

Korkmaz, G. ve Eroğlu, E. (2020). “Model Karmaşıklığının Kontrolü”. *İktisadi ve İdari Yaklaşımlar Dergisi*, 2 (2), s.146-162.

MODEL KARMAŞIKLIĞININ KONTROLÜ¹

Control of Model Complexity

Gökhan KORKMAZ², Ergün EROĞLU³

ÖZET

Model karmaşıklığı, modellerin başarısındaki en önemli ölçütlerden birisidir. Bu çalışmada bugüne dek model karmaşıklığının kontrolünde öne çıkan yaklaşımlar başlıklar halinde incelenmiştir. Bunlar Occam’ın usturası, Popper’in yanlışlanabilirliği ve istatistiksel öğrenme teorisidir. Occam’ın usturası ve Popper’in yanlışlanabilirliği model karmaşıklığının kontrolünde, evet bir felsefi yaklaşım sağlamaktadırlar ve kabul de görmektedirler. Fakat model karmaşıklığının nasıl kontrol edileceği konusunda matematiksel bir formülasyon sağlamamaktadırlar. Fakat istatistiksel öğrenme teorisinin (diğer adıyla, VC teorisi) konuya yaklaşımı yalnızca felsefi bir düzeyde kalmamakta, aynı zamanda şimdiye dek geliştirilen modellerde kullanılan ampirik risk minimizasyonu (ARM) ilkesine VC katsayısını ilave ederek yeni bir risk minimizasyonu (yapısal risk minimizasyonu, YRM) ilkesi getirmektedir. Sonuç olarak Vapnik ve Chervonenkis tarafından geliştirilen VC teorisi, bir kontrol modeli olarak, ispatlanmış matematiksel arka planı ve oldukça başarılı olan sonuçları itibariyle, model karmaşıklığının kontrolü konusunda, günümüz çerçevesinde, en tutarlı ve güvenilir bir yaklaşım olarak, model geliştiriciler için iyi bir ilham kaynağı olabilir.

Anahtar Kelimeler: Model Karmaşıklığı, Occam’ın Usturası, Popper’in Yanlışlanabilirliği, İstatistiksel Öğrenme Teorisi

1 “Kredi Riskinin Belirlenmesinde Yapay Zekâ Yöntemleri ve Bir Uygulama” başlıklı, basılmamış doktora tezinden üretilmiştir.
2 İstanbul Üniversitesi İşletme Fakültesi, Sayısal Yöntemler Anabilim Dalı Doktora Programı Öğrencisi, Şırnak Üniversitesi, İktisadi ve İdari Bilimler Fakültesi İşletme Bölümü Araştırma Görevlisi ORCID: 0000-0002-1702-2965 gkhnkrkmz3873@gmail.com

3 İstanbul Üniversitesi İşletme Fakültesi, Sayısal Yöntemler Anabilim Dalı, Öğretim Üyesi, ORCID: 0000-0003-4454-6251 eroglu@istanbul.edu.tr

ABSTRACT

Model complexity is one of the most important criteria for the success of models. In this study, the prominent approaches to controlling model complexity have been examined under headings. These are Occam's razor, Popper's falsifiability, and the statistical learning theory. Occam's razor and Popper's falsifiability in the control of model complexity, yes they provide a philosophical approach and they are also accepted. However, they do not provide a mathematical formulation on how to control model complexity. However, the statistical learning theory (aka VC theory) approach to the subject is not only at a philosophical level, but also introduces a new principle of risk minimization (structural risk minimization, YRM) by adding the VC coefficient to the empirical risk minimization (ARM) principle used in the models developed so far. . As a result, the VC theory developed by Vapnik and Chervonenkis as a control model, with its proven mathematical background and highly successful results, can be a good source of inspiration for model developers as the most consistent and reliable approach to the control of model complexity in today's framework.

Key Words: Model Complexity, Occam's Razor, Popper's Falsifiability, Statistical Learning Theory

GİRİŞ

Dünyada artık iş süreçlerinin hemen tamamında, gerçekleştirilen işlemlerin tamamı bilgisayarlar ve düşük maliyetli sensörler yardımıyla çeşitli veri tabanlarında kayıt altına alınmakta ve kayıt altına alınan bu veriler zamanla veri tabanı yönetim sistemleri vasıtasıyla sistematik bir şekilde, depolanmakta, saklanmakta ve ihtiyaç duyulduğunda çok kolay bir şekilde erişim sağlanabilmektedir. Fakat kurumlar karar aşamasına geldiklerinde bu büyük bilgi yığınları, pek bir anlam ifade etmemektedir. Önemli olan bu bilgi yığınlarının, kapsamlı olarak değerlendirilip, kurum için ne anlama geldiğinin irdelenmesi gerekmektedir. Bu aşamada veri madenciliği devreye girmekte ve salt haliyle son derece kullanışsız olan bu büyük veri yığınları, bir anlamda özetlenerek anlamlı çıktılar üretilebilmektedir. Bu süreç ise veri yığınlarından adeta yeni bir hesap makinesi geliştirme çalışması da diyebileceğimiz bir “model kurma” yöntemi sayesinde gerçekleştirilmekte ve makine öğrenmesi sayesinde veri madenciliği yöntemleri robotize edilmektedir.

Şimdiye kadar, büyük veri yığınlarından anlamlı sonuçlar üretebilmek için çok sayıda modelleme yöntemi geliştirilmiştir. Bunlardan bir kısmı geleneksel de diyebileceğimiz istatistik yöntemler ve diğer bir kısmı ise, nispeten daha yeni olan, yapay zekâ yöntemleri başlıklarıyla sınıflandırılmaktadır. Bu iki sınıflandırma arasındaki temel fark şöyle ki: istatistik teknikler arka planda analitik ispatlara dayandırılırken yapay zekâ yöntemleri, bunlardan farklı olarak, veriden öğrenme (learning from data) de diyebileceğimiz, var olan veriler arasındaki örüntüleri deşifre etmeye ve bulunan bu örüntüler üzerinden model kurmaya çalışan yöntemlerdir. Dolayısıyla yapay zekâ yöntemleri, örneğin diferansiyel denklemler üzerinde formül oluşturmaya çalışırken, herhangi bir matematik kuralını ya da diller arası tercüme yapmaya çalışırken herhangi bir gramer kuralını bilmeden bunu yapar, yani gördüğü veriyi genelleştirme yoluyla kurallaştırır . Bu iki yaklaşım arasındaki tek fark elbette bununla sınırlı değildir. İstatistik yöntemler, regresyon, başta olmak üzere, modellerini belli birkaç formülü kullanarak

oluştururlar. Dolayısıyla bu modeller uygulanırken yapılan işlem sayısı, yapay zekâ yöntemlerine kıyasla, oldukça azdır. Bu yüzden çok kısa sürede sonuç alınabilir. Yapay zekâ yöntemleri ise iteratif yani bir bakıma deneme-yanılma yöntemiyle arka planında o kadar çok işlem gerçekleşir ki bu işlem kapasitesi çoğu zaman bilgisayarları zorlar, bazı durumlarda analizler birkaç gün sürebilir, hatta hiç sonuç da alınamayan ileri düzey karmaşık problemler de bulunmaktadır. Bu yüzden yapay zekâ yöntemlerinin gelişim süreci, bilgisayarların ve işlemci kapasitelerinin gelişimiyle oldukça paralellik arz eder. Yapay zekâ yöntemlerinin bu handikabını aşmak için çok farklı yöntemler geliştirilmektedir. Sezgisel yöntemler başta olmak üzere, destek vektör makinelerinde kullanılan çekirdek fonksiyonları, yapay sinir ağlarında kullanılan delta öğrenme kuralı bunlara örnek olarak verilebilir. Analitik mantığa dayanan istatistikî yöntemlerde genellikle modelleme işlemleri birkaç formülle yapıldığı için bu şekilde bir sorunla karşılaşmamaktadır. Fakat söz konusu olan yapay zekâ olunca, bu kadar işlem yükünün yanı sıra kullanılan model de karmaşıkça, hem sonuç almak çok uzun sürecektir. Hem de, modelin sonuçları iyiye değil kötüye gidecektir. Bunun sebebi, veriden öğrenen modelin, bu eğitim verisini aşırı öğrenmesi, yani modelin içindeki fonksiyonel örüntünün yanısıra verinin içindeki gürültüyü de öğrenmesi hatta bir bakıma ezberlemesinden kaynaklanır. Ezberleme ise, modeli eğitim verisine bağımlı kılar. Bu durumda model eğitim hatasını minimize ederken, genelleme hatasını minimize etmek için herhangi bir önlem almaz. Sonuç olarak modelin, eğitim (ya da test) verisinin dışındaki veri kümelerine uygulandığında, başarı oranı oldukça düşecektir. Bu sorunun önüne geçebilmek için yapay zekâ yöntemlerinin, model kurma çalışmasını, yani eğitim sürecini kısıtlamak, durdurmak ya da bir yere kadar eğitilmesini sağlamak gerekmektedir. Çünkü iyi bir yapay zekâ yöntemi, hazırdaki eğitim verisine olan odaklanma kabiliyeti ile, bunun dışındaki veri kümelerine yapılacak genelleme yeteneği arasındaki dengeyi iyi kurabilmelidir. Aksi takdirde eksik uyum/öğrenme (under fitting) ya da aşırı uyum/öğrenme (over fitting) sorunlarıyla karşılaşılacaktır. Bu sorun ise eğitim ve genelleme hatasının her ikisinin birlikte minimize edilmesiyle giderilebilecek bir sorundur.

