Bursa Uludağ Üniversitesi Mühendislik Fakültesi Dergisi, Cilt 26, Sayı 3, 2021 ARAŞTIRMA

DOI: 10.17482/uumfd.974353

DÜZGÜN OLMAYAN DESTEK VEKTÖR REGRESYONU KISITLI İKİNCİL PROBLEMİNİ ÇÖZMEK İÇİN İKİNCİ DERECEDEN BENZER BİLGİLERE SAHİP BİR ARDIŞIK ASGARİ ENİYİLEME ALGORİTMASI

Aykut KOCAOĞLU [*]

Alınma: 25.07.2021; düzeltme: 30.09.2021; kabul: 13.11.2021

Öz: ε-duyarsız Destek Vektör Regresyonu (ε-DVR), ε-duyarsızlık özelliğine sahip düzenlenmiş l_1 hata kayıp fonksiyonu ile ifade edilir ve l_1 kayıp fonksiyonunun sahip olduğu gürbüz olma özelliği yanında küçük hatalara karşı duyarsız olma özelliğine de sahiptir. Ayrıca, düzenlenmiş hata ile çözümün düzlüğü üzerinde kontrol sağlanır. Bu çalışmada, ε-DVR ikincil problemi, klasik pürüzsüz DVR ikincil probleminin yarısı kadar eniyileme değişkenine sahip olma avantajıyla eşitlik ve eşitsizlik kısıtları altında düzgün olmayan dışbükey parçalı ikinci dereceden problem olarak türetilmiştir. Türetilen bu dışbükey düzgün olmayan ikincil eniyileme problemi, ardışık kayıp fonksiyonu değerleri arasındaki farka ilişkin bir üst sınırın en aza indirilmesine dayanan bir çalışma kümesi seçimi (ÇKS) kullanan verimli bir Ardışık Asgari Eniyileme (AAE) algoritması ile çözülmüştür. Daha önce düzgün olmayan ikincil ε-DVR probleminin AAE algoritması ile çözümünde ÇKS için Karush-Kuhn-Tucker (KKT) koşullarını en fazla ihlal eden çiftler alınarak birinci dereceden bilgiler kullanılmıştır. Önerilen ÇKS'de ise ikinci dereceden benzer bilgiler kullanılmaktadır ve bu düzgün olmayan eniyileme problemini çözmek için birinci dereceden emsaline göre üstünlüğü bir dizi gerçek dünya veri kümesi üzerinde elde edilen sonuçlarla gösterilmiştir. Ayrıca, sonuçlar klasik pürüzsüz DVR ile de karşılaştırılmıştır.

Anahtar Kelimeler: Destek vektör regresyonu, Ardışık asgari eniyileme, Düzgün olmayan eniyileme, Çalışma kümesi seçimi

A Sequential Minimal Optimization Algorithm with Second-Order like Information to solve a Non-Smooth Support Vector Regression Constrained Dual Problem

Abstract: The ε-insensitive Support Vector Regression (ε-SVR) is expressed by the ε-insensitive regularized l_1 error loss function, and has the robustness of the l_1 loss function, as well as being insensitive to small errors. Also, the regularization of error loss provides a control over the flatness of the solution. In this study, the ε-SVR dual problem is derived as a non-smooth convex piecewise quadratic problem under the equality and inequality constraints, with the advantage of having half the optimization variables of the classical smooth SVR dual problem. This convex non-smooth dual optimization problem is solved by an efficient Sequential Minimal Optimization (SMO) algorithm using a working set selection (WSS) based on minimizing an upper bound for the difference between consecutive loss function values. Previously, in the solution of the non-smooth dual ε-SVR problem with SMO algorithm, first order information was used by selecting the pairs in a manner that violates the Karush-Kuhn-Tucker conditions the most. In the proposed WSS, second-order like information is employed and its superiority over the first-order counterpart to solve this non-smooth optimization problem has been demonstrated by the results obtained on a set of real-world datasets. Additionally, the results are compared with the classic smooth SVR problem.

Keywords: Support vector regression, Sequential minimal optimization, Non-smooth optimization, Working set selection

[*] Elektrik ve Enerji Bölümü, İzmir Meslek Yüksekokulu, Dokuz Eylül Üniversitesi, 35380, İzmir, TURKİYE
İletişim Yazarı: Aykut Kocaoğlu (aykut.kocaoglu@deu.edu.tr)

