

Topluluk (Ensemble) Öğrenme Yöntemleri: Bagging ve Boosting

Atıl Samancıoğlu

1 Giriş

Topluluk (ensemble) teknikleri, birden fazla makine öğrenmesi modelini bir araya getirerek tahmin performansını artırmayı hedefler. Bu yöntemler özellikle Kaggle gibi platformlarda yaygın olarak kullanılır ve genellikle tek bir modelin elde edemeyeceği yüksek doğruluk oranlarına ulaşır.

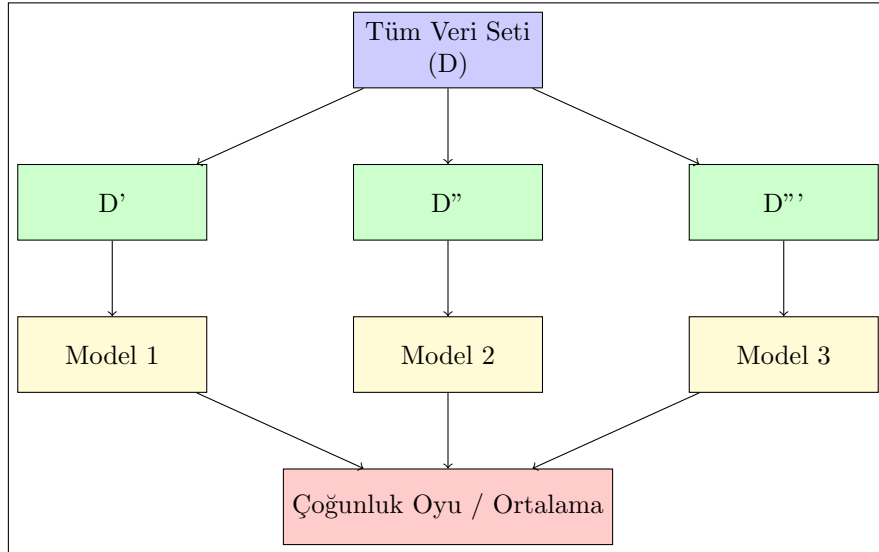
Bu dökümanda iki ana topluluk yöntemi üzerinde duracağız:

- **Bagging (Bootstrap Aggregation)**
- **Boosting**

2 Bagging Nedir?

Bagging, aynı veya farklı türden birden fazla *base learner* (temel model) kullanarak veri setinin farklı alt örnekleri üzerinde bu modelleri eğitir. Her model farklı bir veri örneği üzerinde eğitildiği için çeşitlilik sağlanır. Sonuç olarak:

- Sınıflandırma problemlerinde: **Çoğunluk Oylaması (Majority Voting)** yapılır.
- Regresyon problemlerinde: **Ortalama Alma (Averaging)** yöntemi kullanılır.



Bagging'in Temel Adımları

1. Tüm veri setinden **bootstrap örnekleme** ile alt veri setleri oluşturulur.
2. Her alt veri seti için bir model eğitilir (örneğin: Karar Ağacı, Lojistik Regresyon vb.).
3. Tahmin sırasında tüm modellerin sonuçları birleştirilir.

Bagging’de Paralellik

Bagging yönteminin en önemli özelliklerinden biri, tüm temel modellerin **paralel olarak** eğitilmesidir. Bu, eğitim sürecinin hızlanmasını sağlar.

Örnek Algoritmalar

- **Random Forest:** Bagging yönteminin en bilinen örneklerinden biridir ve temel olarak birden fazla karar ağacını bir araya getirir.

3 Boosting Nedir?

Boosting, zayıf öğrenciler (*weak learners*) olarak adlandırılan modelleri ardışık olarak eğitir. Her model, önceki modelin başarısız olduğu veri örneklerine odaklanır. Böylece hatalar kademeli olarak azaltılır.

Boosting’in Temel Adımları

1. İlk zayıf model tüm veri seti üzerinde eğitilir.
2. Sonraki modeller, önceki modelin yanlış tahmin ettiği veri noktalarına ağırlık verir.
3. Süreç ardışık olarak devam eder ve en sonunda tüm modellerin çıktıları birleştirilerek güçlü bir model oluşturulur.