Literatürde modelleri, kısıtlayan, cezalandıran, durdurma kriteri ekleyen, budayan vs. çeşitli yöntemler kullanılmaktadır. Bu yöntemlerden biri, istatistiksel öğrenme teorisi (statistical learning theory, SLT), sadece deneme-yanılma yöntemine dayanarak bir model iyileştirmesi sunmamakta aynı zaman da eğitim ve genelleme hatalarının birlikte minimize edilmesini sağlamaya çalışmaktadır. Bunun için, halihazırda yaygın olarak kullanılan risk (yanlış tahminin gerçekleşmesi) minimizasyonu yöntemi olan ampirik risk minimizasyonu (ARM) yöntemine, bir VC boyutu sınırlaması dahil ederek, yeni bir risk minimizasyonu yöntemi olan, yapısal risk minimizasyonunu (YRM) geliştirmekte ve bu soruna kuramsal bir çözüm yaklaşımı getirmektedir.

Model karmaşıklığı ne olmalıdır? Modeli basitleştirirken modelin performansından ödün vermek gerekmez mi? vb. bu konudaki birçok tartışma, model karmaşıklığının sınırlarını felsefi olarak yani yaklaşım açısından basitleştirmiştir. Fakat bunu matematiksel olarak başarabilmek ve modelleri terbiye etmek aynı şekilde kolay olmamaktadır. Yine de konunun nerelerden geçerek geldiğini anlamak ve olgunlaşan bu model karmaşıklığı sınırlarını iyi çizmek için konu üzerinde şimdiye dek etkili olan yaklaşımları ele almak gerekmektedir. Model karmaşıklığı ve optimizasyonu ile alakalı olarak şimdiye dek yapılan çok farklı tartışmalar bulunmakta ve hatta bunlar birkaç yüzyıl ötesine kadar gitmektedir. Occam's razor bunlardan en çok bilineni ikincil olarak ise Popper'ın yanlışlanabilirliği, felsefi

yaklaşımlar olmakla birlikte, bilim tarihine damga vurmuş iki önemli akımdır. Bilim felsefesi alanında yapılan tartışmaların, bilimin gelişmesi adına faydalı olacağı kabul edilecek olursa, bu yaklaşımların da konumuz çerçevesinde tartışılmasının literatüre katkı sağlamış olduğunu kabul etmek gerekmektedir.

Model karmaşıklığını kontrol etmenin uygulama sonuçlarına yadsınamaz bir katkısının olacağı, bu problemle başa çıkamayan yöntemlerin, olumsuz analiz sonuçları dikkate alındığında, netlik kazanmaktadır. Bu çalışma da konunun geçmişi ve bugün geldiği en son nokta tespit edilmeye çalışılacaktır. Bunun için ilk olarak model yaklaşımı, sonra konunun arka planında yer alan makine öğrenmesi ve bu konuya dair kavramsal çerçeve etrafıca ele alınacak ardından probleme getirilen felsefi ve matematiksel yaklaşımlar tartışılacaktır. Konunun geçmişi ve geldiği nokta tespit edildikten sonra geleceğine dair bir tahminleme yapılacaktır.

1. MODEL

Model, Fransızca “modèle” kelimesinden gelmekte olup kelime karşılığı, biçim, örnek, tip, projeksiyon vb. şeklindedir (TDK, Genel Türkçe Sözlük: 11/11/2019). Konumuzla alakalı olarak ise model, büyük veri yığınlarından anlamlı karar çıktıları üretebilmek için, bu verileri sistematik bir şekilde özetleyen, azaltan, indirgeyen ve belli bir biçime (örneğin matematiksel bir fonksiyona) dönüştüren matematiksel ya da istatistiksel faaliyetlerin bütünüdür. Makine öğrenmesi, “veriden örüntüleri (desenleri) bulup çıkararak, otomatize edilmiş bir süreç” olarak tanımlanabilir (Kelleher, 2015).

Model kavramı, en bilinen şekliyle anlatılacak olursa, matematiksel olarak bir fonksiyonun türevinin nasıl alındığını zihninde modelleyebilen bir matematik öğrencisinin, artık binlerce farklı fonksiyonun türevini ezberlemek ya da yanında taşımak zorunda kalmadan herhangi bir fonksiyonun türevini rahatlıkla alabilmesine benzetilebilir. Bunun gibi, bir veri kümesi üzerinde model geliştirebilen bir kişi de, artık veri yükünden kurtulur ve bu modeli konuyla ilgili tüm veri setlerine uygulayabilir (Emir, 2013: 41-42). Dolayısıyla her bir modeli, farklı bir hesap makinesi olarak görmek, özellikle konumuz açısından oldukça faydalı bir benzetme olacaktır.

Modelleme, aynı zamanda bir soyutlama sürecidir. Bir sonraki problemin çözümü için model yeterli kıvama gelmişse geriye kalan tüm verilerden ya da unsurlardan oluşturulan modelin soyutlanması söz konusudur.

Bir modelin modelliğinden bahsedebilmek için bu modelin, üç temel özelliğe sahip olması beklenmektedir (Kühne, 2005):

1. Haritalama özelliği (mapping future): Bir modelin, “orijinal veriyi temel alması”,
2. Küçültme özelliği (reduciton future): Bir modelin, “yalnızca orijinalin seçimini yansıtması”,
3. Faydalılık özelliği (pragmatic future): Bir modelin, bir amaç için “orijinali yerine kullanılabilir” olması.

Sonuç olarak unutulmamalıdır ki her modelleme çalışması, belli bir oranda, orijinal verinin kapsayıcılığından ya da bilgi haznesinden tavizi gerektirir. Çünkü bir gerçek hayat verisini %100

temsil edebilecek bir model, orjinal verinin yine kendisidir. Yani modelin hafifletici ve hesaplayıcı faydasından yararlanabilmek için, muhakkak veride yer alan bilgiden belli bir miktar fedakârlık etmek gerekmektedir. Aksi takdirde, modelin başka veri kümelerine de uyarlanabilme yani genelleme yeteneği oluşmayacaktır. Bununla birlikte son bir iki on yılda geliştirilen ve oldukça olgunlaşan birçok istatistik ve yapay zekâ yöntemleriyle oluşturulan modeller artık %90'ın üzerinde bir doğruluk oranı (accuracy rate) yakalayabilmekte ve böylelikle söz konusu modelleme kayıpları, minimize edilebilmektedir.

Farklı farklı sektörlerde kullanılan modelleme çalışmaları ile:

- Veri madenciliği kapsamında; müşteri ve pazar profilleri oluşturma, mal değeri ve satış tahminleri üretme, üretim ihtiyacını belirleme, pazar performansını ve çevresel ekonomik riskleri tahmin etme
- Bankalarda, kredi müracaatlarını değerlendirme, kredi kartı sahtekârlıklarını saptama, borçlanma ve risk değerlendirmeleri
- Araçlar ve robotlar için optimum güzergah tespiti, robot hareketlerinin kontrol edilmesi
- Hava tahminleri
- At yarışı, futbol vb. müsabaka sonuçlarının tahmin edilmesi
- Güvenlik sistemlerinde ve soruşturmalarda, kart, göz, yüz, avuç içi, parmak, imza ve yazı tanıma, çek okuma, şekil tanıma, spektrum tanımlaması
- Mekanik parçaların ömürlerinin ve kırılmalarının tahmin edilmesi, meydana gelebilecek, arıza, patlama, yıkılma vb tahmin edilmesi
- Kalite kontrol süreçleri, işlem kontrolü, sistem kontrolü, kimyasal yapılar, dinamik sistemler, işaret karşılaştırma, plastik kalıpcılık, kaynak kontrolü
- İş çizelgelerini ve iş sıralamasını belirlenmesi
- İletişim kanallarındaki geçersiz ekoların filtrelenmesi yine iletişim kanallarındaki trafik yoğunluğunu kontrol etme ve anahtarlama
- Radar ve sonar sinyallerinin sınıflandırılması
- Üretim planlama ve çizelgeleme
- Kan hücreleri reaksiyonları, hücre tiplerinin tanımlanması ve kan analizlerinin sınıflandırılması, kanser ve diyabet gibi dâhil değişik tıbbi teşhisler ve kalp krizinin tedavisi, vb. birçok çalışma yürütülmektedir (Elmas, 2016: 88; Öztemel, 2012: 36).

