

Adaptif Sigmoid, Lojistik Sigmoid ve Tanjant Hiperbolik Aktivasyon Fonksiyonların Tam Bağlı ve Konvolüsyonel Sinir Ağlarında Kıyaslanması

Mehmet Murat TURHAN¹, Muhammed Fatih TALU²

¹ Muş Alparslan Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Muş, Türkiye, ORCID: 0000-0003-4497-9102

² İnönü Üniversitesi, Mühendislik Fakültesi, Malatya, Türkiye, ORCID: 0000-0003-1166-8404

Sorumlu Yazar: mm.turhan@alparslan.edu.tr

Geliş tarihi: 23.12.2022

Kabul tarihi:30.12.2022

Özet

Aktivasyon fonksiyonunda serbest parametrelerin bulunması öğrenme aşamasında fonksiyonun adaptif olarak şeklinin değişmesini sağlar. Bu çalışmada 3 serbest parametresi bulunan adaptif sigmoidal aktivasyon fonksiyon incelenmiştir. Bu fonksiyon serbest parametrelerinin 2 farklı şekilde başlatıldığı durum için tam bağlı ve konvolüsyonel sinir ağ mimarili yapılarında standart lojistik sigmoid ve tanjant hiperbolik fonksiyonla kıyaslanmıştır. Mnist ve Fashion Mnist veri kümeleri üzerinde elde edilen deneysel sonuçlar, aktivasyon fonksiyonunda bulunan serbest parametreler sayesinde eğitim veri kümesi üzerinde ağların daha hızlı ve daha başarılı eğitilebildiğini ancak bu aktivasyon fonksiyonlu ağların aşırı uyumlamaya daha yatkın olduğunu göstermektedir. Adaptif aktivasyon fonksiyonunun serbest parametrelerinin 2 farklı şekilde başlatıldığı bütün mimarilerde Mnist veri kümesi üzerinde %100 eğitim başarısı elde edilirken, test veri kümesinde sabit aktivasyon fonksiyonlu mimarilerle daha başarılı sonuçlar elde edilmiştir. Fashion Mnist veri kümesi sınıflandırma problemi için en yüksek eğitim başarısı %97,36 ile adaptif aktivasyon fonksiyon kullanılarak elde edilirken, en yüksek test başarısı %89,87 ile standart lojistik sigmoid fonksiyon kullanılarak elde edilmiştir.

Anahtar Kelimeler: Adaptif Aktivasyon Fonksiyonu, Konvolüsyonel Sinir Ağ, Tam Bağlı Sinir Ağ

Comparison of Adaptive Sigmoid, Logistic Sigmoid and Tangent Hyperbolic Activation Functions in Fully Connected and Convolutional Neural Networks

Abstract

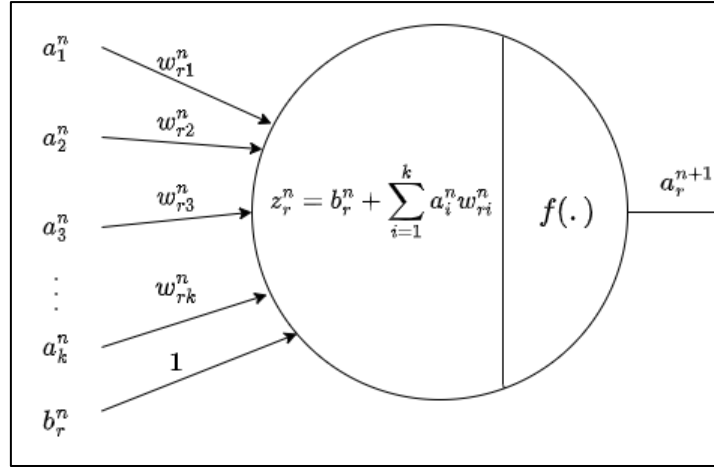
The presence of free parameters in the activation function allows the function to adaptively change its shape during the learning phase. In this study, adaptive sigmoidal activation function with 3 free parameters was investigated. This function is compared with the standard logistic sigmoid and tangent hyperbolic functions in fully connected and convolutional neural network architectures for the case where the free parameters are initialized in two different ways. Experimental results obtained on Mnist and Fashion Mnist datasets show that networks can be trained faster and more successfully on the training dataset thanks to the free parameters in the activation function, but networks with this activation function are more prone to overfitting. While 100% training success was achieved on the Mnist dataset in all architectures where the free parameters of the adaptive activation function were initialized in two different ways, more successful results were obtained with the architectures with fixed activation function in the test dataset. For the Fashion Mnist dataset classification problem, the highest training success was obtained by using the adaptive activation function with 97.36%, while the highest test success was obtained by using the standard logistic sigmoid function with 89.87%.

Keywords: Adaptive Activation Function, Convolutional Neural Network, Fully Connected Neural Network

1. GİRİŞ

Tam bağlı sinir ağ mimarisi temel olarak: Giriş katmanı, gizli katman veya katmanlar ve çıkış katmanından oluşur. Bu mimaride katmanlar arasında yol ağırlıkları olarak adlandırılan bağlantılar vardır. Katmanlarda hesaplama birimleri olarak nöronlar bulunur. Her bir nörona giriş olarak bir önceki katmanda bulunan nöronların çıkışının ilgili yol ağırlıklarıyla çarpımlarının toplanması ve bu toplama yanlılık değerinin eklenmesiyle elde edilen sonuç gelir. Bu sayısal değer nöron için seçilen aktivasyon fonksiyonundan geçirilir ve nöronun çıkış değeri elde edilir. Sinir ağ mimarilerinde kullanılan temel nöron yapısı Şekil 1'de sunulmuştur. Kullanılan aktivasyon fonksiyonu ağ mimarisinin başarısında önemli rol oynar [1]. Ağa öğrenebilme ve karmaşık problemleri çözebilme yeteneği kazandıran en önemli parametre aktivasyon fonksiyonudur. Lineer aktivasyon fonksiyonu tüm ağ mimarisinde kullanılırsa, ağın öğrenebilme yeteneği çok sınırlı olur. Çoklu giriş ve tek çıkışlı derin mimarili bir ağda kullanılan bütün aktivasyon fonksiyonları lineer olursa bu ağ çoklu girişe ve tek çıkışa sahip sadece bir

nörondan oluşan ve yine bu nöronda lineer aktivasyon fonksiyonunun olduğu mimari yapıyla aynı temsiliyete sahip olur. Yani her iki ağ mimarisinde de girişlerin lineer birleşimi çıkışa yansıtılmış olur. Dolayısıyla ağ mimarilerinde karmaşık problemleri çözebilmek için lineer olmayan aktivasyon fonksiyonlarına ihtiyaç duyulur. Literatürde, tek gizli katmana sahip olan tam bağlı sinir ağ mimarisinin temsil gücü araştırılmıştır. Evrensel yaklaşıklık teorisi olarak bilinen ve gizli katmanında sonlu sayıda nöron içeren ağ mimarisinin, keyfi sürekli fonksiyonları belirli bir hata değerine kadar yaklaşıklık gücü kanıtlanmıştır. Gizli katman için seçilen aktivasyon fonksiyonunun sahip olması gereken özelliklerle bu teoremin geçerli olması sağlanır [2-4]. Evrensel yaklaşıklık teorisine göre tek gizli katmana sahip ağın temsil gücü kanıtlanmış olup, ağın uygun parametrelerinin nasıl elde edileceği yani ağın nasıl öğreneceğiyle ilgili çalışma kapsam dışı tutulmuştur.