1. GİRİŞ

Destek vektör makineleri (DVM), başlangıçta sınıflandırma problemlerini çözmek için ortaya çıksa da (Boser ve diğ., 1992; Cortes ve Vapnik, 1995) sonrasında regresyon problemlerini ele almak için genişletilmiştir (Vapnik, 1998; Smola ve Schölkopf, 2004). DVR problemini formüle ederken düzenlenmiş ε-duyarsız kayıp fonksiyonu kullanılır ve bu kayıp fonksiyonu genellikle l_1, l_2 ve Huber kayıp fonksiyonlarıdır. Bunlar içerisinde en az mutlak sapmalar diye bilinen l_1 kayıp fonksiyonu aykırı verilere karşı dayanıklı olması ve yaygın olarak kullanılmasından dolayı ön plana çıkmaktadır. ε-duyarsız kayıp fonksiyonundaki ε parametresi çözümün seyrekliği ile ilişkili olan ε-duyarsız bir tüp tanımlar ve bu tüpün dışındaki eğitim örnekleri destek vektörleri oluşturur. Aynı zamanda bu ε-duyarsız tüp küçük hataları yok sayarak bu hatalara karşı da dayanıklı olmasını sağlamaktadır. Ayrıca ε-DVR'de düzenlenmiş hata fonksiyonu ile çözümün düzlüğü üzerinde de kontrol sağlanır. Yukarıda bahsedildiği gibi düzenlenmiş ε-duyarsız hata kayıp fonksiyonu ile formüle edilen ε-DVR problemi, genellikle çekirdek hilesi de kullanılarak Lagrange ikincil problemi olarak ifade edilerek çözümü yapılır. Bu ikincil problem, eğitim örnek sayısının iki katı olan 2L adet Lagrange çarpanının çözümünü gerektiren ikinci dereceden dışbükey bir problemdir. Özellikle doğrusal olmayan durum için çekirdek hilesi DVR'nin önemli bir özelliğidir, yüksek boyutlu bir uzayda örtük olarak çalışmayı ve bu ikincil uzayda çözümü elde etmeyi sağlar.

DVM/DVR'lerin eğitimleri için çeşitli eniyileme yöntemleri mevcuttur (Bottou ve Lin, 2007; Shawe Taylor ve Sun, 2011). Bu yöntemler arasında eniyileme problemlerini ayrıştırarak daha küçük ölçekli problemler haline getiren ve her adımda yalnızca iki Lagrange çarpanını güncelleyen yinelemeli bir algoritma olan AAE, DVM/DVR ikincil problemlerinin çözümünde etkili bir şekilde kullanılmaktadır. Özellikle ortaya çıkan büyük ölçekli verilerin oluşturduğu büyük çekirdek matrisi içeren eniyileme problemlerinin ayrıştırılması çözümde avantaj sağlamaktadır. DVM ikinci dereceden probleminin çözümünde AAE algoritması ilk defa Platt (1998) tarafından geliştirilmiştir. Daha sonra ÇKS için KKT koşullarını en fazla ihlal eden çiftlerin seçimine dayanan birinci dereceden bilgilerin kullanılması Keerthi ve diğ. (2001) tarafından önerilmiştir. Fan ve diğ. (2005), SVM'lerde ortaya çıkan ikinci dereceden problemi çözmek için ÇKS için ikinci dereceden bilgileri kullanmayı önermişlerdir. Ancak bu pürüzsüz ikinci dereceden problem regresyon durumu için 2L eniyileme değişkeni içermektedir. Bazı çalışmalarda ise L adet eniyileme değişkeni ile çalışma avantajı sağlayan düzgün olmayan dışbükey problemin çözülmesi tercih edilmiştir (Flake ve Lawrence, 2002; Guo ve diğ., 2006; Takahashi ve diğ., 2006; Kocaoğlu, 2019). Guo ve diğ. (2006), Takahashi ve diğ. (2006) çalışmalarında ÇKS için en fazla ihlal eden çiftlerin seçimi ile birinci dereceden bilgiler kullanılmıştır. Kocaoğlu (2019) ise l_2 kayıp fonksiyonu içeren düzgün olmayan ikincil DVR probleminin çözümünde ikinci dereceden benzer bilgiler kullanmıştır ve aynı problemin çözümünde ikinci dereceden benzer bilgilerin daha iyi performans sağladığını göstermiştir. Ayrıca, AAE algoritmasının performansını iyileştiren başka çalışmalar da mevcuttur (Barbero ve diğ., 2009; Barbero ve Dorronsoro, 2011; Abe, 2015; Abe, 2016; Noronha ve diğ., 2019; Kumar ve diğ., 2021).

Düzgün olmayan dışbükey DVR ikincil probleminin çözümü için sadece eşitlik kısıtları altında çözüm sağlayan Kocaoğlu (2019) çalışmasında önerilen AAE algoritması, bu çalışmada hem eşitlik hem eşitsizlik kısıtları altında çözüm sağlayacak şekilde geliştirilmiştir. Bu sayede l_1 hata kayıp fonksiyonu içeren DVR ikincil düzgün olmayan dışbükey probleminin çözümünde kullanılmıştır. Bu problem, klasik DVR ikincil probleminin yarısı kadar yani L adet eniyileme parametresiyle uğraşma avantajına sahiptir. Önerilen AAE algoritmasında ÇKS için ikinci dereceden benzer bilgiler kullanılmaktadır ve bu düzgün olmayan eniyileme problemini çözmek için birinci dereceden emsaline göre üstünlüğü bir dizi gerçek dünya veri kümesi üzerinde elde edilen sonuçlarla gösterilmiştir. Ayrıca, sonuçlar klasik pürüzsüz DVR ile de karşılaştırılmıştır.

Bölüm 2'de l_1 hata kayıp fonksiyonu içeren ε-duyarsız Destek Vektör Regresyonu özetlenmiş ve buna ait düzgün olmayan ikincil problem türetilmiştir. Eşitlik ve eşitsizlik kısıtları altında düzgün olmayan dışbükey ε-DVR ikincil problemini çözmek için geliştirilen AAE algoritması Bölüm 3'te anlatılmıştır. Son olarak, kıyaslama veri setleri ile elde edilen sonuçlar Bölüm 4'te sunulmuştur.

2. ε-DUYARSIZ DESTEK VEKTÖR REGRESYONU

Klasik ε-DVR ikincil problemi ve bunun düzgün olmayan versiyonunun türetilmesi aşağıdaki alt bölümlerde gösterilmiştir.