Modelleme aşamasındaki konuyla alakalı tüm kavramları sırasıyla, kısaca özetleyecek olursak;

Veri; insanların daha sonra çeşitli amaçlar için kullanabilmek amacıyla saklama ihtiyacı duydukları, salt, yorumsuz, ham bilgilerdir. Bir başka tanımla veri “bir problemde bilinen anlatımlardan yola çıkarak bilinmeyi bulmaya yarayan şey, bilgi, data” şeklinde tarif edilmektedir (TDK, Genel Türkçe Sözlük, 11/11/2019).

Veri tabanı kavramı bilgi-işlem süreçlerinde, “sistemik erişim imkânı olan, yönetilebilir, güncellenebilir, taşınabilir, birbirleri arasında tanımlı ilişkiler bulunabilen bilgiler kümesi” olarak tanımlanmaktadır (Kocamaz, 2007: 22).

Veri tabanı sistemi, çeşitli amaçlar için toplanan verileri düzenli bir şekilde saklayan ve bu verileri kullanarak yararlı bilgi üreten bilgisayarlı kayıt tutma sistemidir (Alp vd., 2011).

Veri tabanı yönetim sistemi ise, kullanıcı ile veri tabanı arasındaki iletişimi sağlamakla birlikte veri tabanlarının oluşturulmasını, yedeklenmesini, kullanıcı erişimlerinin düzenlenmesini kısaca yönetilmesini sağlayan yazılımlardır. Bu yazılımlar, kişi ve kurumların veriyi etkin bir şekilde erişimine imkân tanıyan farklı tür bilgilerin, aynı ortamda tutulmalarını, bu bilgilere erişimin hızlı ve kolay olmasını, bilgi tekrarlarının önlenmesini, sürekli yapılması gereken işlemlerin hızlı ve doğru yapılmasını, aynı bilgi kümelerinin birden fazla kişi ya da kurum tarafından, aynı anda kullanılabilmelerini ve verilerin güvenliğini sağlamaktadırlar (Alp vd., 2011).

Veri madenciliği, büyük veri yığınları içerisinde, gelecekle alakalı tahminler yapılabilmesi için bağlantıların, kuralların veya örüntülerin çeşitli paket programları aracılığıyla ortaya çıkarılmasına dönük çalışmaları kapsamaktadır. Veri madenciliği, büyük miktarlarda toplanan verinin, bir kısım analizler neticesinde anlamlı ve yararlı çıktılara dönüştürülmesidir. Veri madenciliği iki temel faaliyet üzerine kuruludur. Bunlar çok fazla miktarlarda verinin veri ambarlarında kullanışlı bir şekilde depolanması ve bu verilerden anlamlı ve yararlı çıktılarının üretilmesidir (Gürsoy, 2009: 27).

Makine öğrenmesi, “daha önceden elde edilmiş örneklere veya vakalara dayanan, açıklayıcı değişkenlerden oluşan tanımlayıcı özellikler (descriptive features) ve bağımlı değişken dediğimiz hedef özellik (target feature) arasındaki ilişkilerin bir modelini öğrenen otomatik bir teknik” şeklinde de tanımlanabilir (Kelleher, 2013: 3).

2. MAKİNE ÖĞRENMESİ

Makine öğrenmesi, temelde bir veri madenciliği yöntemidir ve verilerden anlamlı çıktılarının üretilmesi için gerçekleştirilen “veriden öğrenme” çalışmalarının, yazılımlar aracılığıyla ve algoritmalar yardımıyla makinelerle yaptırılması sürecini içeren bir yapay zekâ yöntemidir. Veriden öğrenme yöntemlerinin genel yaklaşımı; değişkenler arasındaki ilişkilerin tespiti yoluyla bir model oluşturmak, sonra bu modeli yeni verilere uygulayarak muhtemel sonuçları doğru bir şekilde tahmin etmektir (Kelleher, 2013:3).

Yazılımlar modelleme yaparken, makine öğrenmesini farklı şekillerde geliştirirler. Örneğin bir onkoloji hastanesinde, bir doktorun elinde onkoloji hastanesine kan vermiş hastaların kan değerlerinin verisinin bulunduğunu varsayalım. Hastaların bu veride yer alan kan değerlerine karşılık yine aynı hastaların kanser olup olmadıklarının bilgisi de yer alıyorsa, yani hastaya (gözleme) ait sonuç değeri de biliniyorsa “denetimli öğrenme (supervised learning)” modelleri geliştirilebilir. Hastaya ait sonuç değerleri sayesinde hata payları atanarak fonksiyon katsayılarının yeniden düzenlenmesiyle optimum bir model (fonksiyon) oluşturulabilir. Fakat hastaların bir kısmının kanser hastası olup olmadığı biliniyor, diğer kısmının ise sonuç değerleri bu şekilde bilinmiyorsa, bu durumda, hiç bilinmeyen duruma kıyasla iyi

sonuç vereceği düşüncesiyle, tüm gözlemler modele girilir ve bu yöntemin adı “yarı-denetimli öğrenme (semi-supervised learning)” olur. Sonuç değerlerinin hiçbiri bilinmiyorsa, gözlem değerleri olduğu gibi modele girilir. Bu durumda model “denetimsiz öğrenme (unsupervised learning)” gerçekleştirir. Bir de güçlendirici öğrenme (reinforcement learning) yöntemi bulunmaktadır. Bu yöntem de doğru sonuca ödül yanlış sonuca ceza vermek yöntemi kullanarak öğrenme aşaması gerçekleştirilmektedir.

Denetimli öğrenmede, yanlış sonuç alma durumunda hata payını kullanarak geri döndürme ve böylelikle fonksiyon katsayılarını optimize etme yöntemi kullanılırken, denetimsiz öğrenme de, açıklayıcı (bağımsız) değişkenler arasındaki korelasyonlar dikkate alınır. Sınıflandırma ve regresyon problemleri denetimli ya da yarı denetimli öğrenme ile çözülebilirken, kümeleme gibi, daha çok boyut indirgeme (dimension reduction) problemleri, denetimsiz öğrenme yöntemleri kullanılarak çözülür.

Eğitim aşaması için mevcut verinin genellikle %80 gibi bir kısmı, model geliştirirken kullanılır. Bu veriye eğitim verisi (training data) adı verilir. Yani model bu verilerle eğitilir. Buradaki verinin kalitesi, ya da anakütle temsil kabiliyeti, modelin başarısını belirleyen temel unsurlar arasındadır. Verinin geri kalan %20 lik kısmında, ki bu veriye de test verisi (testing data) adı verilir, geliştirilen model uygulanır ve gerçek sonuçlarla, modelin bulduğu sonuçlar kıyaslanır. Buradan modelin doğruluk oranı (accuracy rate) hesaplanır. Geliştirilen modelin başarısı düzeyine, yine benzer bir yöntem ya da birden çok yöntem ilgili veri setine uygulandıktan sonra, modellerin elde ettikleri doğruluk oranlarının mukayese edilmesi yoluyla karar verilir.

Literatürde kullanılagelen sayısız model bulunmaktadır ve bu modeller uygulama sahasında doğruluk oranları göz önüne alınarak zamanla bir doğal seleksiyon geçirmektedirler. Doğruluk oranları yüksek olan modeller zamanla daha çok kullanılmakta ve doğal olarak da daha çok geliştirilmektedir. Gelişen modeller, daha çok kabul görmekte ve literatür ve uygulama bazlı olarak hızla yaygınlaşmaktadır. Modellerin geniş ölçüde kabul görmesinin, doğruluk oranları dışında da bazı nedenleri bulunmaktadır. Bunlardan en önemlisi, birçok çalışmada kullanılan bu makine öğrenmesi modellerin varsayım gerektirmemesi ve herhangi bir dağılıma bağımlı olmaması yani bir bakıma uygulamasının geleneksel istatistiki yöntemlere kıyasla daha kolay olmasıdır.