Şekil 1: Sinir ağında hesaplama birimi olan nöronun yapısı

Ağ mimarilerinde genellikle öğrenilmesi gereken parametreler yol ağırlıkları ve yanlılık değerleridir. Bu tarz problem çözme yaklaşımında aktivasyon fonksiyonları öğrenme sürecinde sabit tutulur ve fonksiyonun şekli değişmez. Literatürde aktivasyon fonksiyonlarının diğer parametreler gibi iterasyonlar boyunca öğrenildiği yani fonksiyonun şeklinin uyarlandığı çalışmalar vardır [5]. Bu çalışmalarda aktivasyon fonksiyonlarına serbest parametreler eklenir ve serbest parametrelerin iterasyonlar boyunca değişip, uygun değere ayarlanmasıyla aktivasyon fonksiyonunun şekli değişir. Böylelikle ağ mimarisi daha esnek bir duruma getirilmiş olur. Chen vd. heterojen işlem birimi olarak adlandırdıkları ve her bir nöronun 3 serbest parametreye sahip olduğu sigmoidal aktivasyon fonksiyonunu önermişlerdir. Genişletilmiş geri yayılım algoritması kullanılarak ağdaki serbest parametrelerin güncellenmesi gerçekleştirilmiştir [6]. Chen vd. fonksiyon yaklaşımına için adaptif ve sabit aktivasyon fonksiyonunun performansını kıyaslamıştır. Tanjant hiperbolik fonksiyona, 2 serbest parametre ekleyerek adaptif aktivasyon fonksiyonu önermişlerdir. Kıyaslama sonucunda daha az nöron kullanarak, adaptif aktivasyon fonksiyonlu ağ yapısı, sabit aktivasyon fonksiyonlu yapıya göre daha başarılı sonuç üretmiştir [7]. Wu vd. çalışmasında ilgili nörona ait çıkış değerini, birden fazla temel fonksiyonun (sigmoid, Gauss, üstel) ağırlıklandırılmasının toplamının oluşturduğu modeli önermişlerdir ve bu ağırlık parametreleri geri yayılım algoritmasıyla öğrenilmiştir [8]. Nakayama vd. her bir nörondaki aktivasyon fonksiyonunu, çok sayıda sigmoidal aktivasyon fonksiyonunun toplamı şeklinde ifade etmiştir. Her bir sigmoidal fonksiyonun öğrenilmesi gereken 4 parametresi vardır [9]. Xu vd. aktivasyon fonksiyonunu 3 temel fonksiyonun toplamı şeklinde ifade etmişlerdir. Bu temel fonksiyonlar: Lojistik, sinüs ve üstel fonksiyonlardır. Bu fonksiyonların her birisinin 2 serbest parametresi vardır. Toplamda her nöronun 6 serbest parametresi bulunmaktadır. Önerdikleri fonksiyon, sabit lojistik fonksiyonla deneysel çalışmalarda kıyaslanmış ve düğüm sayısının az olmasına rağmen önerdikleri fonksiyona sahip olan ağın daha başarılı olduğu gösterilmiştir [10]. Yu vd. lojistik fonksiyona 2 serbest parametre eklemişlerdir. Serbest parametrelili fonksiyon, problemlerin çözümüne erken yakınsayarak daha başarılı sonuç üretmiştir [11]. Chandra vd. standart lojistik aktivasyon fonksiyonuna 1 serbest parametre ekleyerek fonksiyonun türevinin çarpık hale gelmesini sağlamıştır. Böylelikle fonksiyonun bükülme noktası x ekseninde yer değiştirmiş olur. Önerdikleri aktivasyon fonksiyonunu 4 farklı fonksiyon yaklaşımına problemi için standart lojistik fonksiyonla kıyaslamışlardır ve problemlerin çözümünde erken yakınsama sağlamışlardır [12]. Kunc vd. gen ifade çıkarımı için dönüştürücü adaptif aktivasyon fonksiyonu önermişlerdir. Bu fonksiyon, seçilen sabit aktivasyon fonksiyonuna 4 serbest parametre eklenmesiyle oluşturulur. Deneysel çalışmalarında önerdikleri aktivasyon fonksiyonlu mimari sabit aktivasyon fonksiyonlu mimariye göre daha iyi sonuç üretmiştir [13]. Yukarıda

bahsedilen çalışmaların birçoğu tam bağlı sığ sinir ağ mimariler için sabit ve adaptif aktivasyon fonksiyonlarını kıyaslamışlardır. Bu çalışmada ise tam bağlı derin mimariler ve konvolüsyonel mimari için adaptif ve sabit sigmoidal aktivasyon fonksiyonları sınıflandırma problemleri üzerinde kıyaslanmıştır.

2. SİGMOİDAL AKTİVASYON FONKSİYONLARI

Sigmoidal aktivasyon fonksiyonları sinir ağlarında sıklıkla kullanılır. Bu fonksiyonlar S şeklindedir. Tüm reel sayılar için tanımlıdır ve giriş değerini belirli aralığa sıkıştırırlar. Giriş değerlerinin çok büyük ya da çok küçük olmasıyla doygunluğa uğrarlar. Monoton artan, sürekli ve her noktada türevlenebilirler. Fonksiyonun her noktada türevlenebilir olması, geri yayılım algoritmasının [14] uygun bir şekilde çalışabilmesi sağlar. Bu başlık altında standart lojistik, tanjant hiperbolik ve adaptif sigmoidal aktivasyon fonksiyonunun özelliklerinden bahsedilecektir.

2.1 Standart Lojistik Fonksiyon

En temel aktivasyon fonksiyonudur. Yaygın olarak ikili sınıflandırma problemlerinde çıkış katmanında kullanılır. Fonksiyonun 2 asimptotu bulunmaktadır ve giriş değerlerini (0, 1) aralığına sıkıştırır. Giriş değeri artı sonsuza yaklaşırken fonksiyonun çıkışı 1'e, giriş değeri eksi sonsuza yaklaşırken ise fonksiyonun çıkışı 0'a yaklaşır. Fonksiyon denklem 1'de, grafiği ise Şekil 2'de sunulmuştur. Bu fonksiyonun türevine ait denklem ise denklem 2'de, türevine ait grafik ise Şekil 3'te sunulmuştur. Türevi, $x = 0$ iken maksimum değeri olan 0.25'e ulaşır. Fonksiyonun üstel terim içermesi hesaplama maliyeti yönünden olumsuz bir durum iken, fonksiyonun türev hesabında değerinin bulunması hesaplama maliyeti yönünden olumludur.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \quad (2)$$