2.1. Klasik ε-duyarsız Destek Vektör Regresyonu

Normalde ε-DVR, ε-duyarsızlık özelliğine sahip düzenlenmiş l_1 hata kayıp fonksiyonu ile ifade edilir. Fakat bu hata kayıp fonksiyonu düzgün bir fonksiyon değildir ve 1. dereceden türevi bulunmamaktadır. O yüzden genellikle eşitsizlik kısıtları kullanılarak aşağıdaki gibi formüle edilir.

$$\min_{\substack{w\in R^n,\\ b\in R}} \frac{1}{2}\parallel \boldsymbol{w} \parallel_2^2 + \frac{C}{2}\sum_{s=1}^{L} e_s + e_s' \tag{1}$$

$$\text{ö}yleki\ \ y_s - (\boldsymbol{w}^T\varphi(\boldsymbol{x}_s) + b) \le e_s + \varepsilon,$$
$$\boldsymbol{w}^T\varphi(\boldsymbol{x}_s) + b - y_s \le e_s' + \varepsilon,$$
$$\forall s \in \{1, \dots, L\}$$

Burada $\mathbf{x}_s \epsilon R^n, y_s \epsilon R^1$ eğitim için kullanılan giriş-çıkış verileridir, $\varphi(\cdot): R^n \rightarrow R^m$ eğitim verilerini yüksek boyutlu bir özellik uzayına eşleyen doğrusal olmayan bir fonksiyondur ve $C \in R^+$ kullanıcı tarafından belirlenen bir düzenlilik parametresidir, $\mathbf{w}$ ve b ise eniyileme parametreleridir. Birincil eniyileme problemi (1) Lagrange çarpanları yöntemi ile aşağıdaki ikincil probleme dönüştürülerek çözülebilir.

$$\min_{\substack{\boldsymbol{\alpha}\in R^L,\\ b\in R}} J(\boldsymbol{\alpha}) = \frac{1}{2}\sum_{s=1}^{L}\sum_{r=1}^{L} K(\mathbf{x}_s, \mathbf{x}_r)(\lambda_s - \lambda_s')(\lambda_r - \lambda_r') + \epsilon\sum_{s=1}^{L}(\lambda_s + \lambda_s') - \sum_{s=1}^{L} y_s(\lambda_s - \lambda_s') \tag{2}$$

$$\text{ö}yleki \sum_{s=1}^{L}(\lambda_s - \lambda_s') = 0,$$

$$0 \le \lambda_s \le C, 0 \le \lambda_s' \le C, \forall s \in \{1, \dots, L\}$$

Bu problemde 2L adet eniyileme parametresi λ_s, λ_s' $\forall s \in \{1, \dots, L\}$ bulunmaktadır ve bu ikinci dereceden programlama probleminin çözümünde Fan ve diğ. (2005) çalışmasında açıklanan AAE gibi ayrıştırma algoritmalarından yararlanılmaktadır.

2.2. ε-duyarsız DVR probleminin düzgün olmayan versiyonu

ε-DVR, eşitlik kısıtlarıyla ε-duyarsızlık özelliğine sahip düzenlenmiş l_1 hata kayıp fonksiyonu ile (3)'deki gibi yazılır.

$$\min_{\substack{w^*\in R^n,\\ b^*\in R}} \frac{1}{2}\parallel \mathbf{w} \parallel_2^2 + \frac{C}{2}\sum_{s=1}^{L} |e_s|_\epsilon \tag{3}$$

$$\text{ö}yleki\ \ e_s = y_s - ((w^*)^T\varphi(\mathbf{x}_s) + b^*),\ \ \forall s \in \{1, \dots, L\}$$

Burada ε-duyarsız l_1 hata kayıp fonksiyonu aşağıdaki gibi tanımlıdır.

$$|e_s|_\epsilon = \begin{cases} (e-\varepsilon) & \text{if } e \geq \varepsilon \\ 0 & \text{if } -\varepsilon < e < \varepsilon. \\ (e+\varepsilon) & \text{if } e \leq -\varepsilon \end{cases} \tag{4}$$

(3)'ün sürekli ancak türevlenemez bir fonksiyon olduğu unutulmamalıdır. (3) düzgün bir fonksiyon olmamasına rağmen KKT koşullarının subdiferansiyel versiyonu kullanılabilir (Ruszczyński, 2006) ve Lagrangian aşağıdaki gibi elde edilir.

$$L(\mathbf{w}, e, b; \alpha_s) = \frac{1}{2} \| \mathbf{w} \|_2^2 + C \sum_{s=1}^{L} |e_s|_\epsilon - \sum_{s=1}^{L} \alpha_s (\mathbf{w}^T \varphi(\mathbf{x}_s) + b + e_s - y_s) \tag{5}$$

Burada α_s Lagrange çarpanlarını ifade etmektedir ve (5)'deki Lagrange fonksiyonu için birinci dereceden eniyilik koşulları subdiferansiyel kullanılarak aşağıdaki gibi yazılabilir.