3. MODEL KARMAŞIKLIĞI

Bir model, eğitim verisi üzerinde çok yüksek bir doğruluk oranı verebilir. Fakat burada önemli olan, yani bir modelin başarısını belirleyen, o modelin test verisinden sonra (sonuçta test verisi de eğitim verisiyle aynı kümeye ait, karakteristik olarak benzer niteliğe sahiptir), tamamen bağımsız, başka başka veriler üzerinde de aynı başarıyı verebilmesi gerekir ki, bu nokta birçok yöntem için önemli bir kırılma noktasıdır. Peki, neden bir model eğitim ya da test verisi üzerinde gayet yüksek bir oranda doğruluk sağlarken, tanımadığı başka bir veri seti söz konusu olduğunda, bu oran ciddi ölçüde düşer? Bu durum iki önemli kadim sorundan kaynaklanmaktadır. Bunlar “eksik öğrenme (under fitting)” ve “aşırı öğrenme (over fitting)” problemleridir.

İyi bir yapay zekâ yönteminin; üzerinde eğitildiği veriye odaklanma yeteneğiyle, diğer veri setlerine uyarlanabilme yeteneği arasındaki dengeyi iyi kuramayıp eldeki eğitim verisine aşırı yoğunlaşırsa (aşırı öğrenme, ezberleme) diğer veri setlerine olan genelleme kabiliyeti azalacaktır. Diğer durumda ise,

yani eğitim verisine gerekenden daha az öğrenir ya da yoğunlaşırsa eksik uyum problemi yaşanacaktır. Bu sonuç ise genelleme kabiliyetini artıracak fakat veri setinde yer alan ağlar yeterince deşifre edilemeyecektir. Her iki durum optimaliteyi olumsuz etkilerken, iyi bir modelin her iki problemden kaçınılması gerekmektedir.

Eksik öğrenme ve aşırı öğrenme problemlerini bilinen bir örnek üzerinde tartışacak olursak, bir görüntü işleme modelini çeşitli ağaç fotoğraflarıyla eğittiğimizi düşünelim. Sonra eğitim verisinin dışında, yani yeni bir ağaç fotoğrafı modele okuttuğumuzda, eksik öğrenen bir model, sırf yeşil diye her gördüğü fotoğrafı ağaç olarak sınıflandırırken, aşırı öğrenen model ise, sırf yaprak sayısı farklı diye, gördüğü fotoğrafı ağaç olarak kabul etmeyecektir (Burges, 1998:122).

Biraz daha hayatın içinden bir örnek verecek olursak, iki araba satıcısı düşünelim. Arabalara fiyat değerlemesi yaparken bunlardan birincisinin, sadece arabanın dış görünümünü ve markasını dikkate aldığını varsayalım. Diğerinin ise, arabaya test sürüşü yaptığını, konforunu, performansını, hasar kaydını, rengini, sınıfını, yaşını, motor seçeneğini, yakıt tipini, şanzıman seçeneğini, donanımını vs. çok sayıda kriteri göz önüne alarak fiyatlandırma yapmaya çalıştığını varsayalım. Bu iki satıcının yanında yetişen asistanlardan birincisi fiyat değerlendirme işini eksik öğrenirken, ikincisi aşırı öğrenmeye maruz kalacaktır. Birinci satıcı ya da asistanın, yeni araç fiyat değerlemelerinde, varyansı düşük çıkacaktır ve ortalama araç değerlerinin değişmesi için çok fazla yeni ve farklı fiyattaki aracın değerlendirilmesi gerekecektir. Bu ilk asistanın genelleme kabiliyeti yüksek, odaklanma kabiliyeti ise düşük olacaktır. İkinci asistanın değerlemelerinde ise varyans yüksek olacaktır. Yeni araçlar için ortalama değerlerinin değişmesi için çok az sayıda yeni ve farklı fiyat aracın gelmesi yeterlidir. Bu kişinin ise genelleme kabiliyeti düşük, odaklanma kabiliyeti ise yüksek olacaktır. Burada yapılması gereken, birinci tip satıcıyı, aracın daha başka özelliklerini de dikkate almasını sağlamak, ikinci tip satıcıyı da biraz sınırlandırmak ve fazla ayrıntıya girmesine engel olmak olacaktır.

Sonuç olarak optimal sonuç için, modelleme aşamalarında öğrenme düzeyinin optimum düzeyde gerçekleştirilmesi gerekmektedir. Bunu sağlayabilmek içinse modelin aşırı basit ya da aşırı karmaşık bir model olmamasının önüne geçilmelidir. Bunun içinde modelin karmaşıklığını optimize edecek, bir takım kontrol parametrelerinin modele ilave edilmesi gerekecektir.

4. MODEL KARMAŞIKLIĞININ OPTİMİZASYONU

Matematiksel olarak, model karmaşıklığı ayarlaması, ya da model karmaşıklığının kontrolü (control of model complexit), optimizasyonu nasıl sağlanır? Bunun için şimdiye kadar literatürde, karar ağaçlarında ki “budama” tekniği gibi, farklı yöntemler önerilmiş olup, bu yöntem ve yaklaşımları, bugünkü bakış açısıyla üç farklı başlık altında toplamak mümkündür. Bunlar Occam’ın usturası (Occam’s razor), Popper’ın yanlışlanabilirliği (falsifiability of popper) ve VC Teorisi (Vapnik-Chervonenkis Theory) yaklaşımlarıdır.

Model karmaşıklığını kontrol etmek için önerilen yöntemler ciddi ölçüde bilim felsefesi ve bilgi kuramı kavramlarını içermektedir. Bunun nedeni, öngörülü öğrenme ve bilim felsefesi arasında derin bir

bağlantı vardır. Çünkü herhangi bir bilimsel teori ampirik verileri veya geçmiş gözlemleri açıklamak için tümevarımcı (inductive) bir yöntem kullanmaktadır (Cherkassky, 2007:147). Bu çalışmada, tüm bilim çevrelerinin ezberlerini bozan Occam'ın usturası ve Popper'in yanlışlanabilirliği, genel bilim felsefesi bakımından kendi başlıkları altında tartışılmıştır. Daha sonra, bu iki felsefi akımın, geliştirilmesinde ve istatistiksel temeline önemli etkileri olan, pratikteki adıyla "VC teorisi", bilimsel adıyla "istatistiksel öğrenme teorisi" incelenmiştir. VC teorisi bu akımların bilim dünyasına kazandırdıkları yaklaşımı da dikkate alarak yeni ve kapsamlı bir bilimsel paradigma oluşturmuş ve bu yöntem bugün bilim sahasında en yüksek verime sahip model optimizasyonu tekniğidir (Cherkassky ve Mulier, 2007).

4.1. Occam'ın Usturası

William, Occam'ın usturası (Occam's Razor), adıyla anılan ve "açıklamaların azabı (parsimony of explanations) (Rasmussen vd. 2020), olarak da özetlenen prensibini "varlıklar, gerekliliğin ötesinde çoğaltılamaz" görüşüne dayandırmaktadır. Occam'ın usturası ile genel olarak kastedilen, "eğer bir bilimsel problemin çözülmesi için iki farklı model söz konusu ise bunlardan basit olanı, yalın olanı, ayrıntısı az olanı tercih edilmelidir" ve "sistemin verimini artırmayan tüm ayrıntıların budanması" düşüncesidir. Occam'ın usturası, genel bilim felsefesi açısından çok önemlidir ve bilim dünyasında da her zaman geliştirilen yeni fikirleri ve modelleri pragmatik bir bakış açısıyla budamaya devam etmektedir. Modele dönük budama işlemleri, sonuçlara olumsuz tesir etmiyorsa, faydalı ve yerinde görülmektedir (Duda vd., 2000:465).

