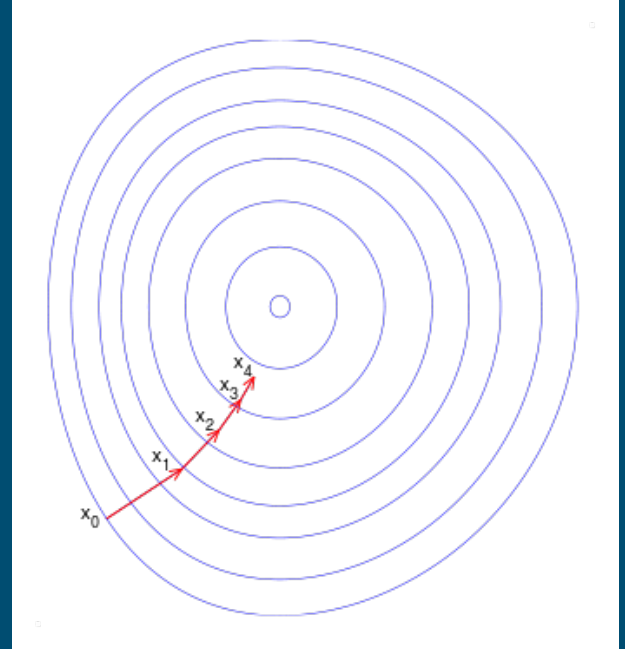
R

- Kütüphaneler

- `data.table`: Hızlı tablo işlemleri
- `Matrix`: Seyrek matris işlemleri

- Matris Ayrıştırma

- Optimizasyon: Rassal bayır inişi algoritması
- Ev yapımı



(Kaynak)

Kullanım Örneği

Verilen Yazar	Araştırma Konuları
Ş. İlker Birbil	Hesaplamalı Yöneylem Araştırması, Eniyileme Algoritmaları, Yapay Öğrenme

Benzerlik Ölçütü	Araştırma Konuları
Mehmet Selman Tamer	Fiber Optik, Sinyal İşleme, Malzeme Bilimi

Benzerlik & Bütünleyicilik Ölçütü	Araştırma Konuları
Ceki Franko	Risk Yönetimi, Rassal Eniyileme, Finansal Ekonomi

Fakat...

Benzerlik Ölçütü	Araştırma Konuları
Kubilay Işık	Ağız, Çene, Diş Sağlığı

Sonuç ve Açık Sorular

- **Yapay Öğrenme**
 - Verinin özniteliklerine bakarak kestirim yapma yaklaşımı
 - R için hazırlanmış pek çok yapay öğrenme algoritması mevcut
- **Akademik Ağ Analizi**
 - Cevaplardan çok sorularımız var
 - Eş yazar tavsiyesinin başarımı artırılabilir mi?
 - Ağ çeteleri bulunabilir mi?
 - Dünya-Türkiye karşılaştırmaları artırılabilir mi?