



HACETTEPE
ÜNİVERSİTESİ

Robust Correlation Scaled Principal Component Regression

Araştırma Yöntemleri Final Ödevi

Ramazan Erduran (21821809)

May 27, 2023

**HACETTEPE
ÜNİVERSİTESİ**





HACETTEPE
ÜNİVERSİTESİ

Robust Correlation Scaled Principal Component Regression

Bu makalede regresyon tahminlerinde beta parametrelerini etkileyen aykırı değerler, uç gözlemler ve çoklu doğrusal bağlantının giderilmesine ilişkin üç yöntem üzerinde durulmaktadır.



HACETTEPE
ÜNİVERSİTESİ

İçerik Şeması

1 Makalenin Yayınlandığı Dergi Hakkında

► Makalenin Yayınlandığı Dergi Hakkında

► Makale Hakkında

► Sonuç ve Tartışma

► Görüş ve Öneriler

► Referanslar



HACETTEPE
ÜNİVERSİTESİ

Dergi Hakkında

1 Makalenin Yayınlandığı Dergi Hakkında

- **Derginin Adı**
Hacettepe Journal of Mathematics and Statistics



HACETTEPE
ÜNİVERSİTESİ

Dergi Hakkında

1 Makalenin Yayınlandığı Dergi Hakkında

- **Derginin Adı**
Hacettepe Journal of Mathematics and Statistics
- **Derginin Indexleri**
Science Citation Index - Expanded (SCI-E)
TrDizin
Scopus
zbMATH
American Mathematical Society



HACETTEPE
ÜNİVERSİTESİ

Dergiye Erişim Bağlantısı

1 Makalenin Yayınlandığı Dergi Hakkında

- Web of Science Bağlantısı
- DergiPark Bağlantısı

**HACETTEPE
ÜNİVERSİTESİ**





HACETTEPE
ÜNİVERSİTESİ

Dergiye Genel Bakış

1 Makalenin Yayınlandığı Dergi Hakkında

Hacettepe Matematik ve İstatistik Dergisi, orijinal araştırma makalelerinin ve ara sıra uygun bir şekilde yüksek standartta anketlerin yayınlanmasına adanmış, uluslararası hakemli, açık erişimli bir dergidir.

Tüm yönleriyle İngilizce olarak yazılmıştır. Matematik ve İstatistik. Fizik, Aktüeryal Bilimler, Finans ve Ekonomi uygulamaları da dahil olmak üzere, Matematik ve İstatistik arasındaki arayüz hakkındaki makaleler özellikle kabul edilir.



HACETTEPE
ÜNİVERSİTESİ

İçerik Şeması

2 Makale Hakkında

► Makalenin Yayınlandığı Dergi Hakkında

► Makale Hakkında

► Sonuç ve Tartışma

► Görüş ve Öneriler

► Referanslar



HACETTEPE
ÜNİVERSİTESİ

Makale Hakkında Genel Bilgiler

Makale İncelemesi

- **Makalenin Adı**

Robust correlation scaled principal component regression



HACETTEPE
ÜNİVERSİTESİ

Makale Hakkında Genel Bilgiler

Makale İncelemesi

- **Makalenin Adı**

Robust correlation scaled principal component regression

- **Yazarlar**

Aiman Tahir

Maryam Ilyas



HACETTEPE
ÜNİVERSİTESİ

Makale Hakkında Genel Bilgiler

Makale İncelemesi

- **Makalenin Adı**

Robust correlation scaled principal component regression

- **Yazarlar**

Aiman Tahir

Maryam Ilyas

- **Makalenin Dili**

İngilizce



HACETTEPE
ÜNİVERSİTESİ

Makale Hakkında

Makale İncelemesi

- **Makalenin Konusu**

Çoklu regresyonda etkili gözlemler ve çoklu bağlantı sorunu ile başa çıkmak için Sağlam Korelasyon Tabanlı Regresyon (RCBR) ve Sağlam Korelasyon Ölçekli Ana Bileşen Regresyonu (RCSPCR) ve MacroRCSPCR adında üç önerinin mevcut yöntemler ile kıyaslanması.



