

Modern Uygulamalı Matematik ve Modern Sürekli Optimizasyon Bağlamında Parametrik Olmayan Regresyon Eğrilerinin Bilim, Mühendislik ve Finans Alanlarındaki Güncel Uygulamaları

Semih Kuter^{1, 2}, Gerhard-Wilhelm Weber², Bülent Karasözen²

¹ Çankırı Karatekin Üniversitesi, Orman Fakültesi, Orman Mühendisliği Bölümü, 18200, Çankırı, Türkiye

² Orta Doğu Teknik Üniversitesi, Uygulamalı Matematik Enstitüsü, 06800, Ankara, Türkiye

semihkuter@karatekin.edu.tr, gweber@metu.edu.tr, bulent@metu.edu.tr

Özet: Modern uygulamalı matematiğin temel ve nihai amacı disiplinler arası işbirliğini teşvik ederek matematik ve matematik uygulamalarının biyoloji, kimya, mühendislik, yer bilimleri, nöro-bilim, fizik, finans ve diğer birçok bilim ve mühendislik dalları ile olan bağlarını ortaya çıkarmaktır. Modern uygulamalı matematik alanındaki son gelişmeler, pek çok etkinin doğrusal olmayan davranış sergilediği gerçek yaşama ve doğal olaya ait problem ve olgunun daha iyi anlaşılması ve modellenmesine yönelik gelecek vaadeden yöntemler sunmaktadır. Bu çalışmada amacımız, yukarıda bahsedilen temeller çerçevesinde, son zamanlarda geliştirilen, istatistiksel öğrenme, ters problemler ve çok amaçlı optimizasyon teorilerinden kaynağını alan ve parametrik olmayan bir regresyon ve sınıflandırma aracı olan konik çok değişkenli uyarlanabilir regresyon eğrileri ve türevlerinin çok farklı alanlardaki başarılı uygulamalarının tanıtılmasıdır.

Anahtar Sözcükler: Parametrik Olmayan Regresyon Eğrileri, Modern Sürekli Optimizasyon, MARS, CMARS.

Recent Applications of Nonparametric Regression Splines in Science, Engineering and Finance within the Context of Modern Applied Mathematics and Modern Continuous Optimization

Abstract: The basic and ultimate goal of modern applied mathematics is to explore the connections between mathematics and its applications in biology, chemistry, engineering, geosciences, physics, neuroscience, finance and many other branches by fostering interdisciplinary collaboration. Recent advances in modern applied mathematics offer promising avenues for a better understanding and modelling of real-life problems and natural phenomena, where many effects often exhibit a nonlinear behavior. In this study, within the above mentioned context, we aim to represent the wide range of successful applications of our recently developed nonparametric classification and regression tool conic multivariate adaptive regression splines and its variations, which originates from statistical learning, inverse problems and multiobjective optimization theories.

Keywords: Nonparametric Regression Splines, Modern Continuous Optimization, MARS, CMARS.

1. Giriş

Öğrenme görevinin temel amacı, eğitim verilerini (sonlu sayıdan oluşan gözlemler) kullanarak gelecekteki (test) veriler için bilinmeyen (belirleyici-tepki) bir bağımlılığı (veya modeli) iyi bir öngörü (genelleme) yeteneği ile tahmin etmektir. Regresyon girdi uzayından, X , çıktı uzayına, Y , bir eşleme öğrenmek olup, tahminleyici olarak adlandırılan bu eşleme, f , nicel çıktıları ($X = R^d, Y = R$) tahmin etmek için kullanılır. Diğer yandan, sınıflandırmanın temel amacı ise, özellikler uzayından, X , etiket uzayına, Y , bir eşleme öğrenmektir. Bu eşleme, f , ise nitel çıktıları ($X = R^d, Y = \{0,1\}$) tahmin etmek için kullanılır ve sınıflandırıcı olarak adlandırılır. Öğrenme görevinin adlandırılması her ne kadar çıktı türüne bağlı olsa da, her ikisi

de ortak özelliklere sahip olup, fonksiyon yaklaşımı olarak kabul edilebilir [18].

Temel bilimler, mühendislik, finans, enerji sektörü gibi pek çok alanda bağımlı (tepki) değişken ile bağımsız (belirleyici) değişkenler arasındaki ilişki çoğunlukla doğrusal olmayan davranış sergiler. Bu nedenle parametrik olmayan regresyon ve sınıflandırma yöntemleri veri madenciliği ve tahmin teorisinde çok yaygın olarak kullanılmaktadır [21].