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{s=1}^{L} \alpha_s \varphi(\mathbf{x}_s) \tag{6}$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{s=1}^{L} \alpha_s = 0 \tag{7}$$

$$0 \in \partial_{e_s} L \rightarrow 0 \in \begin{cases} 0 & \text{if } |e_s| < \varepsilon \\ [-\alpha_s, -\alpha_s + C] & \text{if } e_s = -\varepsilon \\ [-\alpha_s - C, -\alpha_s] & \text{if } e_s = \varepsilon \\ -\alpha_s + C & \text{if } e_s > \varepsilon \\ -\alpha_s - C & \text{if } e_s < -\varepsilon \end{cases} \tag{8}$$

Burada $\partial_{e_s} L$, L'nin e_s'ye göre kısmi subdiferansiyelini belirtir. (8) kullanılarak hata terimleri ile Lagrange çarpanları arasında aşağıdaki ilişki kurulabilir.

$$\begin{cases} e_s = \varepsilon \cdot \text{sign}^\star(\alpha_s) & \text{if } -C < \alpha_s < C \\ e_s > \varepsilon & \text{if } \alpha_s = C \\ e_s < -\varepsilon & \text{if } \alpha_s = -C \end{cases} \tag{9}$$

Burada $\text{sign}^\star(\alpha_s) = \begin{cases} 1 & \text{if } \alpha_s > 0 \\ \lambda \text{ with } \lambda \in [-1,1] & \text{if } \alpha_s = 0 \\ -1 & \text{if } \alpha_s < 0 \end{cases}$ olarak tanımlanmıştır. Ayrıca (4) ve (9)'daki denklemler kullanılarak $C|e_s|_\epsilon - \alpha_s e_s = \varepsilon \cdot \text{sign}^\star(\alpha_s)$ eşitliği yazılabilir. Bu eşitlik ile birlikte (6) ve (7) kullanılıp (5) içerisine yerleştirilirse aşağıdaki ikincil düzgün olmayan dışbükey eniyileme problemi elde edilir.

$$\min_{\alpha_s \in R^L} J(\alpha_s) = \frac{1}{2} \sum_{s=1}^{L} \sum_{r=1}^{L} K(\mathbf{x}_s, \mathbf{x}_r) \alpha_s \alpha_r - \sum_{s=1}^{L} y_s \alpha_s + \varepsilon \sum_{s=1}^{L} |\alpha_s| \tag{10}$$

$$\textit{öyleki} \sum_{s=1}^{L} \alpha_s = 0$$

$$-\mathrm{C} \leq \alpha_s \leq \mathrm{C}, \forall s \in \{1, \dots, \mathrm{L}\}$$

Burada $\alpha_s\ \forall s \in \{1, \dots, \mathrm{L}\}$ Lagrange çarpanlarını ve $\mathrm{K}(\mathbf{x}_s, \mathbf{x}_r) = \varphi(\mathbf{x}_s)^T \varphi(\mathbf{x}_r)$ çekirdek fonksiyonunu temsil etmektedir.

3. AAE ALGORİTMASI

Algoritma 1 Düzgün olmayan kısıtlı dışbükey problemin çözümü için önerilen AAE algoritması

Giriş: Eğitim verileri $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_L, y_L)\}$
Çıkış: $\boldsymbol{\alpha}, b$

Tüm $s \in \{1, \dots, L\}$ için ayarlayarak başlatın: $\boldsymbol{\alpha} = 0$, $\nabla f(\boldsymbol{\alpha}) = \boldsymbol{K\alpha} + \boldsymbol{p}$, $g_s(\boldsymbol{\alpha}) \coloneqq -\nabla f(\boldsymbol{\alpha})_s - \varepsilon sign^+(\alpha_s)$ and $G_s(\boldsymbol{\alpha}) \coloneqq -\nabla f(\boldsymbol{\alpha})_s - \varepsilon sign^-(\alpha_s)$

1: **DÖNGÜ**

2: Çalışma kümesini şu şekilde seçin:

$i \in \underset{t \in \{t | \alpha_t < C\}}{\operatorname{argmax}} \{g_t(\boldsymbol{\alpha})\}$

$j \in \underset{t \in \{t | \alpha_t > -C\}}{\operatorname{argmin}} \{h_{it} |\ G_t(\boldsymbol{\alpha}) < g_i(\boldsymbol{\alpha})\}$

burada $a_{ts} = Q_{tt} + Q_{ss} - 2Q_{ts} > 0$, $b_{ts} = \nabla f(\boldsymbol{\alpha}^k)_s - \nabla f(\boldsymbol{\alpha}^k)_t$ ve

$$h_{it} \coloneqq \begin{cases} -\dfrac{{b_{it}}^2}{a_{it}}, & if\ m_1 < -\dfrac{b_{it}}{a_{it}} < m_2 \\ -\dfrac{(-b_{it} + 2\varepsilon)^2}{a_{it}}, & if\ \dfrac{(-b_{it} + 2\varepsilon)}{a_{it}} < m_1 \\ -\dfrac{(-b_{it} - 2\varepsilon)^2}{a_{it}}, & if\ m_2 < \dfrac{(-b_{it} - 2\varepsilon)}{a_{it}} \\ -a_{it}{m_1}^2, & if\ -\dfrac{b_{it}}{a_{it}} \leq m_1 \leq \dfrac{(-b_{it} + 2\varepsilon)}{a_{it}} \\ -a_{it}{m_2}^2, & aksi\ halde \end{cases}$$

