

Çalışma Notları: Sınıflandırma: İleri Yöntemler (Bölüm 9)

18 Kasım 2025

1 Bölüm 9: Sınıflandırma: İleri Yöntemler (Devam)

1.1 Tembel Öğreniciler (Lazy Learners)

1.1.1 Örnek Tabanlı Akıl Yürütme (Case-Based Reasoning - CBR) (Sayfa 425)

CBR sınıflandırıcıları, yeni problemleri çözmek için bir **problem çözümleri veri tabanı** kullanır.

- **Temsil:** k -En Yakın Komşu yönteminin aksine, CBR, çoklukları veya "örnekleri" **karmaşık sembolik tanımlamalar** olarak depolar.
- **Uygulamalar:** Müşteri hizmetleri yardım masaları, mühendislik, hukuk ve tıbbi teşhis alanlarında kullanılır.
- **Zorluklar:** İyi bir **benzerlik metriği** bulmak ve çözümleri birleştirmek ana zorluklardır. Performansı artırmak için gereksiz örnekler atılabilir (editing method).

1.2 Diğer Sınıflandırma Yöntemleri (Sayfa 426)

1.2.1 Genetik Algoritmalar (Genetic Algorithms) (Sayfa 426)

- Doğal seçim ve genetik mekanizmalardan esinlenen bir **optimizasyon ve arama teknigi**dir.
- Sınıflandırma için **sınıflandırma kuralları** bulmak amacıyla kullanılır.

1.2.2 Kaba Küme Yaklaşımı (Rough Set Approach) (Sayfa 427)

- **Temel Fikir:** Verilen eğitim verileri içinde **denklik sınıfları** (equivalence classes) oluşturulmasına dayanır. Bir denklik sınıfındaki tüm çokluklar, tanımlayıcı öznitelikler açısından **ayırt edilemez** (indiscernible) durumdadır.
- **Yaklaşımalar:** Mevcut öznitelikler açısından tam olarak ayırt edilemeyen sınıfları **yaklaşık** olarak tanımlamak için **Alt Yaklaşım** (Lower Approximation) ve **Üst Yaklaşım** (Upper Approximation) kullanılır.

1.2.3 Bulanık Küme Yaklaşımları (Fuzzy Set Approaches) (Sayfa 428)

- **Avantajı:** Belirsiz veya kesin olmayan olgularla (vague or inexact facts) başa çıkmamızı sağlar.
- **Üyelik Fonksiyonu:** Bir eleman, farklı **üyelik dereceleri** (membership degrees) ile birden fazla bulanık kümeye ait olabilir. Geleneksel olasılık teorisinin aksine, üyelik değerlerinin toplamı 1 olmak zorunda değildir.
- **Kullanım Alanı:** Kural tabanlı sınıflandırma sistemlerinde, sürekli öznitelikler için **keskin sınırları** (sharp cut-offs) önlemede faydalıdır.
- **İşlemler:** Bulanık ölçümelerin birleştirilmesi için operasyonlar (Örneğin, AND → minimum, OR → maksimum) sağlar.

1.3 Sınıflandırma ile İlgili Ek Konular (Sayfa 429 - 436)

1.3.1 Çok Sınıflı Sınıflandırma (Multiclass Classification) (Sayfa 430)

İkiden fazla sınıfın olduğu durumlarda kullanılır (örn: SVM'ler doğal olarak ikili sınıflandırıcıdır).

- **Bire Karşı Hepsi (One-versus-All - OVA):** k sınıf için k ikili sınıflandırıcı oluşturulur.
- **Hata Düzeltici Çıktı Kodları (Error-Correcting Output Codes - ECOC):** Her bir sınıf, benzersiz bir L -bitlik **kod kelimesi** ile ilişkilendirilir. Sınıf ataması, çıktı kod kelimesi ile mevcut sınıf kod kelimeleri arasındaki **Hamming mesafesi** karşılaştırılarak yapılır.

1.3.2 Yarı Denetimli Sınıflandırma (Semi-Supervised Classification) (Sayfa 432)

Etiketli (X_l) ve etiketlenmemiş (X_u) verilerin her ikisi de kullanılır.

- **Öz Eğitim (Self-Training):** Sınıflandırıcı önce etiketli veriyle eğitilir, ardından en güvenilir etiketlenmemiş veriler etiketlenir ve eğitim setine eklenir (Hata takviyesi riski vardır).
- **Birlikte Eğitim (Co-Training):** Özellikler iki karşılıklı dışlayıcı (mutually exclusive) alt kümeye ayrılır ve bu alt kümeler üzerinde iki ayrı sınıflandırıcı eğitilir.

1.3.3 Aktif Öğrenme (Active Learning) (Sayfa 433)

Öğrenme algoritması, bilgi kazancını en üst düzeye çıkarmak için **etiketlenmemiş veri çokluklarından yalnızca birkaçını dikkatlice seçe**r ve bir uzmandan yalnızca bunları etiketlemesini ister.

- **Belirsizlik Örneklemesi (Uncertainty Sampling):** Algoritma, etiketleme konusunda en az emin olduğu çoklukları sorgular.

1.3.4 Transfer Öğrenimi (Transfer Learning) (Sayfa 434)

Bir veya daha fazla **kaynak görevden** (source tasks) bilgi çıkarıp, bu bilgiyi farklı bir **hedef görev'e** (target task) uygulamayı amaçlar.

- **Amaç:** Yeni bir modeli eğitmek için gereken etiketli veri miktarını azaltmaktadır.
- **Yöntem:** **TrAdaBoost** gibi örnek tabanlı yaklaşımlar, kaynak verilerin ağırlıklarını otomatik olarak ayarlar ve hedef verilerden çok farklı olan eski verilerin etkisini filtreler.
- **Zorluk:** **Negatif transfer**, yani bilginin transferinin yeni sınıflandırıcının performansını düşürmesi riski mevcuttur.