Parametrik olmayan bir regresyon ve sınıflandırma yöntemli olan, ve Friedman [12] tarafından geliştirilen “çok değişkenli uyarlanabilir regresyon eğrileri – multivariate adaptive regression splines” (MARS) algoritması, doğrusal olmayan ve etkileşimli olayları otomatik modelleyen

doğrusal modellerin gelişmiş bir halidir. Hem sınıflandırma hem de regresyonda çok büyük bir öneme sahip olan MARS yöntemi özellikle çok sayıda değişkenin karmaşık ilişkilerinin modellendiği ekonomi, teknoloji ve bilim alanlarında başarı ile uygulanmaktadır.

MARS yöntemi regresyon modelini oluştururken ileri ve geri adım algoritması diye adlandırılan iki aşamalı bir algoritma kullanmaktadır. İlk aşamada en yüksek karmaşıklık düzeyine ulaşınca kadar temel fonksiyon (TF)'lar eklenerek model yapılandırılır. İlk aşamada kullanılan ileri adım algoritmasıyla elde edilen model istenilenden daha karmaşık bir yapıya sahip olduğundan dolayı ikinci aşamada geri adım algoritması ile modeldeki TF'ler sırasıyla elenerek optimum modele ulaşılmaktadır [13].

Sınıflandırma ve tahminleme ile ilgili MARS'ta karşılaşılan bazı problemlerin giderilmesi ve mevcut yöntemlerin kullanım kolaylığı ve/veya etkinliğinin artırılması yönünde gerçekleştirilmiş olan çalışmalar neticesinde MARS'a alternatif olarak "konik çok değişkenli uyarlanabilir regresyon eğrileri – conic multivariate adaptive regression splines" (CMARS) geliştirilmiştir [31]. CMARS istatistiksel öğrenme, ters problemler, sürekli ve çok amaçlı optimizasyon teorileri kullanarak geliştirilmiş bir yöntemdir. CMARS yöntemi MARS algoritmasının geri doğru adım aşamasını kullanmak yerine, cezalı hata kareler toplamını (PRSS) esas alarak, MARS modelini bir Tikhonov düzenleştirmesi (TR) [4] problemine dönüştürmekte ve bu problemi iç nokta yöntemi'nin kullanımına imkan veren "konik karesel programlama – conic quadratic programming" (CQP) [3] ile çözmektedir.

Bu çalışmada MARS, CMARS ve türevlerinin temel bilim, mühendislik ve finans gibi farklı alanlardaki başarılı uygulamalarından örnekler verilerek, modern uygulamalı matematik ve modern sürekli optimizasyon alanlarında geliştirilen bu yöntemlerin kazandırdığı potansiyelin gelecekte nasıl kullanılabileceği konusuna ışık tutulmaya çalışılmıştır.

Çalışmanın geri kalanı şu şekilde düzenlenmiştir: Bölüm 2'de MARS ve CMARS algoritmalarının matematik temelleri özetlenmiştir. MARS, CMARS ve türevlerinin farklı uygulamaları Bölüm 3'te verilmektedir. Bölüm 4'te ise sonuçlar verilerek, gelecekteki potansiyel çalışma olanaklarından bahsedilmekte ve öneriler sunulmaktadır.

2. MARS ve CMARS Algoritmaları

Regresyon analizi, istatistiksel öğrenmede çok

sayıda bağımsız değişkenin modellendiği ve analiz edildiği bir yöntemdir. Parametrik olmayan regresyon analiz türleri olan MARS ve CMARS, bağımlı değişken ve bağımsız değişkenler arasında yatan fonksiyonel ilişki hakkında herhangi bir özel varsayım yapmaz. Bu bölümde MARS ve CMARS algoritmaları [12, 13, 23, 31]'a dayanılarak özetlenip, aktarılmaktadır.

2.1. MARS Yöntemi

Değiştirilmiş yinelemeli bölümlleme metodolojisine dayanan MARS algoritması "sınıflandırma ve regresyon ağaçları - classification and regression trees" (CART)'ın bir uzantısıdır ve her ikisi de iki simetrik TF'in düğüm yerinde oluşturulduğu aralıkların bölünmesi işlemi açısından benzerdir. Bununla birlikte, MARS algoritması sürekli parçalı doğrusal fonksiyonları kullanılır ve doğrusal olmaya ilişkileri daha etkili modelleyebilen sürekli bir model oluşturur. TF'lerin seçimi veriye dayalı ve çalışılan probleme özgü olup, bu da MARS'ı çok boyutlu problemlerin çözümünde uyarlanabilir bir regresyon tekniği yapmaktadır. MARS modeli oluşturulurken, kısmi doğrusal TF'ler bağımlı değişkeni belirlemek için bağımsız değişkenlerin katkısız ve etkileşimli etkilerini dikkate alacak şekilde birbirine eklenir.