3: Güncelleme değerini hesaplanması:

$$d_j^* \coloneqq \begin{cases} -\dfrac{b_{ij}}{a_{ij}}, & if\ m_1 < -\dfrac{b_{ij}}{a_{ij}} < m_2 \\ \dfrac{(-b_{ij} + 2\varepsilon)}{a_{ij}}, & if\ \dfrac{(-b_{ij} + 2\varepsilon)}{a_{ij}} < m_1 \\ -\dfrac{(-b_{ij} - 2\varepsilon)}{a_{ij}}, & if\ m_2 < \dfrac{(-b_{ij} - 2\varepsilon)}{a_{ij}} \\ m_1, & if\ -\dfrac{b_{ij}}{a_{ij}} \leq m_1 \leq \dfrac{(-b_{ij} + 2\varepsilon)}{a_{ij}} \\ m_2, & aksi\ halde \end{cases}$$

4: Kutu kısıtlamasını sağlaması: $d_j = \max(-C - \alpha_j, -C + \alpha_i, d_j^*)$

5: Güncelle: $\alpha_j \coloneqq \alpha_j + d_j$, $\alpha_i \coloneqq \alpha_i - d_j$ ve $\nabla f(\boldsymbol{\alpha}) \coloneqq \nabla f(\boldsymbol{\alpha}) - \boldsymbol{K}_i d_j + \boldsymbol{K}_j d_j$
Burada $\boldsymbol{K}_i \in R^L$ ve $\boldsymbol{K}_j \in R^L$ **K** matrisinin i. and j. sütunlarıdır.

6: Tüm $s \in \{1, \dots, L\}$ $için$ $g_s(\boldsymbol{\alpha}) \coloneqq -\nabla f(\boldsymbol{\alpha})_s - \varepsilon sign^+(\alpha_s)$ ve $G_s(\boldsymbol{\alpha}) \coloneqq -\nabla f(\boldsymbol{\alpha})_s - \varepsilon sign^-(\alpha_s)$ güncellenmesi.

7: $m(\boldsymbol{\alpha}) - M(\boldsymbol{\alpha}) \coloneqq \max_{s\in\{s|\alpha_s<C\}} g_s(\boldsymbol{\alpha}) - \min_{s\in\{s|\alpha_s>-C\}} G_s(\boldsymbol{\alpha}) \leq \tau$ **İKEN**

8: Hesapla $b \coloneqq \frac{1}{2}\left(\max_{s\in\{s|\alpha_s<C\}} g_s(\boldsymbol{\alpha}) + \min_{s\in\{s|\alpha_s>-C\}} G_s(\boldsymbol{\alpha})\right)$

9: **BİTİR**

AAE, eniyi çözüme yakınsamayı garanti edecek şekilde her adımda yalnızca iki Lagrange çarpanını güncelleyen yinelemeli bir algoritmadır. AAE algoritmasının türetilmesi için, (10)'da ayrıntıları verilen dışbükey düzgün olmayan eniyileme probleminin (11)'deki matris formu dikkate alınmıştır.

$$\min_{\boldsymbol{\alpha}\in R^L} J(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{K}\boldsymbol{\alpha} + \boldsymbol{p}^T\boldsymbol{\alpha} + \varepsilon\|\boldsymbol{\alpha}\|_1 \quad (11)$$
$$\ddot{o}yleki\ \ \boldsymbol{u}^T\boldsymbol{\alpha} = 0,$$
$$-C \preccurlyeq \boldsymbol{\alpha} \preccurlyeq C$$

Burada $\boldsymbol{p}^T = [\mathrm{y}_1, \mathrm{y}_2, \dots, \mathrm{y}_\mathrm{L}]\boldsymbol{\epsilon}R^L$, $\boldsymbol{K}\boldsymbol{\epsilon}R^{LxL}$ çekirdek matrisi, $\boldsymbol{u}^T = [1 \dots 1]\boldsymbol{\epsilon}R^L$ sabit vektör ve $\boldsymbol{\alpha\epsilon}R^L$ Lagrange çarpanlarıdır. (11)'deki düzgün olmayan problemin sadece eşitlik kısıtı altındaki versiyonu daha önce Kocaoğlu (2019) çalışmasında çözülmüştür. Bu çalışmada ise bir önceki çalışmada anlatılan adımlar eşitsizlik kısıtlarını da göz önünde bulunduracak şekilde geliştirilmiş ve (11) için ikinci dereceden benzer bilgiler içeren ÇKS ile verimli bir AAE algoritması önerilmiştir. Bu algoritma her bir adımda eşitlik koşulunu sağlayarak iki Lagrange çarpanını $\alpha_j^{k+1} = \alpha_j^k + d_j$ ve $\alpha_i^{k+1} = \alpha_i^k - d_j$ olacak şekilde aşağıdaki problemin çözümüyle yinelemektedir.

$$\min_{d_j\in R^1} J(d_j) = \frac{1}{2}a_{ij}d_j^2 + b_{ij}d_j + \varepsilon\left|\alpha_i^k - d_j\right| + \varepsilon\left|\alpha_j^k + d_j\right| \quad (12)$$
$$\ddot{o}yleki\ -C - \alpha_j^k \leq d_j \leq -\alpha_j^k,$$
$$-C + \alpha_i^k \leq d_j \leq C + \alpha_i^k$$

Önerilen AAE algoritması için her bir adımda iki adet Lagrange çarpanının seçimi, bunların kutu kısıtları altında eşitsizlik kısıtlarını sağlayacak şekilde yinelenmesi ve algoritmanın KKT koşullarını hemen hemen sağlayacak bir koşul ile bitirilmesini içeren detaylı açıklaması Algoritma 1'de verilmiştir.

