YILDIZ TEKNİK ÜNİVERSİTESİ VERİ BİLİMİ PROGRAMI

-

PROJE RAPORU

Destek vektör makinesi sınıflandırmasına ilişkin kapsamlı bir araştırma: Uygulamalar, zorluklar ve trendler

ALİCAN TUNÇ /235B7055

ÖZET

Bu projede,”destek vektör makine sınıflandırmasına ilişkin kapsamlı bir araştırmayı özetleyin: uygulamalar, zorluklar ve trendler” makalesini özetlemeyi amaçladım. Bu inceleme boyunca destek vektör makinelerinin güçlü ve zayıf yönlerini ele alacak, teoriye değindikten sonra güncel kullanım alanlarını, trendleri ve eksiklerini inceleyeceğiz.

İçindekiler

1. Giriş

Son derece disiplinler arası makine öğrenimi alanı, optimizasyon, bilgisayar bilimi, istatistik ve bilişsel bilim dahil olmak üzere çok sayıda bilimsel ve matematiksel alandan ilham almaktadır. Makine öğreniminde sınıflandırma, belirli bir veri kümesini analiz etmek ve verileri istenen sayıda farklı sınıfa bölen bir model oluşturmak için kullanılan denetimli öğrenmeye yönelik bir yaklaşımdır. Literatürde k-en yakın komşu sınıflandırıcıları, bayes ağları, yapay sinir ağları, karar ağaçları ve DVM gibi çeşitli sınıflandırma teknikleri mevcuttur.(J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua ve diğerleri 2019 s.1) Her ne kadar K-en yakın komşu (KNN) algoritmaları kullanım kolaylıklarından dolayı geniş çapta övülse de, yüksek hesaplama gereksinimleri yüzünden büyük girdi veri kümeleriyle uğraşırken çoğu zaman düşük verimle sonuçlanır. Ayrıca gereksiz parametrelere karşı yüksek hassasiyet gösterirler. Öte yandan, sınıflandırma görevlerinde sıklıkla karar ağaçları (eğitim aşamasında sinir ağlarından daha hızlı olan) kullanılır. Ancak karar ağaçları parametreler arasındaki karmaşık ilişkileri temsil edecek kadar esnek değildir. Ayrıca sinir ağları esnek bir araç olarak birçok farklı uygulamada yaygın olarak kullanılmaktadır. Ancak iyi çalışan bir sinir ağı oluşturmak, mimari, nöron ve katman sayısı, öğrenme algoritması ve veri gösterimi gibi bir dizi faktörün dikkatli bir şekilde düşünülmesini gerektirir. Bu teknikler arasında Destek Vektör Makinesi (DVM), beklenen çözümü optimize etmede en etkili yöntemlerden biri olarak öne çıkmaktadır. Vapnik tarafından tanıtılan DVM, sınıflandırma ve regresyon görevleri için kullanılan çekirdek tabanlı bir makine öğrenme modelidir. Olağanüstü genelleme yeteneği, optimal çözümü ve ayırt edici gücü, son yıllarda veri madenciliği, örüntü tanıma ve makine öğrenimi topluluklarından büyük ilgi görmüştür. DVM, pratik ikili sınıflandırma problemlerini çözmek için diğer denetimli öğrenme yöntemlerinden daha iyi performans gösteren güçlü bir araç olarak geniş çapta benimsenmiştir. Güçlü teorik temelleri ve mükemmel genelleme kapasitesi sayesinde DVM'ler en yaygın kullanılan sınıflandırma yöntemlerinden biri haline gelmiştir.

2. DVM'nin Teorik Arka Planı

Örüntü sınıflandırma modelinin temel amacı eğitim verilerinden en iyi performansı elde etmektir. Geleneksel eğitim yöntemlerinde amaç, her girdi-çıktı çiftini kendi sınıfında doğru şekilde sınıflandırmaktır. Bununla birlikte, eğer bir sınıflandırıcı eğitim verileri üzerinde çok sıkı eğitilmişse, nasıl genelleme yapılacağını öğrenmek yerine verileri ezberlemeye başlayabilir ve bu da sınıflandırıcı genellemesinde bir azalmaya neden olabilir. DVM, Destek Vektör Makineleri anlamına gelir. DVM'nin arkasındaki temel fikir, bir eğitim setindeki farklı sınıfları aralarında en büyük marjı sağlayacak bir yüzey kullanarak bölmektir. Başka bir deyişle DVM'nin amacı modelin genelleme yeteneğini en üst düzeye çıkarmaktır. Bu, genellikle ampirik risk minimizasyon teknikleriyle yapılan, eğitim verilerinin ortalama karesel hatasının en aza indirilmesi yerine, bir modelin genel genelleme hatası üzerindeki sınırın en aza indirilmesine odaklanan SRM ile uyumludur.

2.1 Doğrusal olarak ayrılabilir durum

Bir Destek Vektör Makinesinin (DVM) eğitilmesi bir dizi n örnek gerektirir; burada her örnek, bir giriş vektörü xi ve onunla ilişkili etiket yi'den oluşan bir çifttir. Eğitim seti X şu şekilde temsil edilebilir:

$$(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \ldots, (\mathbf{x}_n, \mathbf{y}_n) \qquad (1)$$

[1]

Burada xi, n'inci örneğin giriş vektörünü temsil eder ve yi, bununla ilişkili etikettir. DVM eğitiminin amacı, bu örnekleri kendi sınıflarına doğru şekilde sınıflandırabilen bir model öğrenmektir. Görselleştirme amacıyla, iki boyutlu bir girdi durumunu ele alalım; burada x ∈ R^2. Veriler doğrusal olarak ayrılabilirse, giriş veri kümesini mükemmel şekilde ayırabilen birçok hiperdüzlem vardır. Şekil 1'de bu ayrımı sağlayan çeşitli karar hiperdüzlemleri gösterilmektedir. Bu ayrımı gerçekleştirebilecek sonsuz sayıda hiperdüzlemin olduğu açıktır. Bununla birlikte, sınıflandırıcının genelleme yeteneği ayırma hiperdüzleminin konumuna bağlıdır; marjı maksimuma çıkaran hiperdüzlem optimaldir. Giriş uzayını ayıran karar sınırı veya hiperdüzlem aşağıdaki denklemle tanımlanabilir:

$$\mathbf{w}^{\mathsf{T}}\mathbf{x} - b = 0,$$

(J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua ve diğerleri, 2019 s. 2)

Burada w hiper düzleme dik olan ağırlık vektörüdür, x giriş vektörüdür ve b bias terimidir.

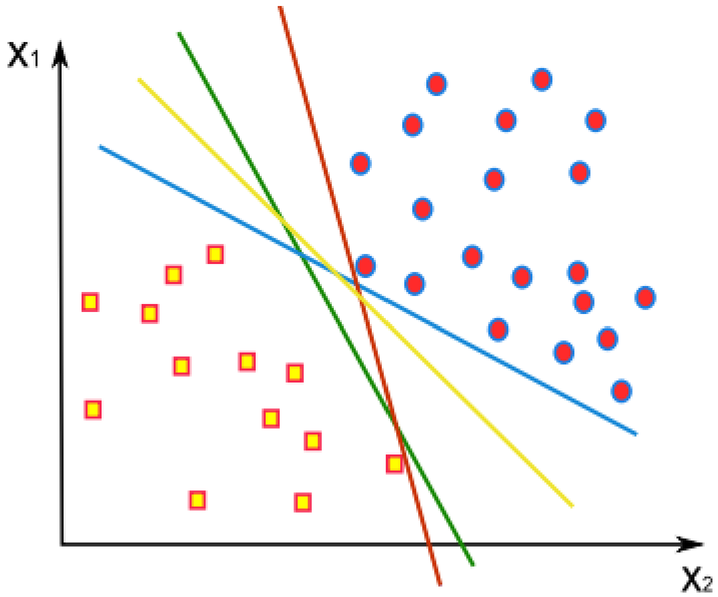


Fig. 1. Separation hyperplanes. [2]

$$\langle w \cdot x^{+} \rangle + b = 1$$
$$\langle w \cdot x^{-} \rangle + b = -1 \quad (2)$$

These can be combined into a set of inequalities:

$$y_i(\langle w \cdot x_i \rangle + b) \geqslant 1 \forall i \quad (3)$$

The geometric margin of x^{+} y x^{-} is

$$\begin{aligned} \gamma_i &= \frac{1}{2}\left(\left\langle \frac{w}{\|w\|} \cdot x^{+} \right\rangle - \left\langle \frac{w}{\|w\|} \cdot x^{-} \right\rangle\right) \\ &= \frac{1}{2\|w\|}\left[\langle w \cdot x^{+} \rangle - \langle w \cdot x^{-} \rangle\right] \\ &= \frac{1}{\|w\|} \end{aligned} \quad (4)$$