MARS aşağıdaki kesik kısmi doğrusal baz fonksiyonlarının açılımlarını kullanır:

$$\begin{aligned} [x - \tau]_+ &= \begin{cases} x - \tau, & \text{if } x \geq \tau, \\ 0, & \text{otherwise,} \end{cases} \\ [\tau - x]_+ &= \begin{cases} \tau - x, & \text{if } x \leq \tau, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Yukarıdaki ifadede, τ tek değişkenli düğüm noktası olup ($x \in \mathbb{R}$), bu iki fonksiyon yansıyan çift olarak adlandırılır ve '+' simgesi sadece pozitif parçaların kullanıldığını, aksi halde sıfır olduğunu gösterir. Genel regresyon modelindeki bağımsız değişkenler ile bağımlı değişken arasındaki ilişki aşağıdaki ifadeyle tanımlanır:

$$Y = f(\mathbf{X}) + \varepsilon. \quad (1)$$

Burada Y bağımlı değişkeni, $\mathbf{X} = (x_1, x_2, \dots, x_p)^T$ bağımsız değişkenler vektörünü, ise ilave sıfır ortalama ve sonlu varyansa sahip stokastik bileşeni göstermektedir. MARS'ın arkasında yatan mantık, p-boyutlu düğüm noktasına

$\tau_i = (\tau_{i,1}, \tau_{i,2}, \dots, \tau_{i,p})^T$ sahip her bağımsız değişken (x_j ($j=1,2,\dots,p$)) için, o model girdisine ait her bir veri vektöründe yansımaları çiftler üretmektir.

Dolayısıyla, MARS'taki 1-boyutlu TF'ler topluluğu aşağıdaki şekilde verilmektedir:

$$C := \left\{ \left[\frac{\mathcal{X}_j}{\tau} - \tau \right]_+, \left[\tau - \frac{\mathcal{X}_j}{\tau} \right]_+, \mid \tau \in \{x_{1,j}, x_{2,j}, K, x_{N,j}\}, j \in \{1, 2, K, p\} \right\}. \quad (2)$$

Yukarıda N toplam gözlem sayısını, p girdi uzayının boyutunu göstermektedir. İfade (1)'de yer alan $f(\bar{X})$, C kümesince art arda oluşturulan ve β_0 kesişim noktalı doğrusal bir kombinasyon olarak aşağıdaki şekilde ifade edilebilir:

$$Y = \beta_0 + \sum_{m=1}^M \beta_m B_m \left(\frac{\mathcal{X}}{\tau} \right) + \varepsilon. \quad (3)$$

İfade (3)'de, B_m C kümesinden bir TF ya da iki veya daha fazla TF'in çarpımı olup, M adet doğrusal olarak bağımsız TF'ler kümesinden alınmıştır. Burada $\mathcal{X}$, fonksiyonuna katkıda bulunan $\bar{X}$ 'in alt vektörü olup, β_0 minci TF'in bilinmeyen katsayısını veya sabiti temsil etmektedir. Varolan bir TF'i başka bir değişkeni içeren diğer bir yansıyan çiftle çarparak, farklı bağımsız değişkenler arasındaki etkileşimi tanımlayan yeni bir TF üretilerek, hem var olan TF'ler hem de yeni oluşturulan TF'ler modele dahil edilir. Bu yolla yüksek boyutlarda eğrilerin eklenmesi sağlanarak aşağıdaki ifadeyle verilen çok değişkenli eğri TF'leri oluşturulur:

$$B_m \left(\frac{\mathcal{X}}{\tau} \right) := \prod_{k=1}^{K_m} \left[s_{K_j^m} \cdot \left(\frac{\mathcal{X}_{K_j^m}}{\tau_{K_j^m}} - \tau_{K_j^m} \right) \right]_+ \quad (i = 1, 2, K, N). \quad (4)$$

Yukarıda verilen İfade (4)'de, minci TF'de çarpılan toplam kesikli doğrusal fonksiyonların sayısı K_m ile, minci TF'de yer alan kinci kesikli doğrusal fonksiyona ait bağımsız değişken $\mathcal{X}_{K_j^m}$ ile, $\tau_{K_j^m}$ için belirlenen düğüm noktası ise $\tau_{K_j^m}$ ile belirtilmekte olup, son olarak $s_{K_j^m} \in \{\pm 1\}$

MARS yöntemi regresyon modelini oluştururken ileri ve geri adım algoritması diye adlandırılan iki aşamalı bir algoritma kullanmaktadır. İlk aşamada en yüksek karmaşıklık düzeyine ulaşıncaya kadar temel fonksiyonlar eklenerek model yapılandırılır. İlk aşamada kullanılan ileri adım algoritmasıyla elde edilen model istenilenden daha karmaşık bir yapıya sahip olduğundan dolayı ikinci aşamada geri adım algoritması ile modelin tahmin etme yeteneğine katkısı yetersiz olan TF'ler sırasıyla elenerek optimum modele ulaşılmaktadır.