4. DENEYLER

Bu bölümde, düzgün olmayan dışbükey kısıtlı ε-DVR ikincil problemini çözmek için İkinci Dereceden Benzer bilgiler kullanan bir ÇKS içeren AAE algoritması (İDB_AAE) önerilmiştir ve aynı problemin çözümünde ÇKS için Birinci Dereceden bilgiler içeren en fazla ihlal eden çiftleri kullanan AAE algoritması (BD_AAE) Guo ve diğ. (2006) çalışması ile karşılaştırılmıştır. Karşılaştırılan her iki problemde L adet eniyileme parametresinin çözümünü yapmaktadır. Bunlara ilaveten 2L adet eniyileme parametresi içeren klasik DVR pürüzsüz ikincil problemini ÇKS için İkinci Dereceden bilgiler kullanarak çözen AAE (İD_AAE) ile de karşılaştırılma yapılmıştır. AAE algoritmalarının kaynak kodları C ++ ile yazılmıştır ve çözüm aşamasında çekirdek matris elemanlarını geçici olarak hafızada tutmak için 100 MB ön bellek kullanılmıştır.

Karşılaştırmalar için aynı platformu sağlamak için tüm deneyler, 8 GB RAM ve 64-bit Windows 10 işletim sistemine sahip Intel Core I7 işlemcili 2,7 GHz bilgisayar üzerinde MATLAB2020b ortamında gerçekleştirilmiştir. Bir çekirdek fonksiyonu olarak, en yaygın kullanılan Gauss fonksiyonu

$K(x_s, x_r) = e^{\frac{-\|x_s - x_r\|_2^2}{2\sigma^2}}$ seçilmiştir. En iyi model parametreleri (çekirdek parametresi σ, düzenlileştirme parametresi C ve duyarsızlık parametresi ε), $\{2^{-3}, 2^{-2}, \cdots, 2^1\} \times \{10^0, 10^1, \cdots, 10^3\} \times \{0, 0.02, 0.05, 0.1, 0.2, 0.5, 1\}$ arasından 5 katlı çapraz doğrulama yaklaşımı kullanılarak bulunmuştur ve Tablo 1'de gösterilmiştir. $f(x_s)$, y_s gerçek değerinin tahmini, x_s giriş örneği ve $\hat{L}$ örneklerin toplam sayısını ifade etmektedir ve karşılaştırma kriteri olarak bunlar cinsinden Karekök Ortalama Karesel Hatası (KOKH) aşağıda tanımlanmıştır.

$$\mathrm{KOKH} = \sqrt{\frac{1}{\hat{\mathrm{L}}} \sum_{\mathrm{s}=1}^{\hat{\mathrm{L}}} \left(\mathrm{y_s} - \mathrm{f}(\mathrm{x_s})\right)^2} \tag{13}$$

Dört veri seti Mpg, Housing, Mg ve Abalone LIBSVM* web sitesinde mevcuttur. Geri kalanlar, iyi bilinen UCI† makine öğrenimi havuzundan alınmıştır. Her veri kümesinin giriş değişkenleri, eğitim sürecinden önce kapalı aralık $[0, 1]$ olacak şekilde normalleştirilmiştir.

Çeşitli veri kümelerine ait 5 katlı çapraz doğrulama yaklaşımı ile elde edilen aynı eğitim verileri ve test verileri üzerinde AAE_BD, AAE_İD ve AAE_İDB yöntemleri ile çözülen ε-DVR ikincil problemine ait ortalama değerler ve standart sapmaları Tablo 2'de sunulmuştur. Bu üç yöntem de ikincil problemi farklı türetmeler ile elde edilen aynı ε-DVR problemini çözdükleri için Destek Vektör (DV) sayıları, eğitim KOKH ve test KOKH değerleri beklenildiği gibi hemen hemen aynı çıkmıştır. Fakat L adet eniyileme parametresi ile uğraşan ε-DVR düzgün olmayan ikincil probleminin çözümünde AAE_BD ve AAE_İDB yöntemleri karşılaştırıldığında, AAE_İDB'nin her bir yinelemede daha fazla işlem yükü olmasına rağmen yineleme sayılarını büyük ölçüde azalttığı ve eğitim süresinin AAE_BD yöntemine kıyasla daha iyi olduğu Tablo 2'de gözlenmiştir. Yineleme sayılarındaki bu iyileştirmenin nedeni önerilen AAE_İDB yönteminin ÇKS için ikinci dereceden benzer bilgiler kullanması ardışık yinelemeler arasındaki kayıp fonksiyon değerleri arasındaki farkın daha fazla olması ile ilişkilidir ve bu durum eniyi değere daha az yinelemede ulaşmasını sağlamaktadır. AAE_İDB ve AAE_İD yöntemleri karşılaştırıldığında ise AAE_İD yönteminin yarısı kadar eniyileme parametresi ile uğraşan AAE_İDB yönteminin yineleme sayıları bakımından biraz daha fazla olmasına rağmen eğitim süresi bakımından daha iyi olduğu Tablo 2'de görülmektedir. Eğitim süresindeki bu iyileşme eniyileme parametresi sayısının yarıya inmesinin yanında AAE_İD yöntemindeki 2Lx2L boyutundaki çekirdek matrisinin önerilen AAE_İDB yönteminde LxL boyutunda olmasıdır. Çekirdek matrisi elemanlarının sürekli hesaplanmasını engellemek amacıyla ön bellek kullanılması AAE yönteminin tercih edilen bir özelliğidir ve daha küçük boyutlu çekirdek matrisi bu elemanların önceden hesaplanması ve önbelleğe alınması olasılığını artırmaktadır. Sonuç olarak ÇKS için ikinci dereceden benzer bilgiler kullanan AAE_İDB yöntemi AAE_BD yönteminin L adet eniyileme parametresi ile uğraşma avantajını ve AAE_İD yönteminin ise ÇKS için İkinci dereceden bilgiler kullanma avantajlarını birleştirerek ε-DVR eğitim süresini azaltan alternatif verimli bir çözüm sunmaktadır.