[3]

Destek Vektör Makineleri (DVM'ler) bağlamında w, optimal ayırma hiperdüzlemini tanımlar ve b sapma(bias) terimidir. Hiperdüzlem ile ona en yakın eğitim veri noktası arasındaki mesafe, kenar boşluğu olarak bilinir. Genelleme yeteneğinin en üst düzeye çıkarılması, ideal ayırma hiperdüzleminin seçilmesiyle sağlanır. Ağırlık vektörünün normunu en aza indirmek, geometrik marjı optimize etmek için gereklidir. İkinci dereceden programlama problemini çözmek için DVM'leri kullanırken amacımız hem optimal hem de iki paralel hiperdüzlemi (H1 ve H2) tanımlamaktır. İki paralel hiperdüzlem (H1 ve H2) arasında veri olmadığından emin olmak için SVM'nin amacı marjı, yani aralarındaki mesafeyi maksimuma çıkarmaktır. H1 ve H2 arasındaki marj maksimuma çıkarıldığında, bazı veri noktaları H1'in üzerinde veya üzerinde yer alabilirken diğerleri H2'nin üzerinde veya altında yer alabilir. Bu veri noktaları ayırma hiperdüzleminin konumu üzerinde doğrudan etkiye sahip olduğundan bunlara destek vektörleri denir. H1 ve H2 arasına girmeyen geri kalan veri noktaları, hiperdüzlemlerin konumunu veya sınıflandırıcının genelleme kapasitesini etkilemeden değiştirilebilir veya kaldırılabilir. Böylece, bu küçük destek vektörleri seti bir SVM'nin çözümünü tanımlar.Kısaca bu yöntemle veriyi en iyi şekilde ayıran vektörü belirleyebildiğimizi söylüyoruz.

Proposition 1. For the linearly separable case $S = [(x_1, y_1) \cdots (x_l, y_l)]$, if (w, b) is the solution

$$\min_{w,b} \langle w \cdot w \rangle = \|w\|^2$$
$$\text{subject to} : y_i(\langle w \cdot x_i \rangle + b) \geqslant 1 \tag{5}$$

then the maximal margin is given by $\gamma = \frac{1}{\|w\|}$.

[4]

İki ana nedenden dolayı Lagrange formülasyonunu kullanarak SVM optimizasyon problemini ikili forma dönüştürüyoruz. İlk olarak, orijinal problemde belirtilen koşullar, üzerinde çalışılması daha kolay olan Lagrange çarpanları ile değiştirilebilir. İkinci olarak, problemin bu şekilde yeniden formüle edilmesi, eğitim verilerinin yalnızca vektörler arasındaki nokta çarpımları biçiminde görünmesine olanak tanır. Bu, DVM prosedürünün doğrusal olmayan duruma genelleştirilmesini sağlayan çok önemli bir özelliktir. DVM için lagrangianı şu şekilde verilir:

$$L(w, b, \alpha) \equiv \frac{1}{2}\langle w \cdot w \rangle - \sum_{i=1}^{l} \alpha_i [y_i(\langle w \cdot x_i \rangle + b - 1] \tag{6}$$

where α_i are the Lagrange's multipliers.

The dual is found in two steps: first, taking the derivative with respect to w and b

$$\frac{\partial L(w, b, \alpha)}{\partial w} = w - \sum_{i=1}^{l} \alpha_i y_i x_i = 0 \rightarrow w = \sum_{i=1}^{l} \alpha_i y_i x_i \tag{7}$$

$$\frac{\partial L(w, b, \alpha)}{\partial b} = -\sum_{i=1}^{l} \alpha_i y_i = 0 \rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{8}$$

and second, substituting Eqs. (7) and (8) in the original Lagrangian (6)

$$\begin{aligned} L(w, b, \alpha) &= \tfrac{1}{2}\langle w \cdot w \rangle - \sum_{i=1}^{l} \alpha_i [y_i(\langle w \cdot x_i \rangle + b) - 1] \\ &= \tfrac{1}{2}\left\langle \sum_{i=1}^{l} \alpha_i y_i x_i \cdot \sum_{i=1}^{l} \alpha_i y_i x_i \right\rangle - \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j (\langle x_j \cdot x_i \rangle + b) - \sum_{i=1}^{l} \alpha_i \\ &= \tfrac{1}{2} \sum_{i,j=1}^{l} \alpha_i y_i \alpha_j y_j \langle x_i \cdot x_j \rangle - \sum_{i,j=1}^{l} \alpha_i y_i \alpha_j y_j \langle x_j \cdot x_i \rangle - \sum_{i=1}^{l} \alpha_i y_i b + \sum_{i=1}^{l} \alpha_i \\ &= -\tfrac{1}{2} \sum_{i,j=1}^{l} \alpha_i y_i \alpha_j y_j \langle x_i \cdot x_j \rangle + \sum_{i=1}^{l} \alpha_i \end{aligned} \tag{9}$$

[5]

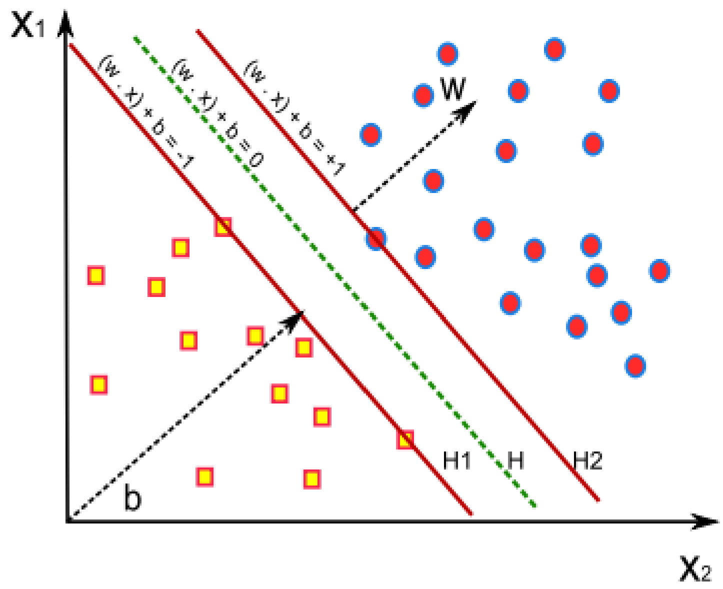


Fig. 2. Optimal classifier.

[6]

Şekil 2, DVM'ler için ikinci dereceden programlama probleminin geometrik bir temsilini gösterir. Bu, optimal ayırıcı H'yi (marjı maksimuma çıkaran hiperdüzlem) ve ayrıca marjı tanımlayan iki paralel hiperdüzlem H1 ve H2'yi gösterir.

Proposition 2. To the linearly separable case $S = [(x_1, y_1) \cdots (x_l, y_l)]$, if α_i^* is the solution to the quadratic optimization problem

$$\begin{aligned} &\max_{\alpha_i} -\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i y_i \alpha_j y_j \langle x_i \cdot x_j \rangle + \sum_{i=1}^{l}\alpha_i \\ &\text{s.t.} : \sum_{i=1}^{l}\alpha_i y_i = 0 \end{aligned} \tag{10}$$

Then $\|w\|^2$ defines the minimum $w^* = \sum_{i=1}^{l}\alpha_i^* y_i x_i$ and the geometric margin $\gamma^* = \frac{1}{\|w^*\|}$ is maximized.

[7]

2.1.1 Karush-Kuhn-Tucker koşulları

Karush-Kuhn-Tucker koşulları (KKT), çok çeşitli optimizasyon problemlerine mümkün olan en iyi çözümü elde etmek için gerekli koşulları sağladığından optimizasyon teorisinin temelini oluşturur.