Yukarıda anlatılan işlemler sonucunda elde

edilen model serilerinden, optimum terim sayısına, α , sahip en iyi tahmin etme performansını veren model, $\hat{f}_\alpha$, aşağıda verilen "genelleştirilmiş çapraz doğrulama - generalized cross validation" (GCV) prensibine dayanılarak bulunur:

$$GCV(\alpha) := \frac{\sum_{i=1}^N (y_i - \hat{f}_\alpha(x_i))^2}{(1 - Q(\alpha) / N)^2}. \quad (5)$$

Burada örnek gözlem sayısını, $Q(\alpha) = u + dK$; K ileri adımda seçilen düğüm noktası sayısını, u modeldeki doğrusal olarak bağımsız fonksiyon sayısını, d ise her bir TF'in optimize edilme maliyetini vermekte olup, genellikle $d = 3$ alınmaktadır (model eklemeli olduğunda $d = 2$ kullanılır).

2.2. CMARS Yöntemi

Şu ana kadar tartışıldığı üzere, MARS iki algoritmadan oluşmaktadır; ileri adım ve geri adım, ve bu iki algoritma ile iki farklı önemli işlevi aynı anda yerine getirmektedir: (i) hem veriye daha iyi uyan bir model oluşturmakta, (ii) hem de modeli mümkün olduğunca basitleştirmektedir.

CMARS yönteminde ise, MARS'ın geri adım algoritmasını kullanmak yerine, ileri adımda elde edilen $M_{\max}$ adet TF'ler alınarak en küçük kareler kestirimine ceza terimleri eklenmekte, ve bu yolla kestirimin karmaşıklığı (doğruluk) ve tutarlılığı arasındaki ödünleşmeye yeni bir yaklaşım getirilmektedir. MARS'ın ileri adım aşamasında elde edilen PRSS aşağıda verilmektedir:

$$PRSS := \sum_{i=1}^N (y_i - f(\mathcal{X}))^2 + \sum_{m=1}^{M_{\max}} \varphi_m \sum_{\substack{\delta=(\delta_1, \delta_2)^T \\ \delta_1 \in \{0,1\}, \delta_2 \in \{0,1\}}} \sum_{r,s \in \mathcal{I}_m} \int_{\mathcal{Q}^m} \lambda_m^2 \left[G_{r,s}^\delta B_m(\mathbf{h}^m) \right]^2 d\mathbf{h}^m. \quad (6)$$

İfade (6)'da, $\mathcal{I}_m := \{K_j^m \mid j = 1, 2, K, K_m\}$ minci TF'le ilgili değişken kümesini, $\mathbf{h}_m = (h_{K_1^m}, h_{K_2^m}, K, h_{K_m^m})^T$ minci TF'e katılan değişkenler vektörünü, $\varphi_m \approx 0$ ceza parametrelerini, $\mathcal{Q}^m$ ise uygun integral alanlarını belirtmektedir. Son olarak, $\delta = (\delta_1, \delta_2)^T$ için $G_{r,s}^\delta B_m(\mathbf{h}^m) := (\partial^{\delta_1} B_{K_1^m} / \partial^{\delta_2} B_{K_2^m} \dots \partial^{\delta_K} B_{K_m^m})(\mathbf{h}^m)$ ve $|\delta| := \delta_1 + \delta_2$ ($\delta_1, \delta_2 \in \{0,1\}$).

Bu optimizasyon probleminde doğruluk ve karmaşıklık arasındaki ödünleşme ceza parametreleri (φ_m) yoluyla kurulmakta olup, ayrıklaştırma ile çok değişkenli integral $\int_{\mathcal{Q}^m} [G_{r,s}^\delta B_m(\mathbf{h}^m)]^2 d\mathbf{h}^m$, yaklaşım gerçekleştirildikten sonra, İfade (6)'daki PRSS aşağıdaki şekilde yazılmaktadır:

$$PRSS \approx \left\| \mathbf{y} - \mathbf{B}(\hat{\theta}) \lambda \right\|_2^2 + \sum_{m=1}^{M_{\max}} \varphi_m \sum_{i=1}^{(N+1)^{K_m}} L_{im}^2 \lambda_m^2. \quad (7)$$

Burada, $\mathbf{y} := (y_1, y_2, K, \dots, y_N)^T$ bağımlı değişken vektörü, $\mathbf{B}(\hat{\theta}) = (\mathbf{B}(\hat{\theta}_1), \mathbf{B}(\hat{\theta}_2), K, \mathbf{B}(\hat{\theta}_N))^T$ ise $(N \times (M_{\max} + 1))$ ise boyutundaki matris, ve son olarak da $\|\cdot\|_2$ Öklit normudur.