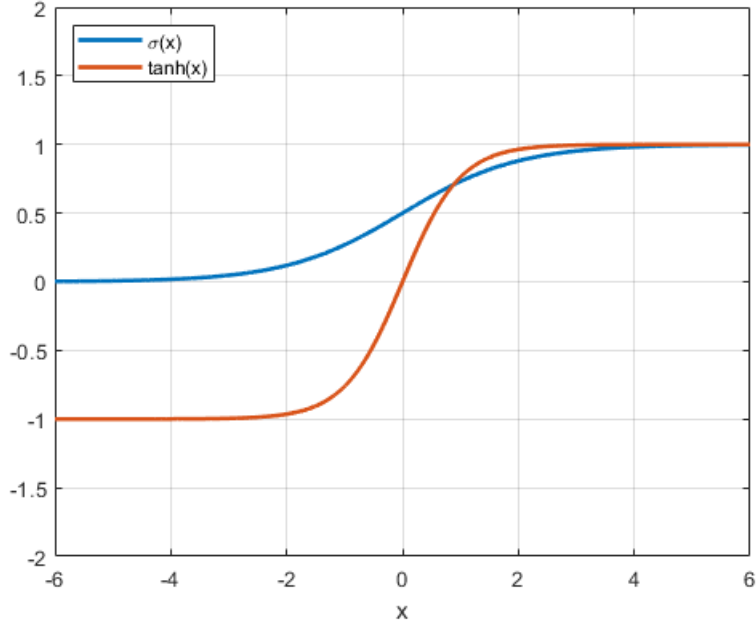
2.2 Tanjant Hiperbolik Fonksiyon

Bir diğer sigmoidal aktivasyon fonksiyonu ise tanjant hiperbolik fonksiyondur. Bu fonksiyon denklem 3'de verilmiştir. Fonksiyonun grafiği ise Şekil 2'de sunulmuştur. Bu fonksiyon giriş değerlerini (-1, 1) aralığına sıkıştırır. Fonksiyonun türevi denklem 4'teki gibidir. Türevinin grafiği ise Şekil 3'te sunulmuştur. Fonksiyonun türevinin maksimum değeri $x = 0$ iken 1 değeridir. Yine fonksiyonun içerisinde üstel terim bulunması hesaplama maliyeti açısından olumsuz iken, fonksiyonun türev hesabında değerinin bulunması maliyet açısından olumludur. Tanjant hiperbolik fonksiyon, standart lojistik sigmoid fonksiyonun ölçeklendirilmiş ve ötelenmiş biçimidir. İki fonksiyon arasındaki ilişki denklem 5'te görülmektedir.

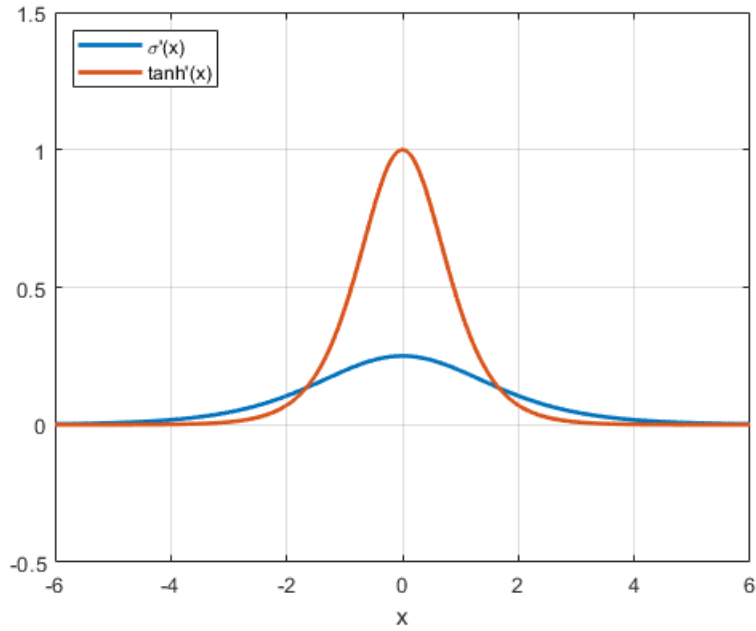
$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

$$\tanh'(x) = 1 - \tanh^2(x) \quad (4)$$

$$\tanh(x) = 2\sigma(2x) - 1 \quad (5)$$



Şekil 2: Lojistik sigmoid ve tanjant hiperbolik aktivasyon fonksiyonlarına ait grafik



Şekil 3: Lojistik sigmoid ve tanjant hiperbolik aktivasyon fonksiyonlarının türevlerine ait grafik

2.3 Adaptif Sigmoidal Fonksiyon

Bu çalışmada kullanılan bir diğer aktivasyon fonksiyonu ise lojistik sigmoid ve tanjant hiperbolik aktivasyon fonksiyonunu genelleştirebilen 3 serbest parametrelili sigmoidal aktivasyon fonksiyondur. Fonksiyon denklem 6'daki gibidir.

$$f(x, U, L, k) = L + \frac{U - L}{1 + e^{-kx}} \quad (6)$$

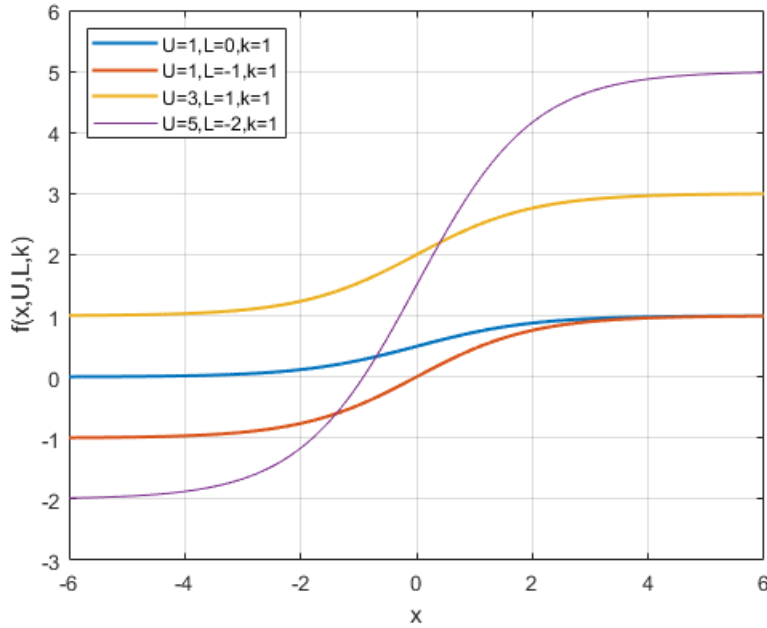
Serbest parametreler: L , U ve k 'dir. L parametresi alt yatay asimptotu, U parametresi üst yatay asimptotu ve k parametresi pozitif olup fonksiyonun sarplığını belirler. k Parametresinin artmasıyla fonksiyon daha sarp hale gelerek üst ve alt asimptot değerlerine erkenden ulaşılmasına etki eder. Fonksiyon kendisine gelen giriş değerini (L , U) arasına sıkıştırır. $L = 0$, $U = 1$ ve $k = 1$ iken fonksiyon standart lojistik sigmoid fonksiyon gibi değer üretir ve $L = -1$, $U = 1$ ve $k = 2$ iken fonksiyon tanjant hiperbolik fonksiyon gibi değer üretir. Fonksiyonun farklı L ve U değerlerine göre çizimleri Şekil 4'te sunulmuştur. Şekil 5'te ise farklı k değerlerine göre fonksiyonun çizimleri sunulmuştur. Fonksiyonun giriş değerine ve serbest parametrelere göre kısmi türevleri sırasıyla denklem 7, 8, 9 ve 10'da verilmiştir. Şekil 6'da farklı L , U ve k parametrelerine göre fonksiyonun türevinin çizimleri sunulmuştur. Fonksiyonun türevinin maksimum değeri $x = 0$ iken $\frac{k(U-L)}{4}$ olur.

$$\frac{\partial f(x, U, L, k)}{\partial x} = k * (f(x, U, L, k) - L) * \left(\frac{U - f(x, U, L, k)}{U - L} \right) \quad (7)$$