Theorem 1. *Given an optimization problem with convex domain* $\Omega \subseteq \mathbb{R}^n$,

$$\begin{array}{lll} \text{minimize} & f(w), \quad w \in \Omega & \\ \text{s.t.} & g_i(w) \leqslant 0, i = 1, \ldots, k, & (11) \\ & h_i(w) = 0, i = 1, \ldots, m, & \end{array}$$

with $f \in C^1$ convex, the necessary and sufficient conditions for a normal point w^* to be optimal are the existence of α^*, β^* such that

$$\begin{array}{ll} \frac{\partial L(w^*, \alpha^*, \beta^*)}{\partial w} = 0 & \\ \frac{\partial L(w^*, \alpha^*, \beta^*)}{\partial \beta} = 0 & \\ \alpha_i^* g_i(w^*) = 0, i = 1, \ldots, k, & (12) \\ g_i(w^*) \leqslant 0, i = 1, \ldots, k, & \\ \alpha_i^* \geqslant 0, i = 1, \ldots, k. & \end{array}$$

From KKT conditions if the training set is linearly separable, it is verified that

$$\|w^*\|^2 = \langle w^* \cdot w^* \rangle = (\sum_{i \in sv} \alpha_i^*)^{-\frac{1}{2}} \quad (13)$$

Therefore, the maximum distance of a hyperplane is:

$$\frac{1}{\|w^*\|} = \left(\sum_{i \in sv} \alpha_i^* \right)^{-\frac{1}{2}} \quad (14)$$

[8]

2.2 Yumuşak marj hiperdüzlemleri ve doğrusal olmayan sınıflandırıcı

Daha önce açıkladığım öğrenme problemi, verilerin doğrusal olarak ayrılabilir olduğunu, yani eğitim veri noktalarının kesişmediğini varsayar. Ancak bu kadar mükemmel bir şekilde ayrılabilen veri kümeleri gerçek hayatta nadirdir. Uygulamada, doğrusal bir ayırma hiperdüzleminin, veri noktaları kesiştiğinde bile hala iyi sonuçlar verebileceği durumlar vardır.

Ancak daha önce tartışılan ikinci dereceden programlama çözümleri, veri noktalarının kesiştiği durumlarda doğrudan uygulanamaz. Bunun nedeni, $y_i(\langle w \cdot x_i \rangle + b) \geqslant 1,$ durumu olduğunda tüm i değerleri için gereken koşulların sağlanamamasıdır.(bkz. Şekil 3). Kesişimde bulunan noktalar doğru şekilde sınıflandırılamaz ve yanlış sınıflandırılmış herhangi bir veri noktası xi için, buna karşılık gelen alfai (Lagrange çarpanı) sonsuza doğru yönelecektir. (J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua ve diğerleri, 2019 s. 4)

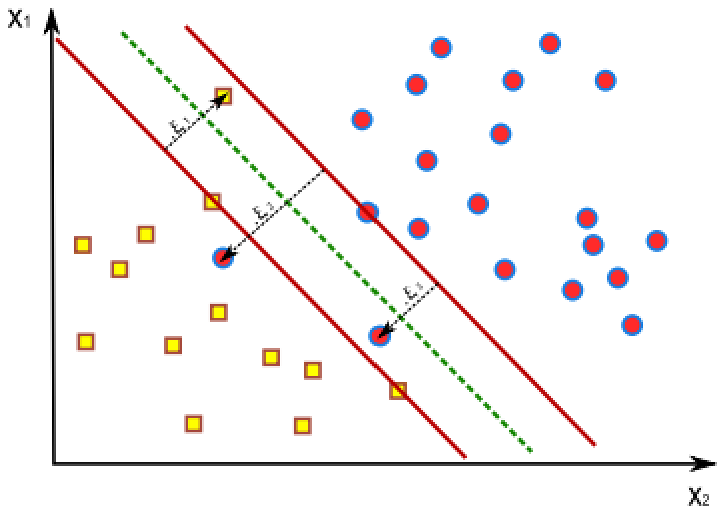


Fig. 3. Soft margin hyperplanes. [9]

Verilerin mükemmel bir şekilde ayrılamadığı durumları karşılamak için algoritma, yumuşak bir kenar boşluğuna izin verecek şekilde değiştirilebilir (bkz. Şekil 4). Bu, denklemde negatif olmayan slack değişkenlerin ni > 0'ın eklenmesini içerir:

$$y_i(\langle w^T \cdot x_i \rangle + b) \geqslant 1 - \xi_i \quad \forall i \qquad (15)$$

[10]

Bu slack değişkenleri kullanarak her zaman uygun bir çözüm bulanbilir. Eğitim verileri xi için, eğer 0< ni < 1 ise, veri noktaları maksimum marjda değildir ancak yine de doğru şekilde sınıflandırılabilir. Bu yumuşak marjın genişliği, marjı maksimuma çıkarmak ve eğitim hatasını minimuma indirmek arasındaki dengeyi belirleyen ceza parametresi C ile kontrol edilebilir.

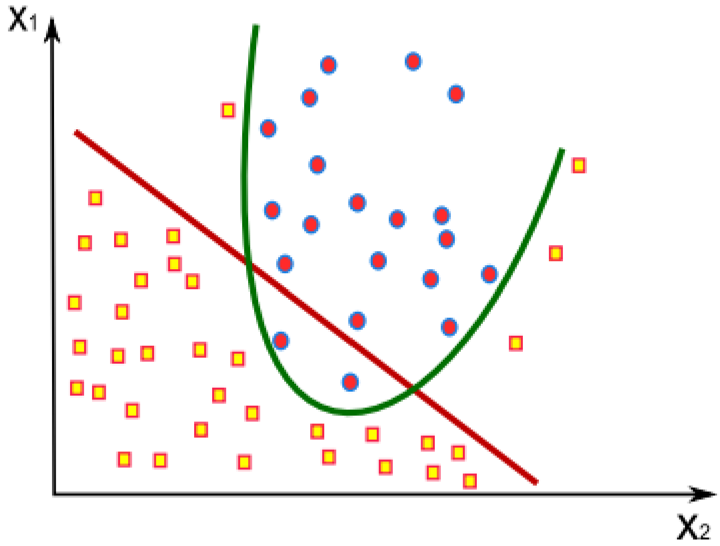


Fig. 4. Non linearly classifier. [11]

Bir öğrenme makinesinin belirli bir dizi işlev için kapasitesi, VC boyutuyla karakterize edilir. İkili sınıflandırma bağlamında 2^h, öğrenen makinenin fonksiyonlarının uygulanması yoluyla akla gelebilecek tüm yollarla iki sınıfa bölünebilecek en fazla nokta sayısını temsil eder.

DVM'ler bağlamında, büyük bir C'nin seçilmesi, az sayıda sınıflandırma hatasıyla sonuçlanır ancak büyük bir ||w||^2 üretebilir. C = 1 değerinin yanlış sınıflandırılmış veri noktalarının olmaması gerektiğini göstermesi dikkat çekicidir. Ancak bu her zaman mümkün olmayabilir çünkü sorun yalnızca bazı C < 1 değerleri için çözülebilir.

$$\begin{gathered}\min_{w,b,\xi_i} \langle w \cdot w \rangle + C\sum_{i=1}^{l} \xi_i^2 \\ \text{s.t.}: y_i(\langle w \cdot x_i \rangle + b) \geqslant 1 - \xi_i \\ \xi_i \geqslant 0\end{gathered} \qquad (16)$$

$$\langle w \cdot x_i \rangle + b \geqslant +1 - \xi_i, y_i = +1, \xi_i \geqslant 0 \tag{17}$$
$$\langle w \cdot x_i \rangle + b \leqslant -1 + \xi_i, y_i = -1, \xi_i \geqslant 0 \tag{18}$$

If $\xi_i < 0, y_i(\langle w \cdot x_i \rangle + b) \geqslant 1 - \xi_i \geqslant 1$, then, we do not consider the condition $\xi_i < 0$.

For the maximum soft margin with Norm-2 (with the diagonal $\frac{1}{C}\delta_{ij}$) the original Lagrangian is given by:

$$L(w, b, \xi_i, \alpha) = \frac{1}{2}\langle w \cdot w \rangle - \sum_{i=1}^{l} \alpha_i [y_i(\langle w \cdot x_i \rangle + b) - 1 + \xi_i] + \frac{C}{2}\sum_{i=1}^{l} \xi_i^2 \tag{19}$$

The dual is found in two steps: in the same way as in the linearly separable case, first differentiating with respect to w and b, and then replacing it in the original Lagrangian

$$\begin{aligned} &\max_{\alpha_i} -\tfrac{1}{2}\sum_{i,j=1}^{l} \alpha_i y_i \alpha_j y_j [\langle x_i \cdot x_j \rangle + \tfrac{1}{C}\delta_{ij}] + \sum_{i=1}^{l} \alpha_i \\ &\text{s.t.}: \sum_{i=1}^{l} \alpha_i y_i = 0 \end{aligned} \tag{20}$$

The Kuhn-Tucker condition is

$$\alpha_i^* [y_i(\langle w^* \cdot x_i \rangle + b^*) - 1 + \xi_i] = 0 \tag{21}$$

[12]

Uygulamada, yumuşak marjlı SVM'ler için ikinci dereceden optimizasyon problemi, ayrılabilir durum için olana çok benzer; temel fark, slack değişkenler ni'nin eklenmesi nedeniyle Lagrange çarpanlarının ai düzeltilmiş değerleridir. C parametresi, daha küçük bir marja sahip olmak ile eğitim noktalarının yanlış sınıflandırılması arasındaki dengeyi kontrol eden kullanıcı tanımlı bir parametredir.

