

Sınıflandırma Algoritmalarının Karşılaştırması ve Hiperparametre Optimizasyonunun Performans Üzerindeki Etkisi

Yazarlar: Salih Özbek, Emre Karaaydın

Tarih: 18.11.2025

1. Özet

Bu çalışma, birleştirilmiş ve temizlenmiş bir veri seti üzerinde üç temel gözetimli sınıflandırma algoritmasının (Lojistik Regresyon, Karar Ağacı ve Yapay Sinir Ağları) performansını karşılaştırmaktadır. Çalışma ayrıca, Yapay Sinir Ağları (YSA) üzerinde **Varsayılan**, **Manuel Deneme** ve **Grid Search (Izgara Araması)** olmak üzere üç farklı optimizasyon senaryosunun etkisini analiz etmektedir. Analiz, modellerin veri yapısına uyumunu ve hiperparametre optimizasyonunun başarıya etkisini ortaya koymayı amaçlamaktadır.

Temel Bulgular:

- Veri Kalitesi:** Veri birleştirme sonrası yapılan analizde kopya kayıtlar tespit edilmiş ve drop_duplicates işlemi ile temizlenerek modelin yanlış öğrenmesi engellenmiştir.
- Temel Modeller:** Lojistik Regresyon ve Karar Ağacı (Decision Tree) modelleri, ölçeklendirilmiş veri üzerinde hızlı ve kabul edilebilir bir başlangıç performansı sergilemiştir.
- Optimizasyonun Gücü:** YSA modelinde varsayılan parametreler ile başlayan süreç, Grid Search optimizasyonu ile zirveye ulaşmıştır. Özellikle katman yapısı ve öğrenme oranının (learning rate) optimize edilmesi, modelin genelleme yeteneğini artırmıştır.

2. Giriş

Gözetimli öğrenme, etiketli verilerden yola çıkarak girdiler ve çıktılar arasındaki ilişkiyi modelleme sürecidir. Bu çalışmada, data.csv ve labels.csv dosyalarının Unnamed: 0 anahtarı üzerinden birleştirilmesiyle oluşturulmuştur.

Bu çalışmada, literatürün en temel algoritmalarından Lojistik Regresyon ve Karar Ağaçları ile daha karmaşık bir yapı olan Çok Katmanlı Algılayıcı (MLPClassifier) test edilmiştir. Rapor, bu algoritmaların temel çalışma prensiplerini karşılaştıracak ve StandardScaler ile ölçeklendirilmiş veri üzerindeki başarılarını accuracy ve classification_report metrikleri ile analiz edecektir.

3. Algoritmaların Çalışma Prensipleri

Karşılaştırılan üç model, sınıflandırma problemini çözmek için farklı matematiksel yaklaşımlar kullanır.

Özellik	Lojistik Regresyon	Karar Ağacı	Yapay Sinir Ağı (MLP)
Temel Felsefe	Olasılıkçı: Sınıfları ayıran doğrusal bir sınır çizer.	Kural Tabanlı: Veriyi özelliklerine göre hiyerarşik böler.	Biyolojik: İnsan beynindeki nöron çalışma prensibini taklit eder.
Çalışma Mantığı	1. Girdilerin ağırlıklı toplamını alır. 2. Sigmoid fonksiyonu ile sıkıştırır. 3. Sınıf atar.	1. Bilgi kazancını (Entropy) hesaplar. 2. En iyi ayrımı yapan özelliği seçer. 3. Dallandırır.	1. İleri besleme ile tahmin yapar. 2. Hatayı hesaplar. 3. Geri yayılım ile ağırlıkları günceller.
Parametreler	max_iter: 1000 random_state: 42	criterion: 'entropy' random_state: 42	solver: 'adam' learning_rate_init: Değişken

4. Deney Tasarımı

Analiz için veri ön işleme ve modelleme aşamaları titizlikle tasarlanmıştır.

- **Veri Hazırlama:** `pd.merge` ile öznitelik ve etiketler birleştirilmiş, `duplicated().sum()` ile tespit edilen kopyalar atılmıştır.
- **Bölme:** Veri %80 eğitim, %20 test olarak ayrılmış, `stratify=y` ile sınıf dengesi korunmuştur.
- **Ölçeklendirme:** `StandardScaler` kullanılarak veriler standart normal dağılıma (ortalama=0, varyans=1) getirilmiştir.
- **Modeller:** Lojistik Regresyon, Karar Ağacı ve YSA (3 farklı aşamada) test edilmiştir.

5. Bulgular

5.1. Lojistik Regresyon ve Karar Ağacı Analizi

Lojistik Regresyon: Model `x_train_scaled` üzerinde eğitilmiş ve test setinde tahminler üretmiştir. Doğrusal ayrıştırılabilirliği yüksek olan sınıflarda başarılı sonuçlar vermesi beklenmektedir.

Karar Ağacı: `criterion='entropy'` kullanılarak eğitilen model, verideki doğrusal olmayan ilişkileri yakalayabilmiştir. `classification_report` çıktısı, modelin her bir sınıf için kesinlik ve duyarlılık değerlerini ortaya koymuştur.

5.2. Yapay Sinir Ağı (MLP) ve Optimizasyon Analizi

YSA analizi, modelin hiperparametre duyarlılığını ölçmek için üç aşamada gerçekleştirilmiştir.

1. **Varsayılan Model:** `MLPClassifier(max_iter=300)` ile yapılan ilk eğitim, temel bir başarımlı skor sağlamıştır.
2. **Manuel Denemeler:** Farklı katman yapıları [(100,), (50, 50, 50)] ve öğrenme oranları (0.01, 0.001) test edilmiş; derin katmanların karmaşık yapıları öğrenmede daha etkili olduğu gözlemlenmiştir.
3. **Grid Search Sonuçları:** `GridSearchCV` kullanılarak katman boyutları, alfa ve öğrenme oranları taranmıştır. `clf_gs.best_params_` çıktısı, modelin en iyi mimarisini belirlemiş ve test setinde en yüksek (`accuracy_gs`) doğruluğa ulaşılmasını sağlamıştır.

6. Sonuç

Bu analiz, sınıflandırma problemlerinde algoritma seçimi ve optimizasyonun önemini net sonuçlarla ortaya koymuştur.

- **Ön İşlemenin Etkisi:** "Kopya Kontrolü" ve "Ölçeklendirme" (Scaling) adımları, modellerin doğru çalışması için kritik rol oynamıştır.
- **Nihai Sonuç:** Veri seti üzerindeki en yüksek performans, **Grid Search ile optimize edilmiş Yapay Sinir Ağı (MLP)** modeli ile elde edilmiştir. Manuel denemeler fikir verse de, sistematik tarama (Grid Search) en uygun mimariyi bilimsel olarak tespit etmiştir.