İfade (7)'deki her bir türev için farklı ceza parametresi kullanmak yerine, aynı ceza parametresi $\varphi > 0$, $\varphi = \phi^2$ kullanıldığında, PRSS aşağıdaki forma dönüşmektedir:

$$PRSS \approx \|y - B(\theta)\lambda\|_2^2 + \varphi \|L\lambda\|_2^2. \quad (8)$$

Yukarıdaki ifadede, L diyagonal $((M_{\max}+1) \times (M_{\max}+1))$ boyutunda matrisi, λ ise veri noktalarının kestirimi yapılan $((M_{\max}+1) \times 1)$ boyutundaki parametre vektörünü vermektedir.

İfade (8)'deki TR problemi, bir sürekli, aslında, dışbükey optimizasyon tekniği olan CQP kullanılarak, uygun bir sınırlama ölçütü θ ile aşağıda verilen şekilde çözülebilmektedir:

$$\begin{aligned} & \underset{h, \lambda}{\text{minimize}} \quad h, \\ & \text{subject to} \quad \|B(\theta)\lambda - y\|_2 \leq h, \quad \|L\lambda\|_2 \leq \sqrt{\theta}. \end{aligned} \quad (9)$$

Bu noktada, θ için sınırlama değerlerinin elde edilebilmesi için, dikkatli bir öğrenme süreci takip edilmelidir. Sürekli optimizasyon tekniklerinin modern yöntemleri uygulandığında, CQP aşağıdaki temel gösterimle ifade edilebilir:

$$\begin{aligned} & \underset{x}{\text{minimize}} \quad c^T x, \\ & \text{subject to} \quad \|G_i x - g_i\|_2 \leq p_i^T x - q_i \quad (1, 2, K, k). \end{aligned} \quad (10)$$

3. MARS, CMARS ve Türevlerinin Uygulama Alanları

Bölüm 2.1'de ayrıntıları verilen MARS yönteminin, bilim ve mühendislik alanlarında çok farklı disiplinler tarafından gerçekleştirilmiş başarılı uygulamaları bulunmaktadır. Deichmann v.d. [10] tarafından yapılan finans alanındaki doğrudan pazarlama uygulamaları, Leathwick v.d. [19]'nin gerçekleştirdiği ekolojide çevre ve türlerin dağılımı arasındaki ilişkilerin incelenmesi, Krzyścin v.d. [15]'nin jeofizik alanında atmosferik içerik ve dinamiklerin modellenmesi çalışması, Kuter v.d. [17], [18] tarafından yapılan uzaktan algılamada uydu görüntülerindeki atmosferik etkilerin modellenmesi ile, Quirós v.d. [26] ve Kuter [16]'in çok spektral uydu görüntülerinin sınıflandırılması, Anaklı [2]'nin kalite verilerinin tahmin ve sınıflandırma modellerinin geliştirilmesi çalışmaları bu uygulamalara örnek olarak verilebilir.

Ancak MARS veriye uyum ve modelin karmaşıklığını dengelemede katı bir yaklaşım izlemektedir. Bölüm 2.2'de özetlenen CMARS yaklaşımı bu konuda kullanıcıya esneklik sağlamaktadır. CMARS yer alan optimizasyon problemdeki sınırların çok amaçlı optimizasyon yaklaşımı [27] kullanılarak belirlenmesiyle çok sayıda

alternatif çözüm elde edilebilmektedir. Böylece kullanıcının amacına en uygun çözüme ulaşılması hedeflenmektedir [29]. Yapılan çalışmalarda MARS ve CMARS metodlarının performansları çeşitli ölçütler kullanılarak karşılaştırılmıştır. Karşılaştırma sonuçları CMARS çözümlerinin MARS'a göre bir çok ölçüt bakımından başarımının yüksek olduğunu göstermiştir [31, 33].