Uygun bir C'nin seçilmesi tipik olarak çapraz doğrulama teknikleri kullanılarak deneysel olarak yapılır. Çapraz doğrulama, veri kümesini birden fazla alt kümeye bölmeyi, modeli bir alt kümede eğitmeyi ve diğerinde doğrulamayı içerir. Bu süreç, model karmaşıklığı ile genelleme performansı arasında en iyi dengeyi sağlayan C değerinin bulunmasına yardımcı olur.

2.3 Kernerler

Bir DVM'de, modelin görünmeyen verilere genelleme yeteneğini en üst düzeye çıkarmak için en uygun hiperdüzlem seçilir. Bununla birlikte, eğitim verileri doğrusal olarak ayrılabilir olmadığında, hiperdüzlemler sınıflar arasındaki marjı maksimuma çıkaracak şekilde optimal olarak belirlense bile sınıflandırıcı yüksek genelleme yeteneğine sahip olmayabilir. Bu gibi durumlarda orijinal girdi uzayı, "özellik(feature) uzayı" adı verilen daha yüksek boyutlu bir uzaya dönüştürülür.(J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua vd., 2019 s. 4)

Feature uzayı dönüşümü, orijinal girdi alanını örtülü olarak verilerin daha kolay ayrılabileceği daha yüksek boyutlu bir alana eşleyen bir kernel fonksiyonu kullanılarak gerçekleştirilir. Bu, DVM'nin, sınıflar orijinal uzayda doğrusal olarak ayrılabilir olmasa bile, dönüştürülmüş uzayda sınıfları daha büyük bir kenar boşluğuyla ayıran bir hiperdüzlem bulmasına olanak tanır. Yaygın kernel fonksiyonları arasında doğrusal, polinom, radyal temel fonksiyon (RBF) ve sigmoid kerneler bulunur.

The basic idea in designing non-linear SVMs is to transform the input vectors $x \in \mathbb{R}^n$ into vectors $\Phi(x)$ of a highly dimensional feature space [30] F (where Φ represents the mapping: $\mathbb{R}^n \to \mathbb{R}^f$) and solve the problem of linear classification in this feature space

$$x \in \mathbb{R}^n \to \Phi(x) = [\phi_1(x), \phi_2(x), \ldots, \phi_n(x)]^T \in \mathbb{R}^f \tag{22}$$

The set of hypotheses considered will be

$$f(x) = \sum_{i=1}^{l} w_i \phi_i(x) + b \tag{23}$$

[13]

Öğrenme prosedürü iki adımdan oluşur: ilk olarak, doğrusal olmayan bir haritalama, verileri özellik alanı F'ye dönüştürür ve ardından bu özellik alanındaki verileri sınıflandırmak için bir Destek Vektör Makinesi (DVM) kullanılır. Doğrusal öğrenen makinelerin önemli bir özelliği ikili biçimde temsil edilebilmeleridir. Bu, karar kuralının eğitim veri noktalarının doğrusal bir kombinasyonu olarak ifade edilebileceği anlamına gelir. Sonuç olarak karar kuralı, giriş veri noktaları arasındaki nokta çarpımları kullanılarak değerlendirilebilir.

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i \langle \phi(x_i) \cdot \phi(x) \rangle + b \quad (24)$$

[14]

Orijinal girdi verilerinin bir fonksiyonu olarak özellik uzayında <ϕ(xi).ϕ(x)> çarpımını doğrudan hesaplamak mümkünse, bu bize doğrusal olmayan bir öğrenme makinesi oluşturmadaki iki temel adımı birleştirmemize olanak tanır. Bu doğrudan hesaplama yöntemine çekirdek(kernel) fonksiyonu adı verilmektedir.(J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua ve diğerleri, 2019 s. 4)

$$K(x, z) = \langle \phi(x) \cdot \phi(z) \rangle \quad (25)$$

where ϕ is a mapping of $X$to a feature space F.

The key to the approach is to find a kernel function that can be evaluated efficiently. Once we have such a decision function, the rule can be evaluated

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i K \langle x_i \cdot x_j \rangle + b \quad (26)$$

A kernel function must respect the following properties: for any $x, y, z \in X$ and $\alpha \in R$

1. $x \cdot x = 0$ only if $x = 0$
2. $x \cdot x > 0$ otherwise
3. $x \cdot y = y \cdot x$
4. $(\alpha x \cdot y) = \alpha(x \cdot y)$
5. $(z + x) \cdot y = (z \cdot y) + (x \cdot y)$

[15]

DVM'lerde kullanılan çekirdek fonksiyonları Mercer teoreminin verdiği koşulu karşılamalıdır. En sık kullanılan çekirde(kernel) fonksiyonlarından bazıları şunlardır:

1. **Linear kernel:** $K(x_i, x_j) = (x_i \cdot x_j)$;
2. **Polynomial kernel:** $K(x_i, x_j) = (x_i \cdot x_j + 1)^p$;
3. **Gaussian kernel:** $K(x_i, x_j) = e^{\frac{-\|x_i - x_j\|^2}{2\sigma^2}}$;
4. **RBF kernel:** $K(x_i, x_j) = e^{-\gamma(x_i - x_j)^2}$;
5. **Sigmoid kernel:** $K(x_i, x_j) = tanh(\eta x_i \cdot x_j + \nu)$;

[16]

Bu çekirdek fonksiyonları sayesinde verileri o alana açıkça eşlemeden özellik uzayındaki nokta çarpımını hesaplayabilmemiz hesaplama açısından avantajlıdır.

DVM'nin gerçek dünya uygulamalarındaki performansı, çekirdek seçiminden büyük ölçüde etkilenebilir. Çeşitli uygulamalarda iyi çalıştığı için Gaussian RBF çekirdeği en yaygın kullanılanıdır. Bununla birlikte, bir çekirdeğin seçimi, verinin belirli özelliklerine bağlıdır ve tatmin edici sonuçlara ulaşmak için, seçilen çekirdek için ideal parametrelerin belirlenmesi genellikle zorunludur.

Normalde teorik kısmı tamamlayıp DVM'ni daha iyi anlayabilmek için çekirdek fonksiyonunu Mercers teoremi ile açıp, doğrusal olarak ayrılamayan duruma daha teorik olarak bakmamız gerekir. Ancak inceleme için DVM'nin arkasındaki matematiği yeterince açıkladığımız için bir sonraki bölümde güçlü, zayıf yönlerine, trendlere ve kullanım alanlarına bakacağız. Son olarak aşağıdaki tablo dört ana kernelin özetini göstermektedir.

Table 1
Popular kernels for SVM.

Kernel name	Expression	Parameters	Characteristics	Some applications
Polynomial	$K(x_i, x_j) = (< x_i, x_j > +1)^r$	$r \in \mathbb{Z}^+$	This kernel allows to map the input space into a higher dimensional space that is a combination of products of polynomials. Despite its high computational load, this kernel is frequently applied to data that has been normalized (norm L_2) [39].	Fault diagnosis of centrifugal pumps [40], natural language processing [41,42
Gaussian radial basis function (RBF)	$K(x_i, x_j) = \exp^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$	σ	This kernel is one of the most widely used in applications, it can be considered a general-purpose translation-invariant kernel. Other related functions are the exponential kernel and Laplacian kernel. Parameter σ must be carefully chosen.	Electric power load forecasting [43], hyperspectral/image classification [44][45][46][47][48], clustering (One-class SVM) [49], bankruptcy prediction [50], classification of electroencephalography signals [51], bioemetric identification [52], health applications [53], intrusion detection [54], stream flow predictions [55]
Linear	$K(x_i, x_j) = < x_i, x_j > +1 = x_i^T x_j + 1$	None	This is the simplest kernel function. It represents the non-kernelized version of SVM. Datasets with many features usually become linearly separable problems. Therefore, the choice of this kernel can be a good option in these cases.	Stock prediction [56], malware detection [57]
Hiperbolic tangent	$K(x_i, x_j) = tanh(< x_i, x_j > \beta + b)$	β, b	This kernel is also known as sigmoid kernel, it is also used as activation function in neural networks. β can be seen as a scaling parameter of the product $x_i^T x_j$, and b a shift. These parameters affect considerably the performance of SVM. [58] showed that using $\beta > 0$ and $b < 0$ guarantees this kernel be conditional positive definite.	Audio classification [59]

[17]

3.DVM'nin Zayıf ve Güçlü Yönleri

J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua ve diğerleri, (2019), genelleştirme kapasitelerine ve birçok avantajlarına rağmen Destek Vektör Makinelerinin (DVM'ler) bazı zayıf yönlerini vurguladılar. Bu zayıf yönler arasında parametrelerin seçimi, büyük veri kümelerinde sınıflandırıcının eğitim süresini etkileyen algoritmik karmaşıklık, çok sınıflı problemler için optimal sınıflandırıcıların geliştirilmesi ve DVM'lerin dengesiz veri kümelerindeki performansı yer almaktadır.