Tablo 1. Veri kümelerinin ve en iyi hiperparametrelerin kısa bir açıklaması.

Veri kümesi	Veri kümesi boyutu	Hiperparametreler (C, σ, ε)
Servo	167×4	10^1 2^{-1} 0.02
Mpg	392×7	10^1 2^{-1} 1

* https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/.
† http://archive.ics.uci.edu/ml/datasets.html.

Boston	506×13	10^3	2^0	1
Mg	1385×6	10^0	2^{-2}	0.1
Airfoil	1503×5	10^3	2^{-3}	1
Abalone	4177×8	10^2	2^{-1}	1

Tablo 2. Karşılaştırma veri kümelerinin deneysel sonuçları.

Veri kümesi	Yöntem	DV sayısı	Eğitim KOKH	Test KOKH	Yineleme sayısı	Eğitim süresi (sn)
Servo	BD_AAE	114,1±2,59	0,513554±0,13	0,566387±0,35	11699,8±1325,28	0,171094±0,12
	İD_AAE	114,3±2,64	0,513572±0,13	0,566366±0,35	**3765,6**±457,36	0,21875±0,12
	İDB_AAE	114,3±2,45	0,513585±0,13	0,566366±0,35	4964,5±2222,93	**0,1625**±0,037
Auto-mpg	BD_AAE	194,15±6,18	2,256665±0,084	2,625365±0,34	2731,55±605,96	0,158594±0,05
	İD_AAE	194,15±6,22	2,256667±0,084	2,625375±0,34	**670,5**±142,25	0,146875±0,038
	İDB_AAE	194,15±6,22	2,256667±0,084	2,62536±0,34	678,85±145,73	**0,136719**±0,05
Boston	BD_AAE	258,6±6,35	2,001211±0,12	3,356768±0,70	191985,9±41939,15	0,59375±0,20
	İD_AAE	258,55±6,24	2,001208±0,12	3,35678±0,70	**26913,8**±4090,30	0,352344±0,19
	İDB_AAE	258,55±6,24	2,001204±0,12	3,35676±0,70	27789,25±5665,55	**0,261719**±0,11
Mg	BD_AAE	388,8±10,18	0,105521±0,001	0,120824±0,005	6639,25±877,76	0,192188±0,066
	İD_AAE	387,75±10,14	0,105514±0,001	0,120818±0,005	**1630,6**±166,32	0,163281±0,040
	İDB_AAE	387,95±10,28	0,105511±0,001	0,12082±0,005	2296,8±476,78	**0,157813**±0,023
Airfoil	BD_AAE	697,55±13,5	1,510086±0,033	2,461149±0,20	853248,2±134813,1	5,989063±2,78
	İD_AAE	697,65±13,25	1,510082±0,033	2,461178±0,20	**73623,55**±9592,94	1,897656±1,34
	İDB_AAE	697,65±13,25	1,510079±0,033	2,461205±0,20	83095,8±17072,98	**1,3125**±0,42
Abalone	BD_AAE	1687,65±16,75	2,037623±0,020	2,108696±0,084	278025,3±40856,79	10,0875±3,24
	İD_AAE	1687,3±16,66	2,03763±0,020	2,108699±0,084	**22429,5**±3787,29	1,846094±0,95
	İDB_AAE	1687,3±17,02	2,037632±0,020	2,1087±0,084	27494,95±8626,56	**1,521094**±0,47

5. SONUÇ

ε-DVR birincil problemi düzenlenmiş ε-duyarsız l_1 hata kayıp fonksiyonu ile ifade edilir. Fakat bu hata kayıp fonksiyonu düzgün bir fonksiyon değildir ve 1. dereceden türevi bulunmamaktadır. Buna rağmen KKT koşullarının subdiferansiyel versiyonu kullanılarak Lagrange ikincil problemi türetilmiştir. Bu ikincil problem hem eşitlik hem de eşitsizlik kısıtları altında L eniyileme parametresi içeren düzgün olmayan bir problemdir. Klasik ε-DVR probleminin yarısı kadar eniyileme parametresi içerme avantajına sahiptir. Ayrıca çözümü için çalışma kümesi seçiminde ikinci dereceden benzer bilgiler kullanan verimli bir AAE algoritması önerilmiştir. Önerilen AAE algoritmasının, çalışma kümesi seçimi için birinci dereceden bilgiler kullanan emsaline ve 2L eniyileme parametresine sahip klasik AAE algoritmasına kıyasla daha iyi performans gösterdiği ve bu varyantlara kıyasla daha düşük eğitim süreleri gerçekleştirdiği gösterilmiştir. Bu performans iyileştirilmesi aynı zamanda çekirdek matrisinin boyutunun önerilen yöntemde daha küçük olması ve her bir yinelemede bu matris elemanlarının önbellekte saklanma olasılığının artması ile de ilişkili olduğu öngörülmektedir. Önerilen AAE yönteminin diğer kayıp fonksiyonları için de kullanılabilir hale getirilmesi ve yazındaki diğer iyileştirilmeler ile birleştirilmesi de sonraki çalışmalarda hedeflenmektedir.