Occam'ın usturası yaklaşımı, diğer yandan işletmelerin yönetim ve organizasyonunda, yalın organizasyon modern yönetim yaklaşımını anımsatmaktadır. Yalın organizasyon stratejisi, işletmelerin yönetim süreçlerinde, "kullanılmayan veya lazım olmayan her şeyi sistemin dışında bırakma ve sistemi daha kompakt ve daha işler hale getirmek" mantığını taşımaktadır. Yalın organizasyon veya Occam'ın usturası, bir başka benzetme ile; fiziki çalışma konusunda elverişli olmayan, hantal, yağlı, kassız, bir insan vücudunu, sağlıktan ödün vermeden, daha hafif, atik, güçlü, kaslı bir çalışan formuna dönüştürme çabası olarak yorumlanabilir. Böylelikle yapılan işlerde hız, esneklik ve işlevsellik kazanmak birinci amaçtır. Fakat yönetim sürecini doğrudan etkileyen unsurların, yalın organizasyon çerçevesinde, sürecin dışına itilmesi, ya da hafif ve esnek olma adına kişinin sağlığına zarar verebilecek bir ekstrem yaklaşım eskiye kıyasla daha olumsuz bir sonucu beraberinde getirecektir.

Kabul edileceği üzere, bilim dünyasında, özellikle bilim üreticileri için, adeta bir sınav niteliği taşıyan bu yaklaşım tamamen felsefidir ve sadece istatistik değil tüm bilim sahaları için geneldir ve yapılacak olan bu budamanın istatistiksel boyutunu veya matematiksel formülünü içermemektedir. "Varlıklar, gerekliliğin ötesinde çoğaltılamaz" derken burada "varlıkların ve gerekliliklerin ölçüsü kime ve neye göre belirlenecektir" sorusu da tartışmaya açık bir sorudur. Evet, prensip olarak güzel, faydalı ama matematiksel temelden yoksun olduğu için, konumuz açısından, ayakları tam olarak yere basmayan ve felsefi çerçevede kalan bir yaklaşım olarak kalmaktadır (Cherkassky ve Mulier, 2007:147).

4.2. Popper'in Yanlışlanabilirliği

20. yüzyılın en önemli bilim felsefecilerinden biri olarak kabul edilen Avusturyalı, Sir Karl Raimund Popper (1902–1994)'a göre tümevarım yöntemi geçersiz ve gereksiz bir yöntemdir. Yani bilimsel bir keşif,

çeşitli uygulamalar neticesinde, doğru sonuçlanıyorsa, bu tecrübeler o keşfin doğruluğunun ispatlandığı anlamına gelmez. Sadece, “teori ciddi şekilde test edilmiş veya iyi bir şekilde doğrulanmıştır” denilebilir. Popper’ın bilimsel yöntem hakkındaki yaklaşımı, “bütün sistemleri zorlu bir sınamadan geçirerek, sonunda nispeten elverişli olanı almak” kaydıyla, her kuramı “yanlışlama” ya tabi tutma şeklindedir. Popper, Einstein’ın görecelik kuramı, Karl Marx’ın tarih anlayışı ve Freud’un psikanaliz kuramı ve Alfred Adler’in bireysel psikoloji kuramı gibi bilimsel köşe taşlarına farklı bir ilgi duydu (wikepeida, 14/04/2020).

Einstein’ın, test edilebilir tahminlerle, cesur varsayımlar yapma isteğinden etkilenen Karl Popper (Corfield vd., 2005, s.1), özellikle Einstein’ın kuramının ileri sürdüğü bir yaklaşımın (güneşin yakınından geçen ışık ışınları, güneşin yerçekimi alanının etkisine girerek eğilmeye uğrarlar) 1919’da güneş tutulmasının olması sırasında doğrulanması Popper için, yanlışlanabilirlik adına, yeni bir ilham kaynağı oldu. Popper’i etkileyen kuramın öndeyişinin doğru çıkması değildi. Söz konusu hipotezin doğru çıkmaması halinde, yanlışlanmış olacak olan teori, derhal reddedilecekti. Önemli olan kuramın yanlışlanmaya açık, yani açık yürekli bir biçimde formüle edilmesiydi. Popper, diğer kuramların (Marx, Freud, Adler) sahiplerinin hangi koşullarda kuramlarından vazgeçeceklerini belirtmediklerine dikkat çekti. Doğrulayıcıları çok fakat yanlışlayıcıları belirsiz olan bu kuramlar, ona göre bilimsel olmayan kuramlardı (wikipedia, 07/04/2020).

Popper, yanlışlanabilirlik (yanlışlanabilirlik tedavisi), yaklaşımını 1934 tarihli Araştırma Mantığı (Logik der Forschung) adlı kitabında, temellendirmiştir. 1959’da ilk kez İngilizceye çevrilen kitabında, Popper; bilimi, diğer iddialardan ayıran tek faktörün yapılacak tahrifata dayanıklılık olduğunu belirtmiştir. Yani öne sürülen bilimsel keşif, eğer doğruysa tahrif edilme riskine karşı koyma yetisine sahipse bilimsel keşif kimliği kazanabilir. Popper, hangi kuram olursa olsun belli koşullarda ampirik destek bulmanın kolay olduğunu; bilimselliğin ampirik destek sağlamada değil, kuramın hangi koşullar altında yanlış olduğunu belirlemeyi esas aldı. Eğer bir kuram yanlışlanabilir ise, bilimseldir. Karl Popper, en iyi kuram “zamana bağlı olarak yanlışlanabilir, çürütülebilir olan kuramdır” demiştir (Corfield vd., 2005:1).

Bu arada Viyana Çemberinde yer alan kişiler ve konuyla ilgilenen mantıkçı ampiristler, doğrulama ve tahrifatın bu simetrik olmayan çözümüne katılmadılar ve olasılıksal bir çerçevede bir doğrulama teorisi geliştirdiler. “Bilimsel Keşif Mantığı (The Logic of Scientific Discovery)” adı verilen bu teorinin önemli bir çoğunluğu, bilginin söz konusu olduğu durumlarda, yapılan keşfin doğruluk olasılıklarını göz önüne alma yaklaşımıyla ele alınmaktaydı. Bu sırada Popper yavaş yavaş, Bayes karşıtı bilim filozoflarının toplanma noktası haline geldi. Bu fikir, Popper’ın yanlışlanabilirliği (Popper’s falsificationism) olarak ün yapmış ve zamanla bu yaklaşımından dolayı da, tümevarımı önemli bir bilimsel yöntem olarak gören ve Viyana Çemberi olarak da bilinen 20. YY’ın önde gelen mantıksal pozitivistlerinden (Hans Hahn, Kurt Gödel, Felix Kaufmann, Philip Frank, Bertnard Russel, Whitehead, A.J. Ayer, Wittgenstein) (wikipedia, 07/04/2020) ayrılmış ve farkını da ortaya koymuştur (Corfield vd., 2005:1).

Meşhur örnek üzerinden şekillendirecek olursak; üç, beş ya da on tane beyaz kuğu görülürse “bütün kuğular beyazdır” diyemeyiz. Bu tümevarımdır ve geçerliliği tartışmalıdır. Sadece “on tane beyaz kuğu görülmüştür” denilebilir. O halde, Popper açısından “bilimsel keşifler doğrulanamaz”. Farklı renkte bir kuğu görüldüğünde ise “bütün kuğular beyazdır” teorisi rahatlıkla yanlışlanabilir. O zaman

“bilimsel teoriler doğrulanamaz, fakat yanlışlanabilir”. Sonuç olarak, bir teörinin bilimsel olabilmesi için evrensel ispatı gerekli değil, fakat yanlışlanabilir olması yeterlidir. Eğer farklı renk bir kuğu ispatlanamamışsa, yani bütün kuğuların beyazlığı, yanlışlama sürecinden başarıyla geçmiş ve “evet, kuğular beyazdır” teörisi, bilimsel keşif niteliği kazanmıştır.

Popper’ın yanlışlanabilirliği aslına bakıldığında, Occam’ın usturası ile aynı yönlü bir teoridir. Occam, gereksiz ayrıntıların budanması gerektiğini belirtmiştir. Popper ise yanlışlama yoluyla, doğru olmayan kısımların budanması gerektiğini ve böylelikle bilimsel keşif kimliği kazanılabileceğini belirtmişlerdir. Yeni bilimsel keşifleri de bir nevi sınava tabii tutan, her iki yaklaşımda felsefi boyuttan öteye geçmemekte, konumuz açısından matematiksel olarak bir çerçeve sunamamaktadırlar. VC teörisi, modeli basitleştirme ya da budama şeklinde değil, modelin öğrenme düzeyini en verimli noktada durdurma yaklaşımıyla model karmaşıklığını optimize etmekte (minimize değil) ve konunun matematiksel sınırlarını da belirleyerek, ispatlamakta ve optimum sonuçlar ürettiği ise bilinmektedir.