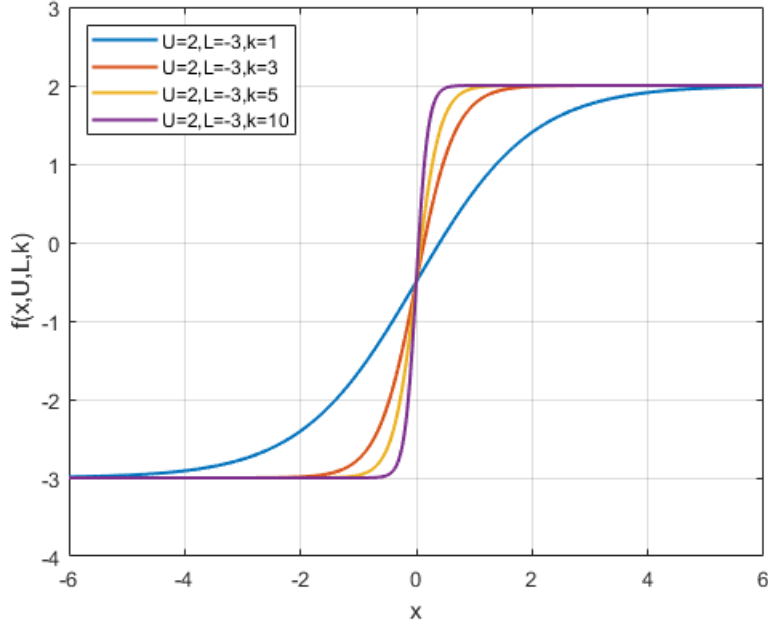
$$\frac{\partial f(x, U, L, k)}{\partial U} = \frac{f(x, U, L, k) - L}{U - L} \quad (8)$$

$$\frac{\partial f(x, U, L, k)}{\partial L} = \frac{U - f(x, U, L, k)}{U - L} \quad (9)$$

$$\frac{\partial f(x, U, L, k)}{\partial k} = x * (f(x, U, L, k) - L) * \left(\frac{U - f(x, U, L, k)}{U - L} \right) \quad (10)$$

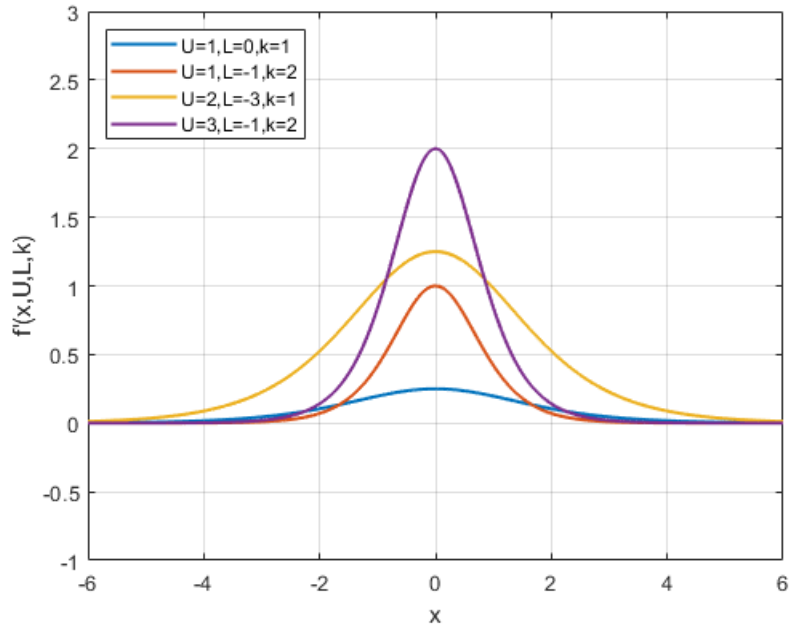


Şekil 4: Adaptif sigmoidal aktivasyon fonksiyonunun farklı U ve L değerlerine göre çizimi



Şekil 5: Adaptif sigmoidal aktivasyon fonksiyonunun farklı k değerlerine göre çizimi

Fonksiyonun serbest parametrelere sahip olmasıyla ağ mimarisindeki yol ağırlıkları ve yanlılık değerleri gibi bu serbest parametreler de iterasyonlar boyunca uygun değere güncellenir. Dolayısıyla fonksiyonun değer üretme aralığı ve sarplığı lojistik sigmoid ve tanjant hiperbolik gibi sabit olmayıp değişkenlik gösterir. Ağ mimarisi oluşturulurken bu tip serbest parametrelili fonksiyonlarla genelde 2 farklı şekilde mimariler oluşturulur. İlki son katman hariç ağdaki bütün nöronların kendine has serbest parametrelerinin olduğu, ikincisi her bir gizli katmandaki nöronların ortak serbest parametrelere sahip olduğu yapılardır. Bu çalışmada ikinci yöntem tercih edilmiştir. H değeri, ağ yapısındaki gizli katman sayısını belirtmek üzere, adaptif sigmoidal fonksiyonlu mimaride sabit aktivasyon fonksiyonlu mimariye göre 3H fazla parametre bulunur. Serbest parametrelerin güncellenmesi, yol ağırlıkları ve yanlılık değerleri gibi geri yayılım algoritmasıyla gerçekleştirilir.



Şekil 6: Adaptif sigmoidal aktivasyon fonksiyonunun farklı U, L ve k değerlerine göre türevlerinin çizimi

Denklem 11, 12 ve 13 sırasıyla U , L ve k parametrelerinin güncelleştirilmesine ait denklemleri belirtir. Bu denklemlerde ortak bulunan $loss$ değeri, ağıın ürettiği çıkış değerleriyle gerçek değerler arasında farkı belirten hata fonksiyonudur. a_i^l değeri, l . gizli katmanın i . nöronunun çıkış değeridir. α , öğrenme katsayısı ve N ise ilgili gizli katmanda bulunan nöron sayısını belirtir.

$$U^l = U^l - \alpha * \sum_{i=1}^N \frac{\partial loss}{\partial a_i^l} * \frac{\partial a_i^l}{\partial U^l} \quad (11)$$

$$L^l = L^l - \alpha * \sum_{i=1}^N \frac{\partial loss}{\partial a_i^l} * \frac{\partial a_i^l}{\partial L^l} \quad (12)$$

$$k^l = k^l - \alpha * \sum_{i=1}^N \frac{\partial loss}{\partial a_i^l} * \frac{\partial a_i^l}{\partial k^l} \quad (13)$$

3. BULGULAR VE TARTIŞMA

Bu bölümde hem tam bağlı mimari hem de konvolüsyonel mimaride standart lojistik sigmoid, tanjant hiperbolik ve adaptif sigmoid aktivasyon fonksiyonları kıyaslanmıştır. Adaptif sigmoid fonksiyonu 2 farklı parametre başlangıcıyla oluşturulmuştur. İlki, parametrelerin standart lojistik sigmoid fonksiyona eşdeğer olacak şekilde başlatıldığı, ikincisi ise parametrelerin tanjant hiperbolik fonksiyona eşdeğer olacak şekilde başlatıldığı durumdur. Dolayısıyla her mimari 4 farklı aktivasyon fonksiyonuyla oluşturulup, kıyaslanmıştır. Deneysel çalışmalarda adaptif sigmoidal aktivasyon fonksiyonlu mimariler için \mathbf{U} , \mathbf{L} ve \mathbf{k} parametreleri iterasyonlar boyunca güncellenirken fonksiyonun monoton artan yapısını korumak için $k > \mathbf{0}$ ve $\mathbf{U} > \mathbf{L}$ sınırlandırması getirilmiştir. Her deneysel çalışmada yol ağırlıkları rastsal olarak başlatıldığı için kıyaslanmanın daha güvenilir olabilmesi amacıyla seçilen aktivasyon fonksiyonlu ilgili mimari uygun öğrenme katsayısıyla 3 kez çalıştırılmıştır. Ortalama eğitim/test başarısı üzerinden kıyaslama işlemi yapılmıştır. Deneysel çalışmalar Core i9-10850K 3.60GHz işlemci, 128GB ram ve RTX 3090 ekran kartına sahip bilgisayarda gerçekleştirilmiştir.