HACETTEPE
ÜNİVERSİTESİ

Makale Hakkında

Makale İncelemesi

- **Makalenin Konusu**

Çoklu regresyonda etkili gözlemler ve çoklu bağlantı sorunu ile başa çıkmak için Sağlam Korelasyon Tabanlı Regresyon (RCBR) ve Sağlam Korelasyon Ölçekli Ana Bileşen Regresyonu (RCSPCR) ve MacroRCSPCR adında üç önerinin mevcut yöntemler ile kıyaslanması.

- **Makalenin Amacı**

Makalenin amacı, yüksek boyutsallık, aykırı değerler ve aynı zamanda eksik gözlemlerin olduğu durumlarla başa çıkabilecek bir yöntem bulmak ve doğruluğundan emin olmaktır.



HACETTEPE
ÜNİVERSİTESİ

Makalenin Metodolojisi

Makale İncelemesi

Makale metodolojik olarak üç adımda ilerlemektedir.

- Sağlam Korelasyon Ölçekli Ana Bileşen Regresyonu (RCSPCR)
- Sağlam Korelasyon Tabanlı Regresyon (RCBR)
- Makro Korelasyon Ölçekli Temel Bileşen Regresyonu (MacrorCSPCR)

RCSPCR ve RCBR, verilerdeki etkili ve aykırı gözlemleri ele almak için tasarlanırken, MacrorcSPCR”nin hem etkili hem de eksik gözlemlerle eş zamanlı çalışması için tasarlanıyor ve öneriliyor.



HACETTEPE
ÜNİVERSİTESİ

Kullanılan Veriler

Makale İncelemesi

İstatistiksel yöntemlerin performansını test etmek için kullanılan yapay olarak oluşturulmuş simüle veriler, normal dağılımdan elde edilmiş simüle veriler olarak ifade ediliyor.

Ayrıca ek olarak önerilen yöntem RCSPCR ve RCBR'nin Data Dictionary'den¹ alınan kanser veri seti üzerindeki performansı da gerçek hayat uygulaması başlığı altında değerlendiriyor.

¹Veri seti artık mevcut görünmüyor ancak yine de bağlantısını ekledim.



HACETTEPE
ÜNİVERSİTESİ

İçerik Şeması

3 Sonuç ve Tartışma

► Makalenin Yayınlandığı Dergi Hakkında

► Makale Hakkında

► Sonuç ve Tartışma

► Görüş ve Öneriler

► Referanslar



HACETTEPE
ÜNİVERSİTESİ

Sonuç ve Tartışma

3 Sonuç ve Tartışma

Sonuç olarak makale, veri matrisindeki etkili gözlemlerle başa çıkmak için iki teknik önermektedir (RCBR, RCSPCR). Önerilen yöntemler mevcut yöntemlerle karşılaştırılıyor ve MacroRCSPCR'nin² çoklu doğrusallık, veri setinin yüksek boyutluluğu, aykırı değerler ve kayıp gözlemlerle başa çıkabileceği öne sürülüyor.

²MacroRCSPCR, çoklu bağlantı sorunu, veri kümesinin yüksek boyutluluğu, aykırı değerler ve eksik gözlemlerle aynı anda başa çıkan, önerilen tekniklerden biridir.



HACETTEPE
ÜNİVERSİTESİ

Sonuç ve Tartışma

3 Sonuç ve Tartışma

Önerilen teknikler, belirli kirlilik seviyelerine sahip (gürültülü) çeşitli simüle edilmiş senaryolar ve kanser verisi dikkate alınarak değerlendiriliyor. Sonuçlar³, MacroRCSPCR'nin eksik ve aykırı verilerin analizinde daha güvenilir olduğunu gösteriyor.

³Sonuçlar değerlendirilirken kullanılan metrikler, R^2 ve APE metriğidir.