Weber v.d. [31] ve Yerlikaya-Özkurt [33] tarafından yapılan üretimde kalite kontrolüne yönelik veri madenciliği çalışmaları, Özmen v.d. [22]'nin yağış verisinin modellenmesi, Alp v.d. [1] ve Weber v.d. [32]'in kredi temerrüt olasılıklarının tahminine yönelik çalışmaları, Yılmaz v.d. [35]'ce yapılan doğal gaz talebinin tahmini, Taylan v.d. [28] ile Taylan ve Weber [30]'in finans, ekonomi ve çevresel süreçlerle ilgili çalışmaları, Yerlikaya-Özkurt v.d. [34]'nin yer hareketlerinin tahminine yönelik alternatif modelleme ve Kuter v.d. [17]'nin yapmış olduğu uydu görüntüleri üzerindeki atmosferik etkilerin giderilmesine yönelik çalışmalar CMARS uygulamalarına örnek olarak verilebilir.

CMARS'ta MARS'ın ileri adım algoritmasında en yüksek karmaşıklık düzeyinde oluşturulan model kullanıldığından, CMARS çok sayıda değişken kullanılmasından kaynaklanan bir karmaşıklığa sahiptir. Bu karmaşıklığı azaltmak için genelleştirilmiş kısmi doğrusal modeller (GPLMs) [20] analiz edilmiş [9, 14] ve sürekli bir regresyon modeli CMARS ile kesikli bir regresyon modeli lojistik regresyon'un katkıları kullanılarak konik genelleştirilmiş kısmi doğrusal model (CG-PLM) geliştirilmiştir [32]. Bu yöntem bağımsız değişkenleri iki kısma ayırarak, klasik doğrusal modellerle doğrusal olmayan modelleri ekleme olarak birleştirip CMARS algoritmasında çok sayıda değişken kullanılmasından kaynaklanan karmaşıklığın azalmasını sağlamıştır.

MARS ve CMARS yöntemleri bağımsız değişkenlerin sabit olduğunu varsaymaktadır fakat gerçek yaşam verilerinin tümünde, yani hem girdi hem de çıktı değişkenlerinde, belirsizlik bulunmaktadır. Buna ek olarak, veriler optimal deney tasarımının içindeki çeşitliliklerden kaynaklanan küçük değişimlere de maruz kalabilirler. Tüm bunlar amaç fonksiyonu ve olası kısıtlarda da belirsizliklere neden olabilmektedir. Bu nedenler sonucunda optimizasyon probleminin çözümleri problem değişkenlerindeki belirsizliklere karşı kayda değer bir duyarlılık gösterebilmektedir. Bu zorluğu aşabilmek için CMARS modeli ve algoritması, verilerdeki belirsizlikleri ele alacak şekilde yeniden yapılandırılmış; çok düzlemli ve

elipsoidale belirsizlik kümeleri esas alınarak Ben-Tal ve Nemirovski [5-7] ile El Ghaoui ve Lebre [11] tarafından geliştirilmiş sağlam optimizasyon yöntemi kullanılarak sağlamlaştırılmıştır [21, 24]. Sağlam optimizasyon, verileri belirsiz (rasgele) olan ve aykırı gözlemler dışındakilerin sadece bazı belirsizlik kümeleri içerisinde tanımlı olduğu optimizasyon problemlerini ele alan bir yöntemidir [8]. Bu yöntemle verilerin belirsizlik içermesi durumunda tahmin varyansı küçültülmeye çalışılmaktadır. Bilgisayar hesaplamalarında sağlamlaştırılan CMARS modelimizin daha basit bir şekline gerek duyulduğundan “zayıf bir sağlamlaştırma” kavramı da oluşturulmuştur. Böylece hem sağlam CMARS (RCMARS) hem de bunun değiştirilmiş hali olan zayıf sağlam CMARS (WRCMARS)’ın teorik olarak geliştirilmesi açıklanıp, yöntem tanıtılmış ve yöntemin duyarlılığı araştırılmıştır [24, 25]. Bu çalışmalarda, verilerde varolan belirsizliklerin üstesinden gelebilmek amacı ile sağlam ve zayıf sağlam olarak isimlendirilen kombinatoriyel yaklaşım kullanılarak CMARS sağlamlaştırılmıştır. Bu yaklaşımla parametrelerin tahmin varyanslarının indirildiği yapılan simülasyon çalışması sonucunda gözlemlenmiştir [21].

4. Sonuç ve Öneriler

Parametre dışı ve etkileşime sahip çok değişkenli davranışların sıklıkla hakim olduğu gerçek yaşam ve doğadaki süreç ve olguların modellenmesinde çok boyutluluğun getirdiği dezavantajlar çoğu zaman baskın gelebilmektedir. Başka bir deyişle, bu tür veriler üzerinde yüksek dereceli polinomlar kullanarak çalışmak zordur. Diğer taraftan, regresyon eğrilerini kullanmak bize her boyutta polinom parçalarının derecelerini oldukça düşük tutma imkanı sağlamaktadır. Regresyon eğrileri aslında çok boyutlu karmaşık veri yapılarına yaklaşmak için oldukça “esnek” tir. Ayırık veriye “yumuşak” bir şekilde yaklaşımda bulunduklarından dolayı, genellikle “yumuşatma eğrileri” olarak da adlandırılırlar.