3.1 Tam Bağlı Sinir Ağ Mimarisi için Aktivasyon Fonksiyonlarının Kıyaslanması

Bu deneysel çalışmada aktivasyon fonksiyonları tam bağlı sinir ağ mimarisi için Mnist veri kümesi üzerinde kıyaslanmıştır. Mnist veri kümesi 60.000 eğitim ve 10.000 test verisinden oluşmaktadır [15]. Bu veri kümesi 0-9 arasındaki 10 rakamın el yazısıyla yazılıp dijitalleştirilmesiyle oluşturulmuştur. Her bir resim 28x28 boyutunda ve tek kanaldan oluşmaktadır. Veri kümesi sınıflar arasında dengeli dağılmıştır ve her rakam için 6000 eğitim verisi ve 1000 test verisi vardır. Bu problemin çözümü için her bir katmanda 128 nöron bulunacak şekilde ve gizli katman sayısının $\{2, 3, 4, 5, 6, 7\}$ olduğu derin yapıli mimariler oluşturulmuştur. Adaptif aktivasyon fonksiyonlu mimarilerin aynı katmandaki nöronlarının ortak U , L , ve k parametrelerini kullanmaları sağlanmıştır. Böylelikle adaptif aktivasyon fonksiyonlu mimarilerin her katmanda sabit aktivasyon fonksiyonlu mimarilere göre fazladan 3 parametresi bulunmaktadır. Öğrenme katsayısı her bir mimari için $\{0.1, 0.01, 0.001\}$ olacak şekilde denenmiştir ve bütün mimarilerde 0.1 öğrenme katsayısıyla en başarılı sonuçlar her aktivasyon fonksiyonu için elde edilmiştir. Epok sayısı 50, her bir iterasyonda kullanılan veri sayısı 128 ve optimizasyon yöntemi olarak stokastik gradyan iniş algoritması belirlenmiştir. Son katmanda Softmax aktivasyon fonksiyonu kullanılmıştır. Ön işlem olarak [0-255] aralığında olan her bir piksel değeri [0-1] aralığına normalize edilmiştir. Yol ağırlıkları, seçilen aktivasyon fonksiyonuna uygun olacak şekilde başlatılmıştır [16, 17]. Tablo 1'de her bir aktivasyon fonksiyonuna ait bulunduğu mimari için elde edilen ortalama eğitim ve test başarısı sunulmuştur. Başarı değerlerinin yanındaki \pm değerler standart sapma değerleridir. Adaptif aktivasyon fonksiyonlu bütün mimarilerde eğitim kümesi üzerinde %100 oranında başarı elde edilirken test kümesinde tanjant hiperbolik aktivasyon fonksiyonlu mimariler en başarılı sonuçları üretmiştir. Adaptif aktivasyon fonksiyonlu mimariler aşırı uyumlama problemiyle karşı karşıya kalmıştır. Adaptif aktivasyon fonksiyonunun başlangıç parametre değerleri de başarıyı etkileyen bir faktördür. Bütün mimarilerde test başarısı, adaptif aktivasyon fonksiyonunun tanjant hiperbolik biçimiyle başlatıldığı düzende ($L = -1, U = 1, k = 2$) standart lojistik sigmoid olarak başlatıldığı düzene ($L = 0, U = 1, k = 1$) göre daha iyi sonuç üretmiştir. Tablo 2'de ilgili aktivasyon fonksiyonlu mimarilere ait öğrenilmesi gereken parametre

sayısı bulunmaktadır. Tablodan da görüldüğü gibi adaptif aktivasyon fonksiyonlu mimariler sabit aktivasyon fonksiyonlu mimarilere göre katman sayısı x 3 fazladan parametreye sahiptir.

Tablo 1: Farklı gizli katman sayılı tam bağlı mimariler için Mnist veri kümesi üzerinde aktivasyon fonksiyonlarının eğitim/test başarı değerlerine etkisi

	Aktivasyon Fonksiyonu				Gizli Katman Sayısı
	Lojistik Sigmoid	Adaptif(L = 1, U = 1, k = 1)	Tanjant Hiperbolik	Adaptif(L = -1, U = 1, k = 2)	
Eğitim Başarısı	0.9786±(0.0003)	1.0±(0.0)	0.9998±(0.00005)	1.0±(0.0)	2
Test Başarısı	0.9700±(0.0003)	0.9681±(0.0014)	0.9797±(0.0006)	0.9752±(0.0006)	
Eğitim Başarısı	0.9829±(0.0002)	1.0±(0.0)	0.9999±(0.00001)	1.0±(0.0)	3
Test Başarısı	0.9720±(0.0019)	0.9694±(0.0011)	0.9797±(0.0005)	0.9770±(0.0003)	
Eğitim Başarısı	0.9857±(0.0001)	1.0±(0.0)	1.0±(0.0)	1.0±(0.0)	4
Test Başarısı	0.9726±(0.0018)	0.9700±(0.0014)	0.9798±(0.0003)	0.9772±(0.0012)	
Eğitim Başarısı	0.9870±(0.0001)	1.0±(0.0)	0.9999±(0.000008)	1.0±(0.0)	5
Test Başarısı	0.9706±(0.0011)	0.9713±(0.0013)	0.9798±(0.0009)	0.9782±(0.0003)	
Eğitim Başarısı	0.9861±(0.0008)	1.0±(0.0)	1.0±(0.0)	1.0±(0.0)	6
Test Başarısı	0.9692±(0.0012)	0.9727±(0.0013)	0.9806±(0.0004)	0.9772±(0.0009)	
Eğitim Başarısı	0.9816±(0.0010)	1.0±(0.0)	1.0±(0.0)	1.0±(0.0)	7
Test Başarısı	0.9642±(0.0011)	0.9719±(0.0005)	0.9801±(0.0004)	0.9774±(0.0006)	

Tablo 2: Farklı gizli katman ve ilgili aktivasyon fonksiyonlu tam bağlı mimariler için öğrenilmesi gereken parametre sayısı

	Aktivasyon Fonksiyonu				Gizli Katman Sayısı
	Lojistik Sigmoid	Adaptif(L = 1, U = 1, k = 1)	Tanjant Hiperbolik	Adaptif(L = -1, U = 1, k = 2)	
118282	118288	118282	118288	2	
134794	134803	134794	134803	3	
151306	151318	151306	151318	4	
167818	167833	167818	167833	5	
184330	184348	184330	184348	6	
200842	200863	200842	200863	7	

3.2 Konvolüsyonel Mimari için Aktivasyon Fonksiyonlarının Kıyaslanması

Konvolüsyonel sinir ağları, tam bağlı sinir ağ mimarilerine kıyasla veri kümesinin görüntü olduğu problemlerde daha başarılı sonuç elde ederler [18]. Bu deneysel çalışmada LeNet-5 [15] olarak adlandırılan konvolüsyonel mimaride adaptif ve sabit aktivasyon fonksiyonları, Mnist ve Fashion Mnist veri kümeleri üzerinde kıyaslanmıştır. LeNet-5 mimarisinin işlem aşamaları Tablo 3'te görülmektedir.