Destek Vektör Makineleri (DVM), özellikle büyük veri kümeleriyle çalışırken yüksek hesaplama yükü nedeniyle öncelikli olarak kusurludur. Bunun ana nedeni, büyük veri kümelerindeki DVM eğitim sürelerinin yavaş olmasıdır çünkü eğitim çekirdeği matrisi, veri kümesi boyutuyla birlikte karesel olarak büyür. Güçlü teorik temellere ve iyi genelleme yeteneklerine rağmen, DVM büyük veri kümelerinin sınıflandırılması için iyi bir seçim değildir. Bir ayırma hiperdüzlemi bulmak için, SVM eğitimi tipik olarak n x n yoğunluk matrisini içeren ikinci dereceden bir programlama (QP) problemi olarak formüle edilir; burada n, veri kümesindeki nokta sayısıdır. Sonuç olarak DVM'nin eğitim karmaşıklığı büyük ölçüde veri kümesinin boyutuna bağlıdır ve bu da büyük veri kümeleri için önemli hesaplama süresi ve bellek gereksinimleriyle sonuçlanır. DVM'ler, özellikle büyük veri kümeleri söz konusu olduğunda hesaplama açısından yoğun olabilir. Veri seçim yöntemleri, ayırma hiperdüzleminin (destek vektörleri) tanımlanması için çok önemli olan örneklere odaklanarak veri kümesi boyutunu azaltmayı amaçlar. Bu yöntemler, basit rastgele örneklemeyi ve yanlış sınıflandırmalara dayalı olarak örnek olasılıklarını güncelleyen daha karmaşık yaklaşımları içerir. Mesafeye dayalı yöntemler, Öklid ve Mahalanobis mesafeleri gibi ölçümlerle örnekler ile hiperdüzlem arasındaki mesafeyi ölçer. Kümeleme yöntemleri ve istatistik tabanlı yaklaşımlar da DVM'lerin büyük veri kümeleri üzerindeki hesaplama yükünü azaltmak için kullanılır, ancak gürültülü verileri işleme veya DVM'lerin seyrek özelliğini koruma konusunda sınırlamaları olabilir. Ayrıştırma yöntemleri, Destek Vektör Makinesini (DVM) hızlandırmayı amaçlamaktadır. Kuadratik Programlama (QP) probleminin aktif kısıtlamalarına odaklanarak eğitim. Parçalama ve Sıralı Minimal

Optimizasyon (SMO) gibi bu yöntemler, daha küçük QP sorunlarını yinelemeli olarak çözer. Paralel optimizasyon ve genetik programlama gibi diğer yaklaşımlar da büyük veri kümelerinin yönetilmesine yardımcı olur. En Küçük Kareler DVM (EKK-DVM) gibi varyantlar, doğruluk için hızdan ödün vermek üzere QP problem formülasyonunu değiştirir.

Ancak tüm bu yöntemler bir nevi değiş tokuş gibidir. Sonuç olarak SVM'nin büyük veri setlerindeki yetersizliğinin diğer yöntemlere göre en büyük dezavantajı olduğunu söyleyebiliriz. Ayrıca, dengesiz veri setlerinde azınlık sınıflarına ait nesnelerin doğru şekilde sınıflandırılması zor bir problem olabilir. Dengesiz verilerle eğitilmiş bir DVM için ideal ayırma hiperdüzleminin elde edilmesi zor olduğundan, bu çarpık veri kümeleri için destek vektör makineleri gibi standart sınıflandırma teknikleri zayıf performans gösterir. Vapnik, regresyon ve sınıflandırma uygulamaları için çekirdek tabanlı bir makine öğrenme modeli olarak Destek Vektör Makinelerini (DVM'ler) tanıttı. Son yıllarda teorisyenler ve uygulayıcılar, ayrım yapma ve genelleme kapasiteleri nedeniyle bunlara daha fazla ilgi duymaya başlamışlardır. DVM'ler sağlam teorik temellere sahiptir ve pratik uygulamalarda genellikle yüksek sınıflandırma doğruluğu gösterir. Bununla birlikte, son deneylere göre, özellikle çoğunluk ve azınlık sınıfı arasındaki oran yüksek olduğunda, dengesiz veri kümelerine uygulandığında DVM performansı önemli ölçüde düşüyor. Dengesiz veri kümelerine uygulandığında DVM'ler, azınlık sınıfının lehine çarpık marjlar üretme dezavantajına sahiptir. Son on yılda, makine öğrenimi topluluğu dengesiz veri kümeleri için sınıflandırma üzerinde daha çok çalışıyor. Sınıflandırıcı performansını artırmak amacıyla dengesiz veri setleri için çeşitli teknikler kullanıyorlar. Bu teknikler tipik olarak iki gruba ayrılır: iç ve dış teknikler. Eğitim veri setlerinin harici yöntemler kullanılarak ön işlenmesi, bunların dengeli olmasını sağlar. Dahili teknikler, öğrenme algoritmalarının eşit olmayan sınıf dağılımına duyarlılığını azaltmak için değiştirerek çalışır. Başka bir deyişle, iç yaklaşımlar yanlış sınıflandırmayla ilgili maliyetleri hesaba katar ve bu maliyetleri modele dahil eder; dışsal yöntemler ise veri setlerini dengelemek amacıyla her sınıf için örnek miktarını dikkate alır. Geniş arama alanı nedeniyle genetik algoritma kullanmadan bu alandaki performansı artırmaya yardımcı olacak örnekleri rastgele seçmek neredeyse imkansızdır. Azınlık sınıfında yeni örnekler oluşturmak DVM sınıflandırmasındaki performansı artırabilirken, bu süreç veri setine gürültü(noise) katabilir.

Tüm bunlara rağmen Destek Vektör Makineleri (DVM'ler), çeşitli makine öğrenimi görevlerindeki güçlü performanslarıyla geniş çapta tanınmaktadır. En önemli güçlü yönlerinden biri, yüksek boyutlu verileri etkili bir şekilde işleme yeteneklerinde yatmaktadır. DVM'ler, sınıflar arasındaki marjı maksimuma çıkarırken, özellik uzayındaki farklı sınıfları en iyi şekilde ayıran hiperdüzlemi bulmayı amaçlar. Bu marj maksimizasyonu, görünmeyen verileri iyi bir şekilde genelleştiren bir modele yol açarak DVM'leri diğer algoritmalara kıyasla aşırı uyum sağlamaya daha az eğilimli hale getirir. Ek olarak DVM'ler, giriş verilerini sınıfların ayrılabilir hale geldiği daha yüksek boyutlu bir alana eşlemek için çekirdek işlevlerini kullanarak doğrusal olmayan karar sınırlarını işleyebilir. Bu esneklik, DVM'lerin verilerdeki karmaşık modelleri yakalamasına olanak tanıyarak onları metin kategorizasyonu, görüntü tanıma ve biyoinformatik dahil olmak üzere çok çeşitli uygulamalar için uygun hale getirir. Hesaplama karmaşıklığına rağmen DVM'ler, verilerin iyi yapılandırılmış olduğu ve hedefin yüksek sınıflandırma doğruluğu elde etmek olduğu senaryolarda oldukça etkilidir.

4.DVM’nin uygulamaları ve gerçek dünya problemleri

Destek Vektör Makineleri (DVM'ler), sınıflandırma ve regresyon görevleri için kullanılan güçlü makine öğrenme modelleridir. Özellikle yüksek boyutlu uzaylarda ve boyutların sayısı örnek sayısından fazla olduğunda etkilidirler. DVM'ler, giriş verilerindeki farklı sınıfları en iyi şekilde ayıran hiperdüzlemi bularak çalışır.