CMARS’ta kullandığımız eğriler ise, birinci ve özellikle ikinci dereceden kare türevlerinin (diğer bir deyişle, karmaşıklıklarının) integrallerinin cezalandırılarak salınımlarının kontrol altında tutulmasından dolayı daha da “yumuşaktır”. Daha sonrasında, bu integraller ayrıklaştırılarak TR programı elede edilmekte ve CQP biçiminde sunulmaktadır.

Şu ana kadar yapmış olduğumuz ve Bölüm 3’te özetlemiş olduğumuz çalışmalarımız, CMARS ve türevlerinin bilim ve mühendisliğin çok farklı dallarında alternatif bir regresyon ve sınıflandırma aracı olarak kullanılabileceğini göstermektedir.

Modern uygulamalı matematik ve modern sürekli optimizasyon alanlarındaki bilimsel ilerlemelere ait dinamik süreçlerin diğer bilim ve mühendislik alanlarıyla entegre bir şekilde kullanılması hiç şüphesiz karmaşık verilerin içsel yapılarını daha iyi anlayabilmemiz ve daha iyi modeller oluşturabilmemizde önemli rol oynayacaktır.

Kaynaklar

- [1] Alp, Ö.S., Büyükbebeci, E., Çekiç, A.İ., Özkurt, F.Y., Taylan, P. and Weber, G.-W., CMARS and GAM & CQP-Modern optimization methods applied to international credit default prediction. *Journal of Computational and Applied Mathematics*, 2011. 235: p. 4639-4651.
- [2] Anaklı, Z. (2009). A Comparison of data mining methods for prediction and classification types of quality problems. MSc Thesis. The Graduate School of Natural and Applied Sciences, Department of Industrial Engineering, Middle East Technical University, Ankara, Turkey.
- [3] Andersen, E.D., Roos, C. and Terlaky, T., On implementing a primal-dual interior-point method for conic quadratic optimization. *Mathematical Programming*, 2003. 95: p. 249-277.
- [4] Aster, R.C., Borchers, B. and Thurber, C.H., Parameter estimation and inverse problems. 2004, USA: Elsevier, Academic Press.
- [5] Ben-Tal, A. and Nemirovski, A., Robust convex optimization. *Mathematics of Operations Research*, 1998. 23: p. 769-805.
- [6] Ben-Tal, A. and Nemirovski, A., Lectures on modern convex optimization: analysis, algorithms, and engineering applications. Vol. 2. 2001: Siam.
- [7] Ben-Tal, A. and Nemirovski, A., Robust Optimization – Methodology and Applications. *Mathematical Programming*, 2002. 92: p. 453-480.
- [8] Bertsimas, D., Brown, D.B. and Caramanis, C., Theory and applications of robust optimization. *SIAM review*, 2011. 53: p. 464-501.
- [9] Çelik, G. (2010). Parameter estimation in generalized partial linear models with conic quadratic programming. MSc Thesis. Graduate School of Applied Mathematics, Middle East Technical University, Ankara, Turkey.
- [10] Deichmann, J., Eshghi, A., Haughton, D., Sayek, S. and Teebagay, N., Application of Multiple Adaptive Regression Splines (MARS) in direct response modeling. *Journal of Interactive Marketing*, 2002. 16: p. 15-27.
- [11] El-Ghaoui, L. and Lebre, H., Robust solutions to least-square problems to uncertain data matrices. *Sima Journal on Matrix Analysis and Applications*, 1997. 18: p. 1035-1064.
- [12] Friedman, J.H., Multivariate adaptive regression splines. *The Annals of Statistics*, 1991.

19: p. 1-67.

[13] Hastie, T., Tibshirani, R. and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics. 2009, NY, USA: Springer.

[14] Kayhan, B. (2010). Parameter estimation in generalized partial linear models with Tikhonov regularization. MSc Thesis. Graduate School of Applied Mathematics, Middle East Technical University, Ankara, Turkey.

[15] Krzyścin, J.W., Eerme, K. and Janouch, M., Long-term variations of the UV-B radiation over Central Europe as derived from the reconstructed UV time series. *Annales Geophysicae*, 2004. 22: p. 1473-1485.