Tablo 3: LeNet-5 mimarisinin işlem aşamaları

Konvolüsyon, (5x5), 6 Filtre
Ortalama Havuzlama, (2x2)
Konvolüsyon, (5x5), 16 Filtre
Ortalama Havuzlama, (2x2)
Tam Bağlı Katman, 120
Tam Bağlı Katman, 84
Tam Bağlı Katman, 10

Fashion Mnist veri kümesi de Mnist veri kümesi gibi 10 farklı sınıfa ait resimler içerir [19]. Veri kümesinde bulunan sınıflar: Tişört, pantolon, kazak, kıyafet, ceket, sandalet, gömlek, spor ayakkabı, çanta ve buttur. Bu veri kümesi 60.000 eğitim verisi ve 10.000 test verisinden oluşmaktadır. Her bir resim 28x28 boyutunda ve tek kanaldan oluşmaktadır. Adaptif aktivasyon fonksiyonlu mimarilerin konvolüsyonel katmanlarında her bir özellik haritasının hesaplanması için farklı U, L ve k parametrelerine sahip olabilecek fonksiyonlar üzerinden işlem yapılmasına müsaade edilmiştir. Dolayısıyla her konvolüsyon katmanı için filtre sayısı x 3 ek parametre adaptif aktivasyon fonksiyonlu mimarilerde bulunur. Tam bağlı katmanda ise ilgili katman için her nöronun ortak U, L ve k parametreleri üzerinden işlem yapması sağlanmıştır. Böylelikle tam bağlı her katman için sabit aktivasyon fonksiyonlu mimarilere göre ilave 3 parametre adaptif aktivasyon fonksiyonlu mimaride bulunur. Mnist ve Fashion Mnist sınıflandırma problem çözümü için seçilen aktivasyon fonksiyonlu her bir mimari, öğrenme katsayısı { 1.0, 0.1, 0.01 } olacak şekilde denenmiştir. Standart lojistik sigmoid aktivasyon fonksiyonu, öğrenme katsayısının 1.0 olduğu durumda her iki veri kümesi için başarılı sonuç üretirken diğer aktivasyon fonksiyonları her iki veri kümesi için öğrenme katsayısının 0.1 olduğu durumda başarılı sonuç üretmişlerdir. Ön işlem olarak piksel değerleri her iki veri kümesi için de [0-1] aralığına normalize edilmiştir. Epok sayısı 50, her bir iterasyonda kullanılan veri sayısı 128 ve optimizasyon yöntemi ise stokastik gradyan iniş algoritması olarak belirlenmiştir. Çıkış katmanında Softmax aktivasyon fonksiyonu kullanılmıştır. Yol ağırlıkları, seçilen aktivasyon fonksiyonuna uygun olacak şekilde başlatılmıştır [16, 17]. Tablo 4'te veri kümeleri için ilgili aktivasyon fonksiyonuyla elde edilen deneysel sonuçlar görülmektedir. Mnist veri kümesi üzerinde test başarısı için en yüksek sonuç tanjant hiperbolik aktivasyon fonksiyonuyla elde edilirken, adaptif biçimde olan aktivasyon fonksiyonlu mimariler bu veri kümesinin eğitim görüntülerinin tamamını başarılı tahmin edebilmiştir. Test kümesinde sabit biçimli aktivasyon fonksiyonları daha başarılı sonuç üretmişlerdir. Aktivasyon fonksiyonlarının LeNet-5 mimarisinde Mnist veri kümesi için epoklar boyunca eğitim/test başarılarına etkisi Şekil 7'de görülmektedir. Adaptif aktivasyon fonksiyonlu mimarilerle yaklaşık 20. epoktan sonra eğitim veri kümesi öğrenilmiştir ve test başarısında ilerleme sağlanamamıştır. Bu durum da aşırı uyumlamanın göstergesidir. Benzer sonuçlar Fashion Mnist veri kümesi için de söz konusudur. Bu veri kümesi için test başarısı en yüksek, aktivasyon fonksiyonunun standart lojistik sigmoid olduğu mimariyle elde edilirken eğitim veri kümesinde en başarılı sonuç parametrelerin tanjant hiperbolik biçimiyle başlatıldığı adaptif aktivasyon fonksiyonlu mimariyle elde edilmiştir. Bu aktivasyon fonksiyonlu mimari test veri kümesinde ise en başarısız sonucu üretmiştir. Yine eğitim ve test başarısı arasındaki fark %9,55 ile en çok bu aktivasyon fonksiyonlu mimaride gerçekleşmiştir. Eğitim ve test başarısı arasındaki farkın ikinci olduğu mimari %7,31 ile serbest parametrelerin standart lojistik sigmoid olarak başlatıldığı adaptif mimari olmuştur. Dolayısıyla her iki veri kümesi için de adaptif aktivasyon fonksiyonlu mimariler aşırı uyumlama problemiyle karşılaşmıştır. Şekil 8'de Fashion Mnist veri kümesi üzerinde LeNet-5 mimari için aktivasyon fonksiyonlarının epoklar boyunca eğitim/test başarı grafiklerine etkisi görülmektedir. İlk epoklarda serbest parametrelerinin tanjant hiperbolik olarak başlatıldığı adaptif aktivasyon fonksiyonlu yapı, tanjant hiperbolik aktivasyon fonksiyonlu yapıya göre test veri kümesinde daha başarıyla ilerleyen epoklarda aşırı uyumlama problemiyle karşı karşıya kalmıştır ve tanjant hiperbolik fonksiyonun gerisinde kalmıştır. Aynı durum parametrelerinin standart lojistik sigmoide eş değer olarak başlatıldığı adaptif aktivasyon fonksiyonlu yapıyla, standart lojistik sigmoid aktivasyonlu yapı arasında gözlenmektedir. Tablo 5'te ilgili aktivasyon fonksiyonlu mimarilere ait öğrenilmesi gereken parametre sayısı bulunmaktadır. Adaptif aktivasyon fonksiyonlu LeNet-5 mimarisi sabit aktivasyon fonksiyonlu mimariye göre fazladan %0,162 parametre bulundurmaktadır. Toplam öğrenilmesi gereken parametre sayısının yanında bu değer çok küçük kalmaktadır.

Mnist veri kümesi sınıflandırma problemi için için Bölüm 3.1'deki tam bağlı her mimarinin parametre sayısı, LeNet-5 konvolüsyonel mimarili her yapının parametre sayısına göre fazla olmasına rağmen konvolüsyonel mimarili her yapı bu veri setinde test veri kümesi için daha başarılı sonuç üretmiştir. 6 Katmanlı ve aktivasyon fonksiyonunun tanjant hiperbolik olduğu tam bağlı mimariyle test başarısı en yüksek değeri olan %98,06'ya ulaşırken LeNet-5 mimarili bütün yapılar test başarısında bu değeri geçmiştir. 6 katmanlı tam bağlı mimarinin parametre sayısı konvolüsyonel mimarili LeNet-5 mimarinin parametre sayısının 3 katından fazla olmasına