4.3. İstatistiksel Öğrenme Teörisi

İstatistiksel öğrenme (VC) teörisine girmeden önce teörinin modelleme sürecine katkılarını netleştirebilmek için, bugüne kadar ki modelleme süreçlerinin ana eksenini oluşturan ARM’nin genel teorik çerçevesini ve optimal sonuç mantığının iyi anlaşılması gerekmektedir. Zaten VC teörisi de ARM yönteminin tutarlılık sınırlarını araştırırken ortaya atılmış bir teoridir. ARM genel bir ilkedir ve regresyonda kullandığımız en küçük hata kareleri ve yoğunluk tahmininde kullandığımız maksimum olabilirlik (maximum likelihood) yöntemleri ARM ilkesinin konusuna göre özelleştirilmiş halleridir. (Vapnik, 1999:989).

Eğitim verisiyle eğitilen modeller, artık gerçek hayatta rastlanılan eğitim verisi dışındaki verilere uygulanır ve modelin tecrübesine göre sonuçlar alınır. Modellerin doğru çıktıyı verememe riskine, “ampirik risk” denilmektedir. Bir başka deyişle model çalıştırmayı, bir nevi modeli deneme şeklinde düşünürsek, ampirik sonuç ile gerçek sonucun farklı olması riski, ampirik risk olarak tanımlanmaktadır. ARM ilkesi doğrultusunda, eğitim hatasını minimize etmek için ampirik riskin minimize edilmesi gerekmektedir. Yani matematiksel olarak bir risk fonksiyoneli belirlenip bu fonksiyonun değerinin minimize edilmesi gerekmektedir.

Ampirik risk minimizasyonu muhtelif formüllerle, birçok modele uygulanabilmektedir. Ampirik riskin formülünü vermeden önce makine öğrenmesinin matematiksel açıklamasına bakmak yararlı olacaktır. Vapnik, veriden öğrenme (konumuz açısından makine öğrenmesi) sürecini üç aşamada özetlemektedir. Makine öğrenmesi denetleyicinin yanıtını mümkün olan en iyi şekilde öngören bir dizi işlevden birini seçmektir ve bu üç adımda gerçekleştirilir (Vapnik, 1999:989):

1. Her birine x denilen tesadüfi vektörler, $P(x)$ gibi sabit fakat bilinmeyen bir dağılımdan açıklayıcı bir şekilde çizilir,
 2. Denetmen, her giriş vektörü (x) için, sabit de olan fakat bilinmeyen bir şartlı dağılım fonksiyonuna göre bir çıkış vektörü (y) döndürür. Bu seçim yine sabit ve aynı dağılımlı gözlemleri $(P(x,y)=P(x)P(y|x))$ oluşturur.
 3. Fonksiyonların bir dizi uygulama yeteneğine sahip bir öğrenme makinesi $(f(x,\alpha)\alpha\in\Lambda)$ oluşturulur.
- Öğrenme problemi, aslında bir fonksiyon $(f(x,\alpha)\alpha\in\Lambda)$ tahmin problemidir. Denetim otoritesinin cevabına

en uygun yaklaşımı seçmek için, denetim otoritesinin verilen bir cevaba verdiği cevap ile öğrenme makinesi tarafından verilen cevap arasındaki kayıp veya tutarsızlık ($L(y, f(x, \alpha))$) ölçülür ve bu tutarsızlığı kullanarak bir beklenen (gerçek) risk fonksiyonu:

$$(4.1) \quad R(\alpha) = \int L(y, f(x, \alpha)) dP(x, y)$$

oluşturulur. Burada amacı, risk fonksiyonunu ($R(\alpha)$) minimize ederek, optimum fonksiyonu ($f(x, \alpha_0)$) bulmaktır. Fakat burada $dP(x, y)$ dağılımının olasılık dağılımı bilinmediği için beklenen riskin doğrudan hesaplanması mümkün değildir. Bu yüzden, eldeki hesaplanabilen tek risk fonksiyoneli olan ampirik risk fonksiyonuna yoğunlaşılır;

$$(4.2) \quad R_{emp}(\alpha) = \frac{1}{l} \sum_{i=1}^l L(y_i, f(x_i, \alpha))$$

Burada L fonksiyonundan kasıt, beklenen sonuç ile ampirik sonuç arasındaki farkın karesine eşittir:

$$(4.3) \quad L(y, f(x, \alpha)) = (y - f(x, \alpha))^2$$

Ortalama karesel hata (OKH) yöntemi kullanıldığında formüller bu şekilde belirlenir. Ortalama mutlak hata kullanıldığında ise L fonksiyonu:

$$(4.4) \quad L(y, f(x, \alpha)) = |y - f(x, \alpha)|$$

Şeklinde olacak ve bu durumda ampirik riskin formülü:

$$(4.5) \quad R_{emp}(\alpha) = \frac{1}{n} \sum_{i=1}^n |y_i - f(x_i, \alpha)|$$

Şeklinde olacaktır. n burada gözlem sayısını temsil etmektedir (Vapnik, 1999:989).

VC teorisinin temelleri, Vladimir Vapnik ve Alexey Chervonenkis tarafından, 1960'lı yıllarda geliştirilmiştir. Rus Matematikçi Vladimir Vapnik ve arkadaşı Alexey Chervonenkis, yukarıda formülasyonu da verilen ARM ilkesinin tutarlılık sınırlarını araştırırken istatistiksel öğrenme teorisini

geliştirmişlerdir. VC teorisi, VC katsayısını temel almaktadır. ile gösterilen ve genel olarak uzay boyutunun bir fazlasına (yani çoğu durumda $h=n+1$) denk gelen VC katsayısının hesaplanması her zaman kolay değildir ve çoğu zaman sonsuz da olabilmektedir. Konuyu aşırı genişletmemek için bu çalışmada VC katsayısının hesaplanması konusuna girilmemiştir. ARM ilkesinin yanı sıra, model karmaşıklığının da VC katsayısı ile kontrol edilme suretiyle aşırı öğrenmeden kaçınarak daha optimal sonuçlar üretmeye odaklanan bu risk yaklaşımına ise yapısal risk minimizasyonu (YRM) adı verilmiştir (Vapnik, 1999:993). Yani YRM'nin ARM'den farkı, VC katsayısı sayesinde, model karmaşıklığını da kontrol edebiliyor olması, bu sayede sadece eğitim hatasını değil eğitim hatasının yanısıra genelleme hatasını da minimize edebilen bir yöntem olmasıdır.

YRM sınıflandırma problemleri için geliştirilmiş olsa da risk fonksiyonunun minimize edildiği herhangi bir problemde kullanılabilir. Bir model için kullanılacak fonksiyonlar kümesi sonsuz olabilir. Dolayısıyla bu kümenin içinden optimal bir fonksiyonun seçilmesi ancak bir parametrenin rehberliğinde gerçekleştirilebilir. YRM, eğitim verisi ile kapasiteyi eşleştirip, VC katsayısı rehberliğinde fonksiyon seçimine dair yeni bir yaklaşımdır. ARM büyük veri kümelerinde kullanılabilirken YRM küçük veri kümelerinde de kullanılabilir (Emir, 2013:122).

Genelleme sınırını minimize etmek için kullanılacak iki YRM stratejisi vardır:

1. Model karmaşıklığını (VC boyutunu) sabit tutup ampirik riski minimize etmek.
2. Ampirik riski sabit (küçük) tutup VC boyutunu minimize etmek.

Birçok istatistiksel yöntem ve yapay sinir ağları ilk stratejiyi uygularken, sadece destek vektör makineleri ikinci stratejiyi kullanmaktadır (Cherkassky ve Mulier, 2007:123).