DVM'lerin temel avantajı, sınıflar arasındaki marjı en üst düzeye çıkaran ve görünmeyen verilerin iyi bir şekilde genelleştirilmesine yardımcı olan en uygun hiperdüzlemi bulma yeteneklerinde yatmaktadır. Bu marj maksimizasyonu, hiper düzleme en yakın olan, destek vektörleri olarak bilinen eğitim örneklerinin bir alt kümesinin tanımlanmasıyla elde edilir. Ancak DVM'ler, özellikle büyük veri kümeleriyle uğraşırken hesaplama açısından pahalı olabilir. DVM'lerin hesaplama karmaşıklığı neredeyse kübiktir; bu, veri kümesinin boyutuyla birlikte eğitim süresinin de hızla arttığı anlamına gelir. Bu zorluğun üstesinden gelmek için çeşitli yaklaşımlar önerilmiştir:

1. Veri Azaltma: DVM çözümleri genellikle küçük bir veri noktası alt kümesine (destek vektörleri) dayanır. Destek vektörü olma olasılığı daha düşük olan veri noktalarının ortadan kaldırılmasıyla eğitim süresi azaltılabilir.

2. Parçalama: Bu yaklaşım, DVM optimizasyon problemini daha küçük, daha yönetilebilir parçalara ayırır. Her parça yinelemeli olarak çözülür ve süreç, tüm veri noktaları dikkate alınana kadar devam eder. Bu, problemin karmaşıklığını azaltır ve eğitimi hızlandırır.

3. Ayrıştırma Yöntemleri: Bu yöntemler, geri kalan değişkenleri sabit tutarken, her yinelemede çalışan veya aktif küme olarak bilinen bir değişken alt kümesini optimize eder. Bu, bellek gereksinimlerini azaltır ve verimliliği artırır, ancak aktif kümenin dikkatli seçilmesi çok önemlidir.

4. Sıralı Minimal Optimizasyon (SMO): SMO, her yinelemede yalnızca iki noktayı optimize eden, büyük veri kümeleri için hesaplama açısından verimli hale getiren, DVM'lerin eğitimi için popüler bir algoritmadır. Çekirdek hesaplamalarına harcanan süreyi azaltır ve daha hızlı yakınsamaya yol açar.

5. Buluşsal Yöntemi Daraltma ve Çalışma Seti Seçimi: Bu teknikler, gradyan vektörünü güncellemek için gereken çekirdek değerlerinin sayısını azaltarak ve optimizasyon için daha hızlı yakınsamaya yol açan bir başlangıç değişken seti seçerek optimizasyon sürecini daha da hızlandırır.

Özetle, DVM'ler güçlü teorik temellere sahip güçlü modellerdir. Özellikle büyük veri kümeleri için hesaplama açısından pahalı ve zorlu olabilseler de, eğitim sürelerini ve verimliliklerini artırmak için çeşitli teknikler kullanılabilir ve bu da onları çok çeşitli gerçek dünya uygulamaları için uygun hale getirir.Destek Vektör Makineleri (DVM'ler), özellikle metin ve hiper metin kategorizasyonu olmak üzere çeşitli gerçek dünya problemlerinde yaygın uygulamalar bulmuştur. Metin sınıflandırması, çeşitli uygulamalarda güvenilir performans sergileyen Destek Vektör Makineleri (DVM'ler) tarafından başarıyla gerçekleştiriliyor. Yöntemlerden biri, sınıflandırma aşamasında DVM'lerin Öklid mesafesiyle ve eğitim aşamasında DVM yaklaşımıyla eğitilmesini içerir. Eğitim sırasında her kategori için destek vektörleri bulunur ve sınıflandırma kararlarını vermek için yeni veri noktasına en yakın destek vektörleri kategorisi kullanılır. DVM'ler, Çin belge sınıflandırmasındaki diğer makine öğrenimi teknikleriyle birlikte değerlendirildiğinde, orta büyüklükteki iyi organize edilmiş örneklerden öğrenmede mükemmel verimlilik sergilemiştir. Bu örnekler, metin sınıflandırma görevlerindeki çeşitli zorlukların üstesinden gelmede DVM'lerin kullanışlılığını ve uyarlanabilirliğini vurgulayarak metin sınıflandırmanın çeşitli yönlerindeki etkinliğini ortaya koymaktadır. DVM'ler metin ve görüntü sınıflandırmasında yaygın olarak kullanılmaktadır. Metin kategorizasyonunda, belirli bir kategori için destek vektörlerinin belirlenmesinde etkilidirler ve sınıflandırma kararlarına yol açarlar. DVM'ler, özellikle iyi organize edilmiş örneklerle Çin belge sınıflandırmasında öne çıkıyor. Metin sınıflandırmasında terim-frekans dönüşümleri genellikle DVM performansını çekirdek seçiminden daha fazla etkiler. DVM'ler, özellikle metin sınıflandırması için toplu mod aktif öğrenmede, veri etiketleme çabalarını azaltmak için aktif öğrenmede kullanılır. Kelimelerin dağılımsal kümelenmesiyle doğrusal DVM'nin kullanılması, sınıflandırma performansından ödün vermeden metin belgelerinin boyutluluğunu azaltır. DVM'ler aynı zamanda çok kısa belge dönem ağırlıklandırma, metin sınıflandırma için aktif öğrenme ve diğer çeşitli metin sınıflandırma görevlerinde de başarıyla uygulanmıştır.DVM’ler Görüntü sınıflandırmasında, böbrek taşı görüntü analizi, yüz ifadesi tanıma ve hiperspektral görüntü sınıflandırması gibi görevler için medyan filtreler ve özellik çıkarma yöntemleri gibi diğer tekniklerin yanı sıra kullanılır. Genel olarak, DVM'ler hem metin hem de görüntü analizi görevlerinde çok yönlü bir araçtır ve çeşitli gerçek dünya sorunlarına etkili çözümler sunar. Destek vektör makineleri (dVM), biyoenformatikte, özellikle protein ve kanser sınıflandırmasında da kullanılmaktadır. DVM, protein sınıflandırmasında yarı denetimli özellik seçimi için kullanılır; Denetimli yöntemlerle karşılaştırıldığında daha yüksek doğruluk elde eder ve daha az işlem süresi gerektirir. DVM, kanser sınıflandırmasında genleri seçmek ve sınıflandırmak ve prostat kanseri gen ekspresyonuna ilişkin verilerdeki önemli özellikleri bulmak için kullanılır. Ayrıca, sınıflandırma doğruluğunu artırmak için DVM ve derin sinir ağları birleştirilir. DVM, el yazısı karakter tanıma alanında Arapça karakterleri ve rakamların yanı sıra Bangla karakterlerindeki hiyerarşik sınıflandırma şemalarını tanımak için kullanılır. DVM'nin diğer uygulamaları arasında çevrimdışı yazar tanımlama ve çevrimiçi el yazısı karakter tanıma hizmetleri yer alır. Sinir ağları, k-NN ve karar ağaçları dahil olmak üzere diğer sınıflandırıcılarla karşılaştırıldığında DVM'nin etkinliği, cesaret verici sonuçlara sahip bir dizi uygulamada gösterilmiştir.

Ayrıca DVM, algılama performansını artırmak amacıyla dijital videolarda yüz algılamaya yönelik bir topluluk ayarında kullanılır. Başka bir yöntemde değişen bakış açıları ve yoğun aydınlatma gibi dış etkenlere karşı dayanıklılık sağlamak için yerelliğe duyarlı DVM kullanılır. Etki alanı uyarlamasına ve çoklu yüz temsiline dayalı çoklu sınıflandırma sistemi, video gözetimindeki sınıf içi farklılıklara karşı sağlamlığı artırır. DVM, protein katlanması ve uzaktan homoloji tespiti alanlarındaki görevleri sınıflandırmak için kullanılır. Örneğin, Golgi protein sınıflandırması ve özellik sıralamasının her ikisi de rastgele ormanla birlikte DVM'yi kullanır. Ek bir çalışmada DVM, hesaplamalı modellemede veri kısıtlamalarını en aza indirmek amacıyla membran protein ayrımcılığı için topluluk tabanlı bir transfer öğrenme modeline uygulanmıştır. DVM, protein yapısı analizi ve yüz algılamadaki zorlu sorunların üstesinden gelme konusunda uyarlanabilirliğini ve verimliliğini gösterir. Ayrıca DVM, verimliliği artırmak amacıyla genelleştirilmiş öngörücü kontrol alanında santrifüj kompresörlere yönelik dalgalanma kontrol stratejilerinde kullanılır. Kompresörün kütle akışı ve tahliye basıncı, en küçük kareler DVM'ye dayanan doğrusal olmayan model tahmin kontrolü kullanılarak tahmin edilir. DVM, bitki türlerinin sınıflandırılması, kredi kartı sahtekarlığının tespiti ve melanomun sınıflandırılması dahil olmak üzere çeşitli alanlardaki karmaşık sınıflandırma problemlerinde kullanılmaktadır. DVM, farklı bitki türlerini sınıflandırmak amacıyla hem kontrollü hem de kontrolsüz ortamlardaki bitkilerin görüntülerini analiz etmek için kullanılır. Dengesiz veri setleri, kredi kartı sahtekarlığının tespitinde DVM için sorunlar yaratmaktadır ve dengesizliği düzeltmek için yeni algoritmaların geliştirilmesini gerektirmektedir. Hesaplamalı maliyetine rağmen DVM, melanom sınıflandırmasında erken teşhis amacıyla dermatoskopik görüntülerin bilgisayarlı görme analizinde kullanılmaktadır.