ÇIKAR ÇATIŞMASI

Yazar, bilinen herhangi bir çıkar çatışması veya herhangi bir kurum/kuruluş ya da kişi ile ortak çıkar bulunmadığını onaylamaktadır.

YAZAR KATKISI

Çalışmanın tamamı Aykut Kocaoğlu tarafından oluşturulmuştur. Çalışmanın son onay ve tam sorumluluğunu yazar üstlenmektedir.

KAYNAKLAR

1. Abe, S. (2015) Optimizing working sets for training support vector regressors by Newton's method, *International joint conference on neural networks, IJCNN,* Killarney, Ireland. doi:10.1109/IJCNN.2015.7280309

2. Abe, S. (2016) Fusing sequential minimal optimization and Newton's method for support vector training, *International Journal of Machine Learning and Cybernetics,* 7(3), 345–364. doi:10.1007/s13042-014-0265-x

3. Barbero, A., Lopez, J, ve Dorronsoro, J.R. (2009) Cycle-breaking acceleration of SVM training, *Neurocomputing,* 72(7–9), 1398–1406. doi:10.1016/j.neucom.2008.12.014

4. Barbero, A. ve Dorronsoro, J.R. (2011) Momentum sequential minimal optimization: an accelerated method for support vector machine training, *International joint conference on neural networks, IJCNN.* doi:10.1109/IJCNN.2011.6033245

5. Boser, B., Guyon, I. ve Vapnik, V. (1992) A training algorithm for optimal margin classifiers, *Proceedings of the fifth annual workshop on Computational Learning Theory,* 144–152. doi:10.1145/130385.130401

6. Bottou, L. ve Lin C.J. (2007) *Support vector machine solvers in Large scale kernel machines,* MIT Press, Cambridge.

7. Cortes, C. ve Vapnik, V. (1995) Support-vector network, *Machine Learning,* 20, 273–297.

8. Fan, R.E., Chen, P.H. ve Lin, C.J. (2005) Working set selection using second order information for training support vector machines, *Journal of Machine Learning Research,* 6, 1889–1918.

9. Flake, G.W. ve Lawrence, S. (2001) Efficient SVM regression training with SMO, *Machine Learning,* 46(1), 271–290. doi:10.1023/A:1012474916001

10. Guo, J., Takahashi, N. ve Nishi, T. (2006) A novel sequential minimal optimization algorithm for support vector regression, *Lecture Notes in Computer Science,* 4232, 827–836. doi:10.1007/11893028_92

11. Keerthi, S.S., Shevade, S.K., Bhattacharyya, C. ve Murthy, K.R.K. (2001) Improvements to Platt's SMO algorithm for SVM classifier design, *Neural Computation,* 13(3), 637–649. doi:10.1162/089976601300014493

12. Kocaoğlu, A. (2019) An Efficient SMO Algorithm for Solving Non-smooth Problem Arising in ε-Insensitive Support Vector Regression, *Neural Processing Letters*, 50, 933–955. doi:10.1007/s11063-018-09975-3

13. Kumar, B., Sinha, A., Chakrabarti, S. ve Vyas, O. P. (2021) A fast learning algorithm for One-Class Slab Support Vector Machines, *arXiv: 2011.03243*. https://arxiv.org/abs/2011.03243 doi:10.1016/j.knosys.2021.107267

14. Lin, C.J. (2001) On the convergence of the decomposition method for support vector machines, *IEEE Transactions on Neural Network*, 12(6), 1288–1298. doi: 10.1109/72.963765

15. Lopez, J. ve Dorronsoro, J.R. (2012) Simple proof of convergence of the SMO algorithm for different SVM variants, *IEEE Transactions on Neural Networks and Learning Systems*, 23(7), 1142–1147. doi:10.1109/TNNLS.2012.2195198

16. Noronha, D.H., Torquato, M.F. ve Fernandes, M.A.C. (2019) A parallel implementation of sequential minimal optimization on FPGA, *Microprocessors and Microsystems*, 69, 138-151. doi: 10.1016/j.micpro.2019.06.007

17. Platt, J.C. (1998) Fast training of support vector machines using sequential minimal optimization *in Kernel methods: support vector machines*, MIT Press, Cambridge.

18. Ruszczynski, A. (2006) *Nonlinear Optimization*, Princeton, NJ: Princeton University Press. ISBN 978-0691119151. MR 2199043.

19. Shawe Taylor, J. ve Sun, S. (2011) A review of optimization methodologies in support vector machines, *Neurocomputing*, 74(17), 3609-3618. doi: 10.1016/j.neucom.2011.06.026

20. Smola, A.J. ve Schölkopf, B. (2004) A tutorial on support vector regression, *Statistics and Computing*, 14(3), 199–222. doi: 10.1023/B:STCO.0000035301.49549.88

21. Takahashi, N., Guo, J. ve Nishi, T. (2006) Global convergence of SMO algorithm for support vector regression, *IEEE Transactions on Neural Network*, 19(6), 971–982. doi:10.1109/TNN.2007.915116

22. Vapnik, V.N. (1998) *Statistical Learning Theory*, Wiley, New York, USA.