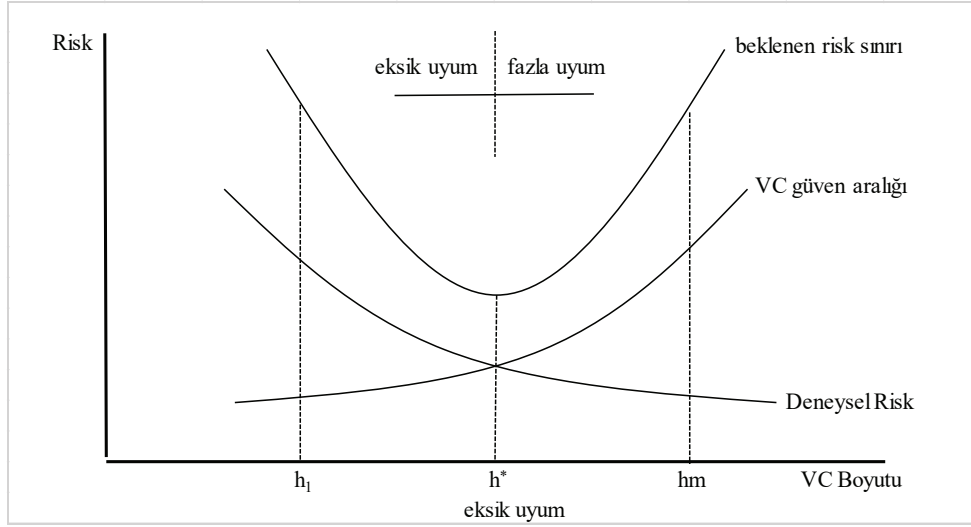
Tümevarımsal öğrenmenin sağladığı en önemli fayda, “doğruluk ve aşırı uyum arasında önemli bir denge kurulması” teorisinin özünü açıklamasıdır. Aşırı öğrenme, bir veri kümesini temsil etmesi için çok zengin bir hipotez kullanıldığında ortaya çıkan bir problemdir. Şimdi, istatistiksel öğrenme teorisinde hipotez alanının zenginliği bir regresyondaki gibi eğrinin derecesi veya bir hipotezin parametre sayısı ile değil, VC boyutu ile kontrol edilir (Corfield vd., 2005:51-52).

VC boyutunda dikkate alarak hesaplanan risk fonksiyoneli ise aşağıdaki gibidir (Çomak, 2008:27):

(4.6)

$$R(\alpha) = R_{emp}(\alpha) + \sqrt{\frac{h \left(\log \left(\frac{2n}{h} \right) + 1 \right) - \log \left(\frac{n}{4} \right)}{n}}$$

Yöntemin tüm bunlar dikkate alındığında risk fonksiyonu yaklaşımı Şekil - 1'de ayrıntılı olarak görülmektedir.

Şekil - 1: VC Boyutu Kullanılarak Beklenen Riskin Minimize Edilmesi

Kaynak: Çomak, a.g.e. s.28.

Şekil - 1'e bakıldığında, VC boyutunu dikkate almayan YSA gibi yöntemlerin, deneysel riski minimize ettikleri nokta, fazla uyuma maruz kalmış, beklenen risk sınırının ise minimize edildiği noktaya denk gelmemektedir. VC boyutunun dikkate alındığı ve deneysel riskle kesiştiği nokta olan ve VC boyutunun h olduğu nokta ise aşırı ve eksik uyumdan kaçınılan ve beklenen risk sınırının minimize edildiği nokta olduğu için model bu karmaşıklık seviyesinde en az hatayı verecektir.

5. MODEL KARMAŞIKLIĞININ KONTROLÜNDE DİKKAT EDİLMESİ GEREKEN UNSURLAR

Belirli bir tekniğin aşırı uyum problemini çözdüğü iddialarına şüpheyle yaklaşılması gerekir. Aşırı öğrenmeyi önlemek için ters sığdırma hatasına (bias) yani eksik öğrenme tuzağına düşmek kolaydır. Her ikisinden, yani hem aşırı öğrenmeden hem eksik öğrenmeden, aynı anda kaçınmak, mükemmel bir sınıflandırıcı öğrenmeyi gerektirir ve önceden bilmeden her zaman en iyisini yapacak bir teknik yoktur (Domingos, 2012, s. 82). Bu yüzden model kurma çalışmalarındaki esas mantık, aslında gerçek hayattaki olayların arkasında bulunan ve tüm akışı belirleyen modele en yakın modeli tespti etmeye dönük bir fonksiyon (model) oluşturma çabasıdır.

Öğrenme algoritmaları sayesinde sınırsız sayıda makine öğrenmesi modeli üretilebilir. Bu algoritmalarından bazıları diğerlerine kıyasla daha çok tercih edilebilmektedir. Bunun önemli bir sebebi hem model olarak uygulaması çok kolay, yani anlaşılabilirliği bakımından çok kolay, hem de bilgisayarlara getirdiği zaman maliyetinin düşüklüğü bakımından tercih sebebidir. Diğer yandan karmaşık algoritmalar ise, diğer verinin içinde yer alan ilişkileri daha basit olan algoritmalara nazaran daha iyi keşfedebilmekte bunlar da modelin ürettiği sonuçların doğruluk oranına olumlu katkıda bulunabilmektedir. Örneğin sınıflandırma konusunda "hangi sınıflandırıcı, neden tercih edilmelidir?", sorusuna verilecek cevap: yöntemin, gereksiz karmaşadan uzak ama doğruluk oranı yüksek, boyut lanetinden etkilenmeyen ama

öğrenme kabiliyeti güçlü, uygulaması basit, az zaman harcayan, güvenilir, tutarlı ve uyarlanabilir olması gibi unsurları sağlamalıdır (Duda vd, 2000:453). Şunu da unutmamak gerekir ki bir modelin uygulaması ne kadar kolay olursa o denli bir yaygın kullanımı da beraberinde getirecektir.

Aşırı öğrenmeden kaçınma, birçok uygulamacı için önemli bir problemdir. Bu problemin adı konmadan önce de düzenleme, budama, ceza şartlarının dâhil edilmesi ve bir açıklama uzunluğunun en iyi hale getirilmesi vb. çeşitli yöntemlerle çözülmeye çalışılmıştır. Uygulama sahasında, kullanılacak olan algoritmayı sınırlama veya detaylarını azaltma gibi bir evrensel eğilim hep söz konusudur. Fakat modelin detaylarını budamak, unutmamak gerekir ki, sonuçlarda hep bir miktar kayıpla sonuçlanacaktır. Bir algoritmayı diğerine tercih etmek için probleme özgü açıklayıcı nedenler yoksa modelin karmaşıklığından kaçınmak evet bir profesyonellik unsuru olsa da meydana getireceği kayıplar daha önemli görülmektedir. “Ampirik başarıyı belirleyen, öğrenme algoritmasının problemle örtüşmesidir; aşırı uyumdan kaçınmanın dayatılması değil”. Ayrıca her model önce en basit şekliyle önerilir. Örneğin YSA perseptron dediğimiz basit algılayıcı ile ilk kez ortaya konulmuş bir problem çözme tekniğidir. Doğrusal bir yapıya sahip olduğu için doğrusal içerikli optimizasyon problemlerini çözebildiği ispanlanmıştır. Fakat perseptron, doğrusal olmayan, meşhur XOR problemini çözemediği gibi örüntü tanıma problemini ise çözmekten çok uzaktır. Hoparlörden açıklayıcı bir ses tanıma işlemi de basit üç katmanlı bir YSA ile çözülemeyen başka bir problemdir. Yani bu modellerin nispeten daha karmaşık (çok katmanlı algılayıcı gibi) formları ihtiyaçlar doğrultusunda geliştirilmiştir. Yoksa model bu şekilde daha etkili olacaktır, şeklinde bir gereksiz beyin fırtınası söz konusu değildir. Fazla kompleks olduğu için problemleri iyi çözemeyen algoritma örnekleri de olmuştur. Sonuç olarak burada daha çok, yapılması gereken, bir modeli nasıl daha basit bir hale dönüştürürüm düşüncesinden çok, problemle en iyi örtüşen mimarinin ya da modelin oluşturulmasıdır (Duda vd., 2000:465). Yani bir bakıma kilolu ve hantal bir gövdeyi, kaslı ve daha esnek bir bünyeye dönüştürmek, evet istenen bir durumdur. Fakat sağlıktan ve işlev kapasitesinden ödün verilmeden yapılırsa, faydalı bir yaklaşım olacaktır.

Güncel yapay zekâ yöntemlerinden destek vektör makineleri ve bu yöntemin temeli olan YRM'nin geliştiricisi Vapnik, teorisini; “önemli olanın sade bir modele değil iyi bir teoriye sahip olmaktır” şeklinde savunmuştur. İstatistiksel öğrenme teorisi çok sağlam bir matematiksel temele sahip olup bu temel, bizzat geliştiricisi Vapnik bunu kitabında ispatlamıştır (Cortes ve Vapnik, 1995:273-297). Yani bir bakıma teorisinin, Occam'ın yaklaşımına bir denge getirdiği, “her şey olabildiğince kolay yapılmalı ama daha basit değil” (Albert Einstein) ölçüsünde bir “model karmaşıklığı kontrolü teorisi” ortaya koyduğunu (matematiksel anlamda) söylemek oldukça mümkün görünmektedir. Popper'ın yanlışlanabilirliği açısından düşündüğümüzde ise, VC teorisinin, bu yaklaşımın sınavından geçtiği ve muhtemelen Popper'ın bu çalışmadan memnun kalacağı Corfield, Schölkopf ve Vapnik tarafından belirtilmektedir (Corfield vd., 2005:51-52). Yöntemin literatürde gözlemlenen başarısı ise bu arka planı doğrular niteliktedir (Korkmaz, 2020:38-39).