HACETTEPE
ÜNİVERSİTESİ

İçerik Şeması

4 Görüş ve Öneriler

► Makalenin Yayınlandığı Dergi Hakkında

► Makale Hakkında

► Sonuç ve Tartışma

► **Görüş ve Öneriler**

► Referanslar



HACETTEPE
ÜNİVERSİTESİ

Makale Hakkındaki Kişisel Görüşlerim

4 Görüş ve Öneriler

Metrikler

Öncelikle makalede gördüğüm en büyük eksik şu ki sadece iki metrik⁴ ile model değerlendirmesi bence yanlış. En az üç metrik beklerdim ki bence herkes tarafından çok hızlı bir şekilde anlaşılan bir metrik olan ortalama yüzde hata (MAPE) metriğini kesinlikle beklerdim. Zira MAPE kişinin aklında direkt bir olgu oluşturabiliyor.

Regresyon katsayılarının varyans-kovaryans matrisleri ile de kıyaslama yaptıklarını söyleseler de bunu metriklerden saymak istemedim. Zira ortada üç model varken X verileri üzerindeki metriklerini incelemek istesek bu matrisi incelemek diğerlerine nazaran gereksiz fazla zaman alacaktır.

⁴ R^2 ve APE (Average Prediction Errors)



HACETTEPE
ÜNİVERSİTESİ

Makale Hakkındaki Kişisel Görüşlerim

4 Görüş ve Öneriler

Veri Setleri

Normal dağılımdan çektiği verilere biraz gürültü ekleyerek modelleri değerlendirmişler. Açıkcası bu adımı gereksiz buldum. Bu adım ile uğraşmak yerine direkt gerçek hayattan çeşitli veriler elde etseler daha verimli olurdu diye düşünüyorum. Ancak yine de simüle edilen veriler ile de çalıştık onlarla da değerlendirme yaptık diyebilmek için bir seçenek olarak belirtilmiş.

Gerçek hayat verilerine referans olarak bir veri seti⁵ belirtmişler. Fakat veri setinin `.head()` üst kısmını bile göremiyoruz. Heralde direkt bağlantı vermiş olmalarından kaynaklı gerek duymamış olabilirler. Ancak şöyle ki 31 Mart 2023 'te yayımlanmış bir makale için veri setinin bağlantısının kaybolabileceğini öngörmelerini beklerdim.

⁵Kanser veri seti: Artık bağlantıya ulaşamıyor.



HACETTEPE
ÜNİVERSİTESİ

Makale Hakkındaki Kişisel Görüşlerim

4 Görüş ve Öneriler

Karmaşıklık (Complexity)

Eleştirmeyi seviyorum sanırsam ki baya bir eleştirmiş oldum buraya kadar. Ancak bu eleştirimi de yapıp bitireceğim. Önerilen modellerin karmaşıklığına dair bir bilgi bizimle paylaşılmamış. Her şey iyi hoş ancak bu modeller ne kadar sürede çalışıyor ve ne kadar karmaşıklığa sahipler bilmiyoruz (O , O^n , O^{2n} , O^{n^2} ...). Bunu da belirtmelerini beklerdim.



HACETTEPE
ÜNİVERSİTESİ

İçerik Şeması

5 Referanslar

► Makalenin Yayınlandığı Dergi Hakkında

► Makale Hakkında

► Sonuç ve Tartışma

► Görüş ve Öneriler

► Referanslar



Referanslar

5 Referanslar

- H. Abdi and L.J. Williams, *Principal component analysis*, Wiley Interdiscip. Rev. Comput. Stat. 2 (4), 433-459, 2010.
- C. Agostinelli, A. Leung, V.J. Yohai and R.H. Zamar, *Robust estimation of multivariate location and scatter in the presence of cellwise and casewise contamination*, Test 24 (3), 441-461, 2015
- M.H. Ahmad, R. Adnan and N. Adnan, *A comparative study on some methods for handling multicollinearity problems*, Matematika (Johor) 22 (2), 109-119, 2006.
- A. Alfons, *Package "robustHD: Robust methods for high-dimensional data"*, R package version: 0.5.1, 2016.



HACETTEPE
ÜNİVERSİTESİ

Teşekkürler

Ramazan Erduran

21821809