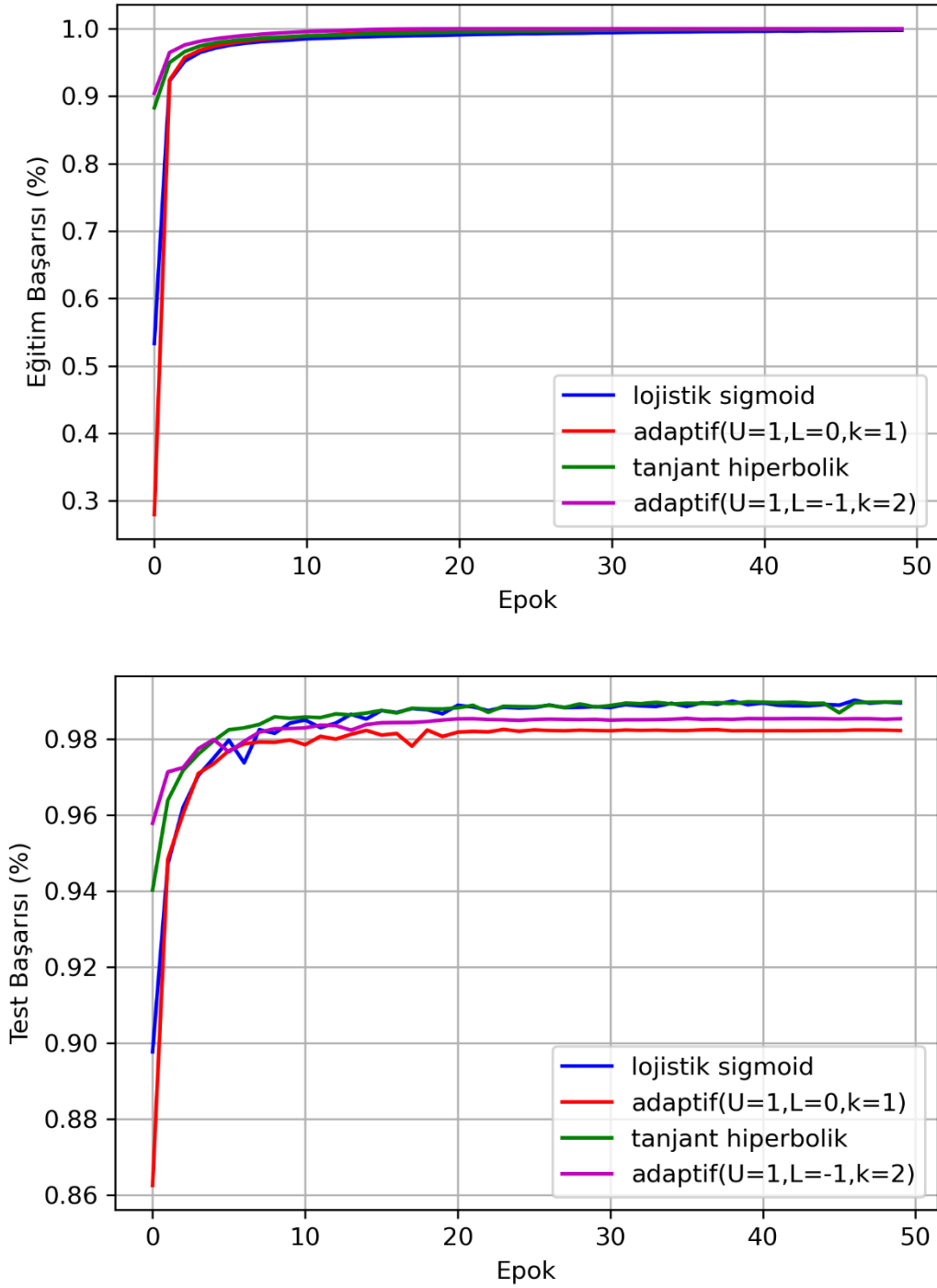
rağmen test başarısında gerisinde kalması, giriş bilgisinin görüntü olduğu sınıflandırma problemlerinde konvolüsyonel mimarilerin ilgili katmanlarında daha iyi özellik çıkartılıp sonraki katmanlara işlenmek üzere aktarılmasıyla açıklanabilir. Adaptif aktivasyon fonksiyonlu mimariler, sabit aktivasyon fonksiyonlu mimarilere göre ek az sayıda parametreye sahip olmalarına rağmen hızlıca eğitim veri kümesini öğrenirken ilerleyen epoklarda sabit yapıli aktivasyon fonksiyonlu mimariler test veri kümesinde adaptif aktivasyon fonksiyonlu mimarileri geçmiştir. Bu da eğitim veri kümesini ezberleme anlamına gelip aşırı öğrenme problemini ortaya çıkartır. Mnist veri kümesinin, Fashion Mnist veri kümesine göre sınıflandırılması daha kolay olduğu görülmektedir. Hem test hem de eğitim kümesi için LeNet-5 mimarili bütün yapılar Mnist sınıflandırma probleminde, Fashion Mnist sınıflandırma probleminde göre daha başarılı sonuç üretmiştir.

Tablo 4: Mnist ve Fashion Mnist veri kümeleri için LeNet-5 mimaride aktivasyon fonksiyonlarının eğitim/test başarı değerlerine etkisi

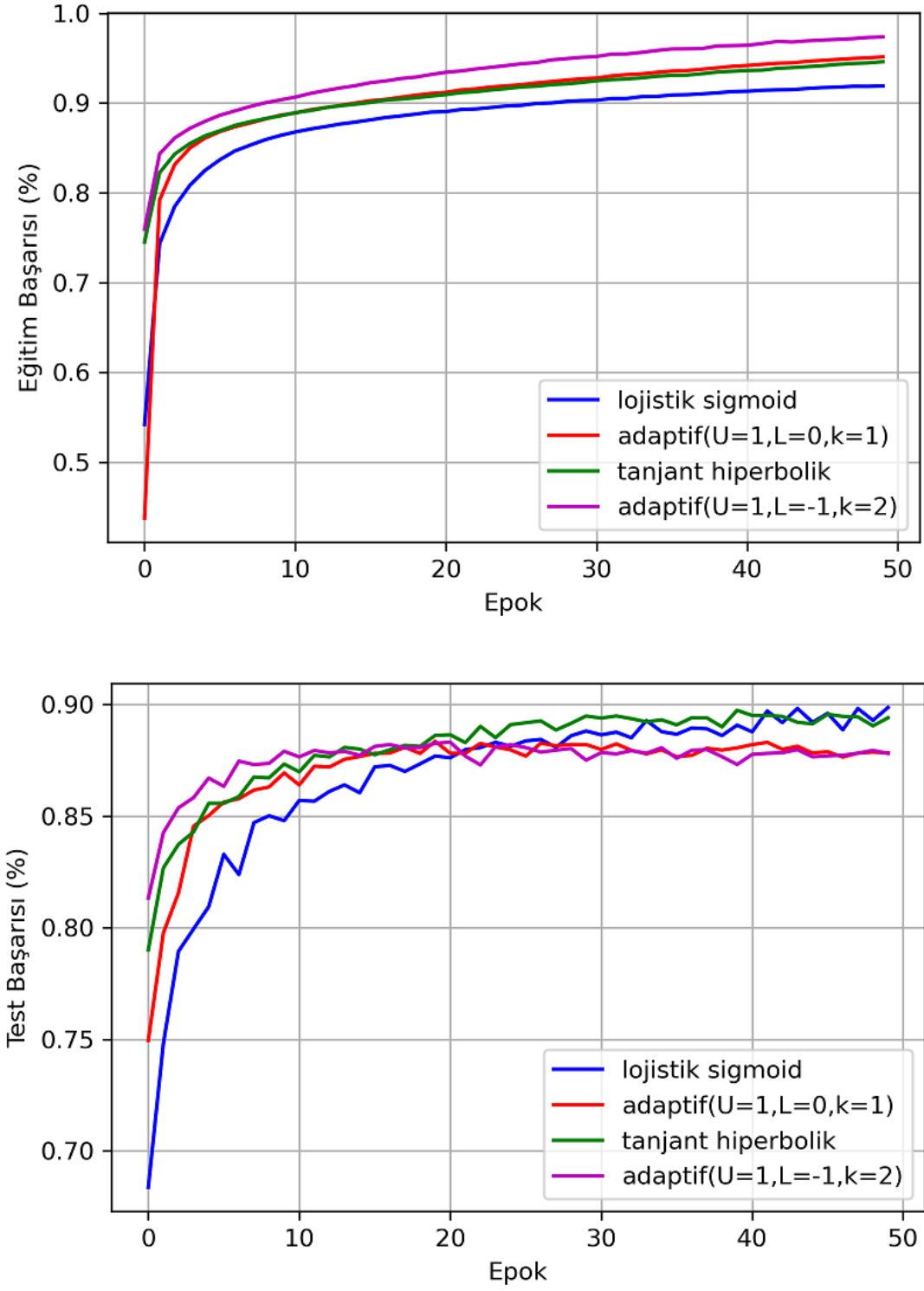
	Aktivasyon Fonksiyonu				Veri Kümesi
	Lojistik Sigmoid	Adaptif (U=1, L=0, k=1)	Tanjant Hiperbolik	Adaptif (U=1, L=-1, k=2)	
Eğitim Başarısı	0.9977±(0.0002)	1.0 ±(0.0)	0.9995±(0.00005)	1.0 ±(0.0)	Mnist
Test Başarısı	0.9896±(0.001)	0.9823±(0.0008)	0.9898 ±(0.0004)	0.9854±(0.0006)	
Eğitim Başarısı	0.9190±(0.0012)	0.9514±(0.0024)	0.9460±(0.0011)	0.9736 ±(0.001)	Fashion Mnist
Test Başarısı	0.8987 ±(0.0017)	0.8783±(0.0041)	0.8940±(0.0025)	0.8781±(0.0016)	