Boosting’de Ardışıklık

Boosting’in en önemli özelliği, tüm modellerin **ardışık olarak** eğitilmesidir. Her yeni model, önceki modelin hatalarından ders çıkarır.

Örnek Algoritmalar

- AdaBoost
- Gradient Boosting
- XGBoost (Extreme Gradient Boosting)

4 Bagging ve Boosting Karşılaştırması

- **Eğitim Şekli:** Bagging paralel çalışırken, Boosting ardışık çalışır.
- **Hedef:** Bagging varyansı azaltmayı hedefler; Boosting ise bias’ı azaltır.
- **Öğrenici Türü:** Bagging, genellikle güçlü öğrenciler kullanır; Boosting ise zayıf öğrencileri güçlü hale getirir.

5 Sonraki Adımlar

Bir sonraki dökümanda **Random Forest** algoritmasını detaylandıracağız. Random Forest, bagging tekniğine dayalı olarak çalışan ve hem sınıflandırma hem de regresyon problemlerinde yüksek başarı sağlayan bir algoritmadır.

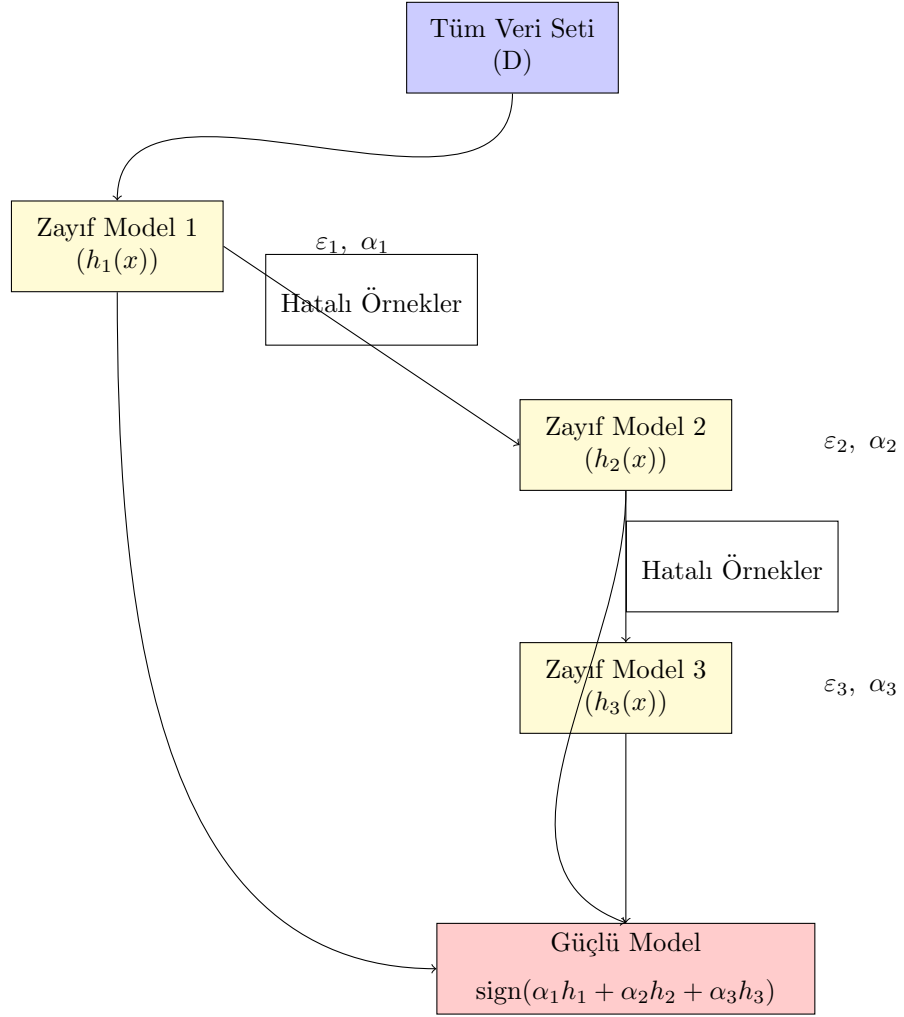


Figure 1: Boosting yönteminin hata katsayıları ve güçlü model ile detaylı görselleştirilmesi