Genel olarak DVM, karmaşık sınıflandırma problemlerini çözmede çok yönlülüğünü göstermektedir, ancak belirli uygulamalardaki performansını artırmak için daha fazla araştırmaya ihtiyaç vardır.

5. Trendler ve zorluklar

Destek Vektör Makinelerinin (DVM) çok sınıflı problemlere, küçük veri kümelerine, dinamik ortamlara ve seyrek etiketli verilere uygulanmasında birçok zor sorun vardır. Bu zorluklar arasında şunlar yer almaktadır:

1. Birleştirilmiş DVM: DVM'nin çok sınıflı problemlere doğrudan uygulanması, ikili sınıflandırma anlamına gelen geleneksel formülasyonu ile sınırlıdır. Bire karşı ve bire bir taktiklerin yanı sıra parçalı doğrusal olmayan bir sınıflandırma fonksiyonu oluşturan veya sınıfları bir simpleks köşelerine haritalayan daha modern tekniklerin de dahil olduğu bir dizi strateji, DVM'yi genişletmek için ortaya konmuştur. Çok sınıflı sınıflandırma.

2.Multi-task DVM: Multi-task öğrenme (MTÖ), bu ortak öğrenmenin, görevlerin bağımsız olarak öğrenilmesinden daha iyi performans gösterebileceği varsayımıyla, birkaç ilgili görevi aynı anda öğrenmeye çalışır. Mevcut DVM tekniklerinin çoğunluğu, sınıf ilişkilerini gözden kaçırabilecek çoklu çoklu görev ikili problemlerini kullanır.

3. Büyük ölçekli sorunlar: DVM eğitimi, büyük veri kümeleriyle uğraşırken çok fazla işlem gücü gerektirebilen ikinci dereceden programlama (QP) sorunlarının çözülmesini gerektirir. Önemli örneklerin bir alt kümesinin seçildiği yetersiz örnekleme ve destek vektörlerinin sayısını azaltmak için kümeleme veya buluşsal tekniklerin kullanılması, büyük ölçekli sorunları çözmeye yönelik iki tekniktir.

4.On-line SVM: Konsept sapması içeren veri akışları için DVM'lerin kolay ve hızlı bir şekilde güncellenmesi gerekir. Bu sorunların üstesinden gelmek için Temsilci Prototip Alanları (TPA) ve veri akışlarından prototiplerin öğrenilmesi gibi teknikler önerildi.

5.Kernel Seçimi ve Parametre Optimizasyonu: DVM'lerin performansı, çekirdeğin(kernelin) seçimine ve parametrelerinin optimizasyonuna bağlıdır. Hiperparametre optimizasyon teknikleri arasında ızgara araması, sınıf ayrılabilirlik analizi ve küresel optimizasyon modelleri bulunur.

6. Transdüktif DVM (TDVM) ve diğer yarı denetimli teknikler: Kısmen etiketlenmiş veri setlerini daha verimli bir şekilde ele almak için, TDVM gibi yarı denetimli teknikler ve TDVM'nin marj dağılımını maksimuma çıkaran algoritmalar önerilmiştir.

Bu yöntemler DVM performansının iyileştirilmesi konusunda umut verici olsa da, DVM'nin çeşitli uygulamalarda ortaya çıkardığı zorluklara daha verimli çözümler geliştirmek için daha fazla araştırmaya ihtiyaç vardır.

5.1 Derin öğrenme ve DVM karşılaştırması

Derin öğrenme, verilerin çoklu doğrusal olmayan ve yinelemeli dönüşümlerini destekleyen hesaplama mimarilerini kullanarak verilerdeki yüksek seviyeli soyutlamaları modelleme yeteneği nedeniyle popülerlik kazanmıştır. Özellikle denetimsiz eğitimlerle yüksek başarı oranları göstermiş ve sağlık, finans, konuşma tanıma, artırılmış gerçeklik, dijital görüntü işleme, 3D ve video uygulamaları gibi çeşitli alanlarda uygulanmıştır.

Derin öğrenme ve Destek Vektör Makineleri (SVM) arasındaki karşılaştırmalı performans çalışmaları birçok yazar tarafından yürütülmüştür. Aşığıdaki Şekil 5’de bu karşılaştırmaları göstermektedir. Birçok durumda derin öğrenme, özellikle karmaşık veri modellerini içeren görevlerde DVM'den daha iyi performans gösterir. Ancak bazı yazarlar performansı daha da artırmak için derin öğrenmeyi DVM ile birleştirmeyi önerdiler. Bu kombinasyon, evrişimli bir sinir ağı (CNN) tarafından öğrenilen özelliklere dayalı olarak DVM sınıflarından birinin eğitilmesini içerir. Bu, doğrusal DVM çekirdeklerinin hassasiyet kaybı olmadan doğrusal olmayan çekirdeklerle değiştirilmesine olanak tanır.

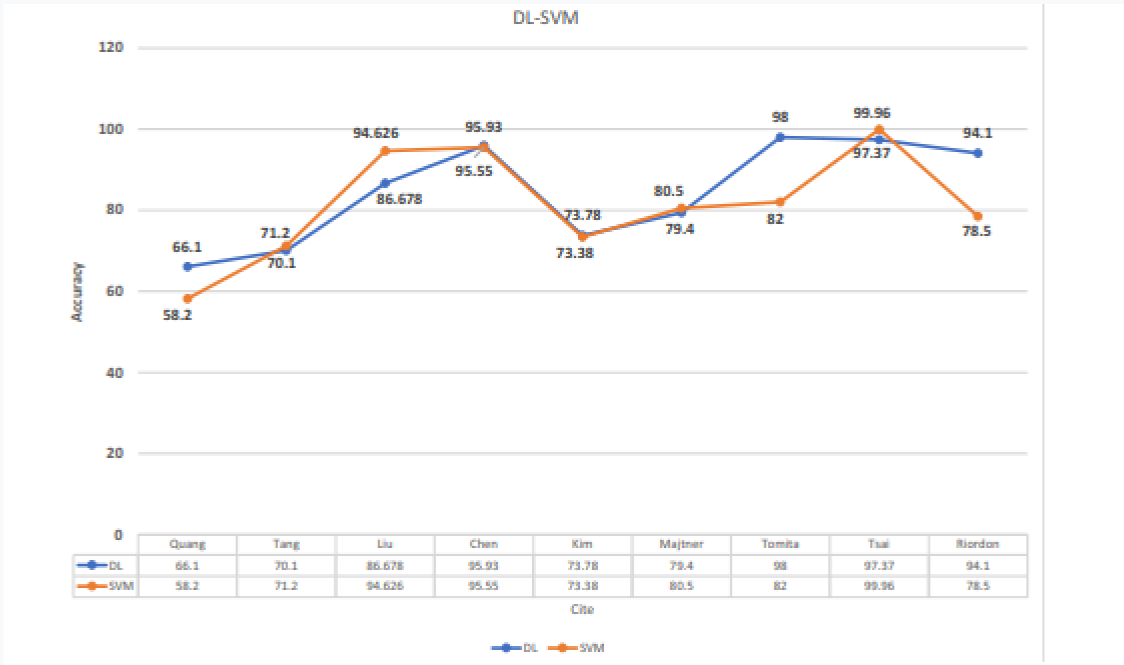
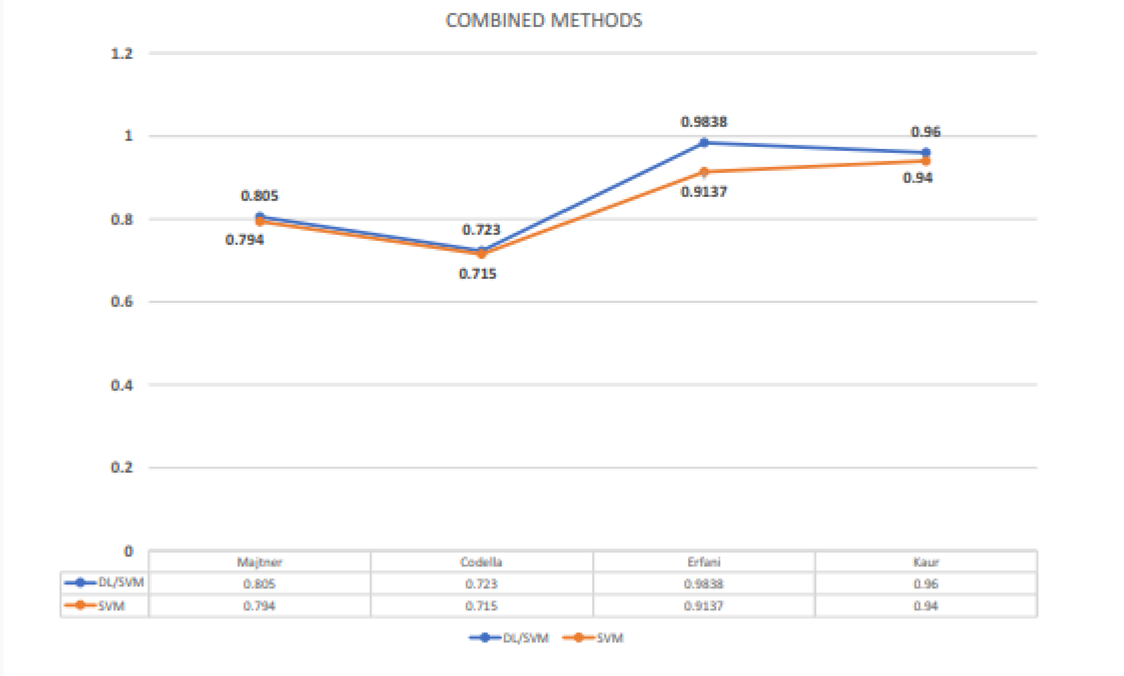