Tablo 5: İlgili aktivasyon fonksiyonlu LeNet-5 mimari için öğrenilmesi gereken parametre sayısı

	Aktivasyon Fonksiyonu				Veri Kümesi
	Lojistik Sigmoid	Adaptif (U=1, L=0, k=1)	Tanjant Hiperbolik	Adaptif (U=1, L=-1, k=2)	
44426	44498	44426	44498	Mnist	
44426	44498	44426	44498	Fashion Mnist	



Şekil 7: Aktivasyon fonksiyonlarının LeNet-5 mimarisi, Mnist veri kümesi üzerinde eğitim ve test başarısına etkisi



Şekil 8: Aktivasyon fonksiyonlarının LeNet-5 mimarisi, Fashion Mnist veri kümesi üzerinde eğitim ve test başarısına etkisi

4. SONUÇ

Bu çalışmada öncelikle adaptif sigmoid aktivasyon fonksiyonu içeren ileri beslemeli sinir ağı çalışmalarından bahsedilmiştir. Aktivasyon fonksiyonuna serbest parametrenin eklenmesiyle ağ mimarilerine esneklik kazandırılmış olur. Serbest parametreler yol ağırlıkları ve yanlılık değerleri gibi iterasyonlar boyunca güncellenir ve bu güncelleme sonucunda fonksiyonların çalışma aralığı ve doyuma uğrama hızı değişir. Mnist ve Fashion Mnist veri kümeleri için farklı mimarilerde standart lojistik sigmoid, tanjant hiperbolik ve 2 farklı başlatma düzeniyle adaptif sigmoid aktivasyon fonksiyonları kıyaslanmıştır. Adaptif aktivasyon fonksiyonlu mimariler hem Mnist hem de Fashion Mnist veri kümelerinin sınıflandırıldığı problemler için eğitim verisinde başarılı sonuç üretirken, test verisinde sabit aktivasyon fonksiyonlu yapıların gerisinde kalarak aşırı uyumlama problemiyle karşılaşmışlardır. İlerleyen çalışmalarda aşırı uyumlama problemini azaltmak için serbest parametrelere L1, L2 gibi düzenleme yöntemlerinin uygulanmasının etkisinin ve yine bu parametrelere öğrenme katsayısı çözelgeleyici uygulanmasının etkisinin incelenmesi planlanmaktadır.

KAYNAKÇA

1. Dubey, S. R., Singh, S. K., Chaudhuri, B. B. Activation functions in deep learning: A comprehensive survey and benchmark, *Neurocomputing*, 2022.
2. Cybenko, G. Approximation by superpositions of a sigmoidal function, *Mathematics of control, signals and systems*, 2(4), 303-314, 1989.
3. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2), 251-257, 1991.
4. Leshno, M., Lin, V. Y., Pinkus, A., Schocken, S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, *Neural networks*, 6(6), 861-867, 1993.
5. Apicella, A., Donnarumma, F., Isgrò, F., Prevete, R. A survey on modern trainable activation functions, *Neural Networks*, 138, 14-32, 2021.
6. Chen, C. L., Nutter, R. S. An extended back-propagation learning algorithm by using heterogeneous processing units, *IJCNN International Joint Conference on Neural Networks, IEEE*, (Vol. 3, pp. 988-993), June, 1992.
7. Chen, C. T., Chang, W. D. A feedforward neural network with function shape autotuning, *Neural networks*, 9(4), 627-641, 1996.
8. Wu, Y., Zhao, M., Ding, X. Beyond weights adaptation: a new neuron model with trainable activation function and its supervised learning, *International Conference on Neural Networks (ICNN'97), IEEE*, (Vol. 2, pp. 1152-1157), June, 1997.
9. Nakayama, K., Ohsugi, M. A simultaneous learning method for both activation functions and connection weights of multilayer neural networks, *IEEE World Congress on Computational Intelligence (Cat. No. 98CH36227)*, (Vol. 3, pp. 2253-2257), May, 1998.
10. Xu, S., Zhang, M. A novel adaptive activation function. *International Joint Conference on Neural Networks*, (Cat. No. 01CH37222), IEEE, (Vol. 4, pp. 2779-2782) July 2001.
11. Yu, C. C., Tang, Y. C., Liu, B. D. An adaptive activation function for multilayer feedforward neural networks, *IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering, TENC0M'02*, (Vol. 1, pp. 645-650), October 2002.
12. Chandra, P., Singh, Y. An activation function adapting training algorithm for sigmoidal feedforward networks, *Neurocomputing*, 61, 429-437, 2004.
13. Kunc, V., Kléma, J. On Transformative Adaptive Activation Functions in Neural Networks for Gene Expression Inference, *bioRxiv*, 587287, 2019.
14. Rumelhart, D. E., Hinton, G. E., Williams, R. J. Learning representations by back-propagating errors. *nature*, 323(6088), 533-536, 1986.
15. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 86(11), 2278-2324, 1998.
16. Bengio, Y. Practical recommendations for gradient-based training of deep architectures, In *Neural networks: Tricks of the trade* (pp. 437-478). Springer, Berlin, Heidelberg, 2012.
17. Glorot, X., Bengio, Y. Understanding the difficulty of training deep feedforward neural networks, In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249-256), March, 2010.
18. LeCun, Y., Bengio, Y., Hinton, G. Deep learning, *nature*, 521(7553), 436-444, 2015.
19. Xiao, H., Rasul, K., Vollgraf, R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, *arXiv preprint arXiv:1708.07747*, 2017.