Yukarıda ARM ve YRM ilkelerini anlatarak mukayese ettik. ARM'nin ampirik riski minimize etmeye odaklandığını, YRM'nin ise bununla birlikte aynı zamanda (VC katsayısı yardımıyla) model karmaşıklığını da optimize ettiğini belirttik. Burada meseleye felsefi açıdan bakılırsa, ARM evet elle tutulur, kabul görmüş ve hemen her sahada çeşitli varyasyonlarıyla kullanılmakta olan bir yöntemdir.

Fakat unutmamak gerekir ki ARM sonuçta bir “minimizasyon” tekniğidir, yani “optimizasyon” tekniği değildir. Hayatta her zaman bir şeyler maksimize ya da minimize edilebilir fakat muhakkak bu yaklaşımların bize getireceği alternatif maliyetlere de katlanmak gerekecektir. Bütüncül ilerleme için, her zaman “optimal” yaklaşımın dikkate alınması gerekliliğini konumuz açısından ifade edecek olursak, belki de optimizasyon yaklaşımı açısından, önemli bir tartışmayı başlatmış olacağız.

SONUÇ

Model karmaşıklığının ayarlanamayışı ve bunun neden olduğu iki istenmeyen sonuç, eksik öğrenme ve aşırı öğrenme durumları, şimdiye dek birçok araştırmacıyı zorlayan ve modellerin kimi zaman çok iyi kimi zamanda çok yetersiz sonuçlar üretmesi gibi istikrarsızlıklara yol açan önemli bir problemidir. Bu problem için önerilen çözümler çok farklı olmakla birlikte yukarıdaki tartışılan üç başlık altında toplanmıştır. Bunlardan ilk ikisi (Occam’ın usturası ve Popper’in yanlışlanabilirliği) tamamen bilim felsefesi kaynaklı olup, bilimsel yaklaşımları açısından önemli fakat matematiksel zemini olmayan bundan dolayı da, konumuz açısından, muğlak kabul edilen yaklaşımlardır. Üçüncü yaklaşım olan VC teorisi ise diğer iki yaklaşımdan farklı olarak, konuya yaklaşımı sadece felsefi düzeyde kalmayıp, matematiksel çerçeve sunmakta ve bu konuda kuram oluşturma yeterliliği kazandığının söylemek mümkündür. Diğer modeller tarafından yaygın olarak kullanılan ARM yaklaşımının sadece eğitim hatasını minimize edebildiği dolayısıyla genelleme hatasını kontrol edecek bir mekanizmasının bulunmadığı bilinmektedir. DVM’nin ise, bu risk minimizasyonu (ARM) ilkesine VC boyutu sınırlaması getirerek model karmaşıklığının önüne geçtiği (YRM), eğitim hatasını bu sınırlama ile genelleme hatasının minimizasyonu ile orantılı olarak gerçekleştirebildiği yapılan inceleme ile görülmüştür. Bu da sonuçlara yapay sinir ağları gibi, çok sayıda veriye ulaşıldığında çok iyi sonuçlar veren, işlemlerin tek merkezden değil de mimarinin içinde yer alan tüm nöronlara dağıtık halde yapıldığı yöntemlere oranla, özellikle çok sayıda veriye ulaşılamadığında DVM önemli bir alternatif haline gelmektedir.

VC teorisini kullanan tek yapay zekâ yöntemi olan destek vektör makineleri bugün en çok kullanılan sınıflandırma (örüntü tanıma) ve regresyon yöntemlerinden birisidir. Birçok problemde de en iyi sonucu vermekte ve yapay sinir ağlarına da önemli bir alternatif teşkil etmektedir. Meseleye bu yönüyle bakıldığında, VC teorisi bugün, model karmaşıklığının optimizasyonu konusunda, matematiksel temele sahip, ispatlanmış ve kabul görmüş, bilim çevrelerinde paradigma değişikliğinin öncüsü olmuş, ve birçok yeni yöntemde de ilham olabilecek potansiyele sahip kapsamlı bir yöntem olarak araştırmacılara hitap etmektedir. Günümüzde, model karmaşıklığını optimize ederek aşırı öğrenme sorununu teorik anlamda çözebilen ve bu sayede eğitim hatasının yanında genelleme hatasını da minimize edebilen, kapsamlı ikinci bir yöntem bulunmamaktadır.

KAYNAKÇA

- Alp, S., Özdemir, S. Ve Kilitci, A. (2011). Veri Tabanı Yönetim Sistemleri. İstanbul: Türkmen Kitabevi.
- Burges, C.J.C. (1998). "A Tutorial On Support Vector Machines For Pattern Recognition", *Data Mining and Knowledge Discovery*, 2(2), s. 121-167.
- Cherkassky, V. and Mulier, F. (2007). *Lerning From Data: Concepts, Theory and Methods*. John Wiley & Sons, Inc., New Jersey.
- Corfield, D. and Schölkopf, B. and Vapnik, V. (1995). "Popper, Falsification And The VC-Dimension", *Max Planck Institute for Biological Cybernetics*, 145, s. 1-4.
- Cortes, C. and Vapnik, V. (1995). "Support vector networks," *Machine Learning*, 20, s. 273–297.
- Çetin, E. (2016). *Yapay Zekâ Uygulamaları*. 3. Baskı, Seçkin Yayıncılık, Ankara.
- Çomak, E. (2008). "Destek Vektör Makinelerinin Etkin Eğitimi İçin Yeni Yaklaşımlar", Doktora Tezi, Selçuk Üniversitesi.
- Duda, R. O. and Hart, P. E. and Stork, D.G. (2000). *Pattern Classification*. 2. Baskı, John Wiley & Sons, Canada.
- Emir, Ş. (2013). "Yapay Sinir Ağları ve Destek Vektör Makineleri Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması: Borsa Endeks Yönünün Tahmini Üzerine Bir Uygulama", Doktora Tezi, İstanbul Üniversitesi.
- Gürsoy, U.T. (2009). *Veri Madenciliği Ve Bilgi Keşfi*. 1. Baskı, Ankara: Pegem Yayınevi.
- Kelleher, J. D. and Namee, B. M. and D'arcy, A. (2015). *Fundamentals Of Machine Learning For Predictive Data Analytics, Algorithms, Worked Examples, And Case Studies*. 1. Baskı, MIT Press, America.
- Kocamaz, K. (2007). "Hastane Bilgi Yönetim Sistemlerinde Veri Madenciliği Ve Konya Meram Tıp Fakültesi Hastanesindeki Hastane Bilgi Yönetim Sistemi Uygulamasının İncelenmesi", Yüksek Lisans Tezi, Selçuk Üniversitesi.
- Korkmaz, G. (2020). "Kredi Riskinin Belirlenmesinde Yapay Zekâ Yaklaşımları Ve Bir Uygulama", Doktora Tezi, İstanbul Üniversitesi.
- Kühne, T. (2005). "What is a model?", *Dagstuhl Seminar Proceedings*, 04101 Online: <http://drops.dagstuhl.de/opus/volltexte/2005/23> 14/04/2020.
- Öztemel, E. (2012). *Yapay Sinir Ağları*. Papatya Yayıncılık, İstanbul.
- Rasmussen, C. And Zoubin G. (2020). "Occam's Razor", Online: <http://mlg.eng.cam.ac.uk/zoubin/papers/occam.pdf>, 14/04/2020.
- TDK, Genel Türkçe Sözlük, Online: <https://sozluk.gov.tr/?kelime=>, 11/11/2019.
- Vapnik V. N. (1999). "An Overview of Statistical Learning Theory", *IEEE Transactions On Neural Networks*, 10(5).
- Wikipedia, Çevrimiçi: https://tr.wikipedia.org/wiki/Karl_Popper, 14/04/2020.
- Wikipedia, Çevrimiçi: https://tr.wikipedia.org/wiki/Mant%C4%B1ksal_pozitivizm, 07/04/2020.