Fig. 5. Performance of SVM vs Deep Learning.

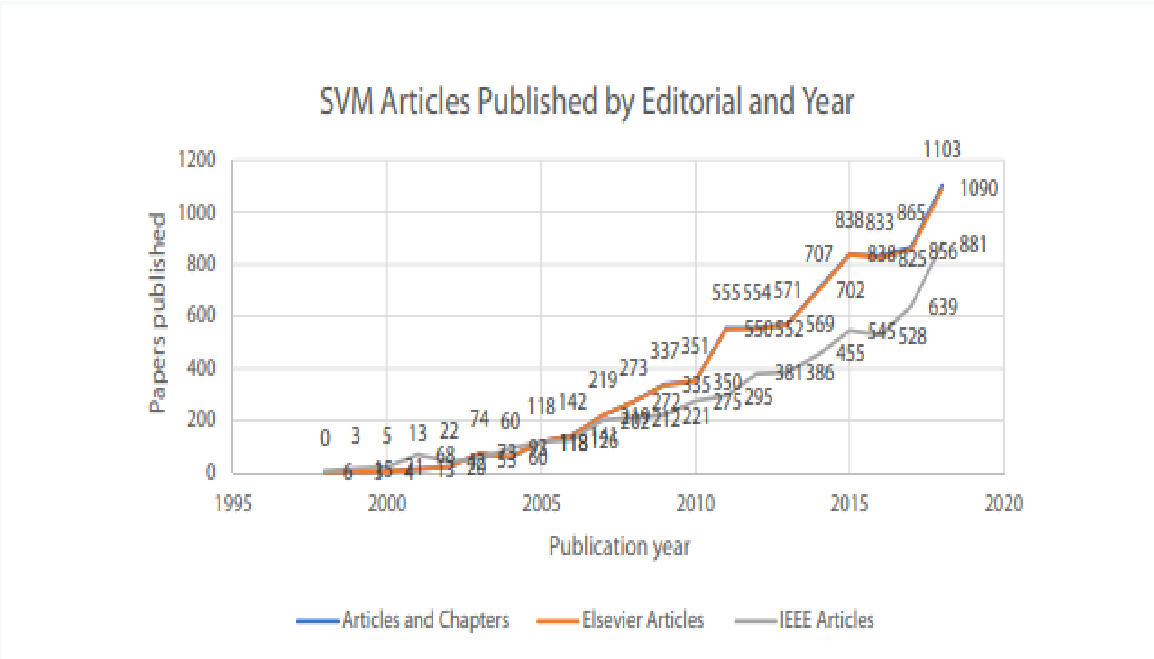


[18]

Genel olarak çalışmalar, DVM ve derin öğrenmenin benzer ortalama performanslara sahip olduğunu ve birlikte kullanıldıklarında farklı uygulamalardaki performansı sinerjik olarak artırabileceğini öne sürüyor.

5.2 DVM'nin Etkisi

Destek Vektör Makinelerinin (DVM) akademik literatürdeki etkisi, 1998'den 2018'e kadar yayınlanan araştırma makalelerine odaklanılarak analiz edildi. DVM ile ilgili araştırma makalelerinin yıllara göre dağılımı aşağıdaki şekilde gösterilmektedir. Analiz, DVM makalelerinin çoğunluğunun, DVM'nin iyi performans gösterdiği normal boyutlu ve dengeli veri kümeleriyle ilgili sorunları ele aldığını ortaya çıkardı. ScienceDirect ve IEEE Xplore arama motorları kullanılarak dergilerden "DVM" terimini içeren 13.000'den fazla yayına ulaşıldı. Büyük ve dengesiz veri kümelerinde SVM uygulamalarını belirlemek için, dergilerde veya kitap bölümlerinde yayınlanmış olmak ve başlıkta, özette veya anahtar kelimelerde ilgili arama terimlerini içeren belirli kriterleri karşılayan yayınlar seçildi.



[19]

6. Sonuç

Destek Vektör Makinelerinin (DVM'ler) güçlü teorik temelleri ve güçlü genelleme yeteneği, bunların çok çeşitli gerçek dünya uygulamalarında yaygın şekilde uygulanmasına yol açmıştır. Metin sınıflandırması, protein kıvrımı(fold) ve uzaktan homoloji tespiti, görüntü sınıflandırması, biyoinformatik (protein ve kanser sınıflandırması), el yazısı karakter tanıma, yüz tespiti, genelleştirilmiş tahmin kontrolü ve diğer alanlar bunların kullanıldığı alanlar arasındadır. Çalışmalar, DVM'lerin sıklıkla alternatif sınıflandırma yöntemlerinden daha iyi performans gösterdiğini göstermiştir.

Bununla birlikte, çok sınıflı veri kümelerinin işlenmesi, parametrelerin seçilmesi, algoritmik karmaşıklık ve dengesiz veri kümelerinin işlenmesi söz konusu olduğunda DVM'lerin sınırlamaları yoktur. Bu dezavantajlara rağmen, DVM'lerin güçlü teorik temelleri ve güçlü genelleme yetenekleri, bunların çok çeşitli gerçek dünya sınıflandırma sorunlarına başarıyla uygulanmasına olanak sağlamıştır. Potansiyel olarak uzun eğitim süreleri nedeniyle, çok büyük veri kümelerine yönelik destek DVM'ler için daha az yaygındır. Ek olarak, DVM doğruluğu dengesiz veri kümelerinden olumsuz etkilenebilir. Bu engellerin üstesinden gelmek için, bu kısıtlamalar karşısında DVM performansını artırmayı amaçlayan özel algoritmalar gibi bir takım stratejiler ortaya konmuştur.

Bu makale, DVM'nin ana dezavantajlarını gidermek için tasarlanmış algoritmaları açıklamakta ve bunların kapsamlı bir tartışmasını sunmaktadır. Bu uygulama zorluklarını DVM ile ele alan araştırmacılar tarafından yürütülen çalışmalara atıfta bulunulmaktadır.

Referanslar

1-19. J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua ve diğerleri, Destek vektör makinesi sınıflandırmasına ilişkin kapsamlı bir araştırma: Uygulamalar, zorluklar ve eğilimler, Nörobilgisayar, https://doi.org/10.1016/j.neucom. 2019.10.118