[16] Kuter, S. (2014). Atmospheric Correction and Image Classification on MODIS Images by Nonparametric Regression Splines. PhD thesis (Research). The Graduate School of Natural and Applied Sciences, Department of Geodetic and Geographic Information Technologies, Middle East Technical University, Ankara, Turkey.

[17] Kuter, S., Weber, G.-W., Akyürek, Z. and Özmen, A., Inversion of top of atmospheric reflectance values by conic multivariate adaptive regression splines. *Inverse Problems in Science and Engineering*, 2014: p. 1-19. DOI: 10.1080/17415977.2014.933828.

[18] Kuter, S., Weber, G.-W., Özmen, A. and Akyürek, Z., Modern Applied Mathematics for Alternative Modeling of the Atmospheric Effects on Satellite Images, in *Modeling, Dynamics, Optimization and Bioeconomics I*, A.A. Pinto and D. Zilberman, Editors. 2014, Springer International Publishing: Switzerland. p. 469-485.

[19] Leathwick, J.R., Elith, J. and Hastie, T., Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modeling*, 2006. 199: p. 188-196.

[20] Müller, M., Estimation and testing in generalized partial linear models - A comparative study. *Statistics and Computing*, 2001. 11: p. 299-309.

[21] Özmen, A. (2010). Robust Conic Quadratic Programming Applied to Quality Improvement - A Robustification of CMARS. MSc Thesis. Institute of Applied Mathematics, Middle East Technical University, Ankara.

[22] Özmen, A., Batmaz, İ. and Weber, G.-W., Precipitation Modeling by Polyhedral RCMARS and Comparison with MARS and CMARS. *Environmental Modeling & Assessment*, 2014. 19: p. 425-435.

[23] Özmen, A., Kropat, E. and Weber, G.-W., Spline regression models for complex

multi-modal regulatory networks. *Optimization Methods and Software*, 2014. 29: p. 515-534.

[24] Özmen, A., Weber, G.-W. and Batmaz, İ., The new robust CMARS (RCMARS) method. *ISI Proceedings of 24th MEC-EuroPT*, 2010: p. 362-368.

[25] Özmen, A., Weber, G.-W., Batmaz, İ. and Kropat, E., RCMARS: Robustification of CMARS with different scenarios under polyhedral uncertainty set. *Communications in Non-linear Science and Numerical Simulation*, 2011. 16: p. 4780-4787.

[26] Quirós, E., Felicísimo, Á.M. and Cuartero, A., Testing Multivariate Adaptive Regression Splines (MARS) as a Method of Land Cover Classification of TERRA-ASTER Satellite Images. *Sensors*, 2009. 9: p. 9011-9028.

[27] Steuer, R.E., *Multiple criteria optimization: Theory, computation, and application*. 1986, New York: Willey.

[28] Taylan, P., Weber, G.-W. and Beck, A., New approaches to regression by generalized additive models and continuous optimization for modern applications in finance, science and technology. *Optimization: A Journal of Mathematical Programming and Operations Research*, 2007. 56: p. 675-698.

[29] Taylan, P., Weber, G.-W. and Yerlikaya-Özkurt, F., A new approach to multivariate adaptive regression splines by using Tikhonov regularization and continuous optimization. *Top*, 2010. 18: p. 377-395.

[30] Taylan, P. and Weber, G.W., Organization in finance prepared by stochastic differential equations with additive and nonlinear models and continuous optimization. *Organizacija (Organization - Journal of Management, Information Systems and Human Resources)*, 2008. 41: p. 185-193.

[31] Weber, G.-W., Batmaz, İ., Köksal, G., Taylan, P. and Yerlikaya-Özkurt, F., CMARS: a new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization. *Inverse Problems in Science and Engineering*, 2011. 20: p. 371-400.

[32] Weber, G.-W., Çavuşoğlu, Z. and Özmen, A., Predicting default probabilities in emerging markets by new conic generalized partial linear models and their optimization. *Optimization: A Journal of Mathematical Programming and Operations Research*, 2012. 61: p. 443-457.

[33] Yerlikaya-Özkurt, F. (2008). A new contribution to nonlinear robust regression and classification with MARS and its applications to data mining for quality control in manufacturing. MSc Thesis. Graduate School of Applied Mathematics, Middle East Technical University, Ankara,

Turkey.

[34] Yerlikaya-Özkurt, F., Askan, A. and Weber, G.-W., An alternative approach to the ground motion prediction problem by a non-parametric adaptive regression method. *Engineering Optimization*, 2014. 46: p. 1651-1668.

[35] Yilmaz, Y., Özmen, A. and Weber, G.-W., A Natural Gas Consumption Forecasting Model for Residential User. Submitted to *Optimization and Engineering*, 2014.