



AdaBoost 심층 분석 보고서 Adaptive Boosting Algorithm Analysis

1. 개요 및 핵심 철학

AdaBoost (Adaptive Boosting)는 1995년 Freund와 Schapire가 개발한 최초의 실용적인 부스팅 알고리즘입니다.

이 알고리즘의 핵심 철학은 "**약한 학습기(Weak Learner)들의 협력**"입니다. 혼자서는 성능이 조금 부족한 모델들을 모아, 서로의 실수를 보완해주며 강력한 성능을 내는 '팀'을 만드는 과정과 같습니다.



핵심 컨셉: "오답 노트" 전략

"이전 모델이 틀린 문제(데이터)를 다음 모델이 집중적으로 공부한다."

AdaBoost는 오분류된 데이터에 **가중치(Weight)**를 부여하여, 후속 모델이 어려운 데이터에 **적응(Adaptive)**하도록 만듭니다.

⚙️ 2. 작동 메커니즘

AdaBoost는 데이터를 순차적으로 학습하며 모델을 정교화합니다. 이 과정은 크게 3단계로 나뉩니다.

STEP 1. 초기화

모든 데이터에 동일한 가중치($\$1/N\$$)를 부여하여 시작합니다.

STEP 2. 반복 학습 (Boosting)

약한 학습기(Stump) 생성 → 에러율 계산 → 틀린 데이터 가중치 증가

STEP 3. 결합 (Voting)

성능이 좋은 모델에 더 큰 투표권을 주어 최종 결과를 판별합니다.

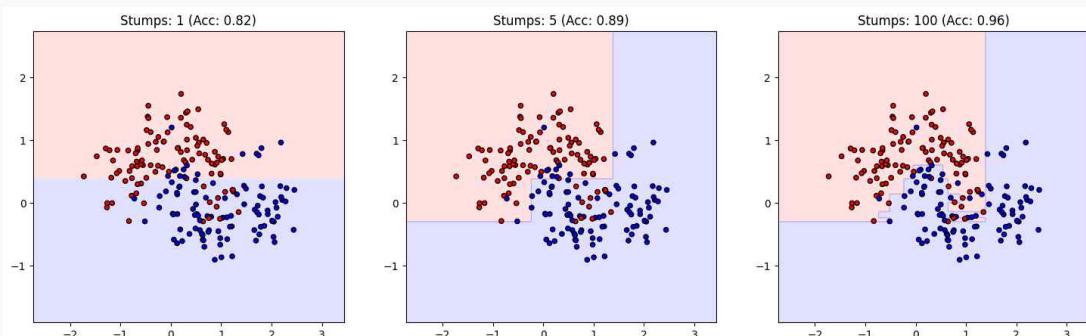


그림 1: 반복 학습을 통한 분류 경계면의 변화

3. Decision Tree vs AdaBoost

일반적인 Decision Tree와 AdaBoost는 모델의 복잡도와 학습 방식에서 큰 차이를 보입니다.

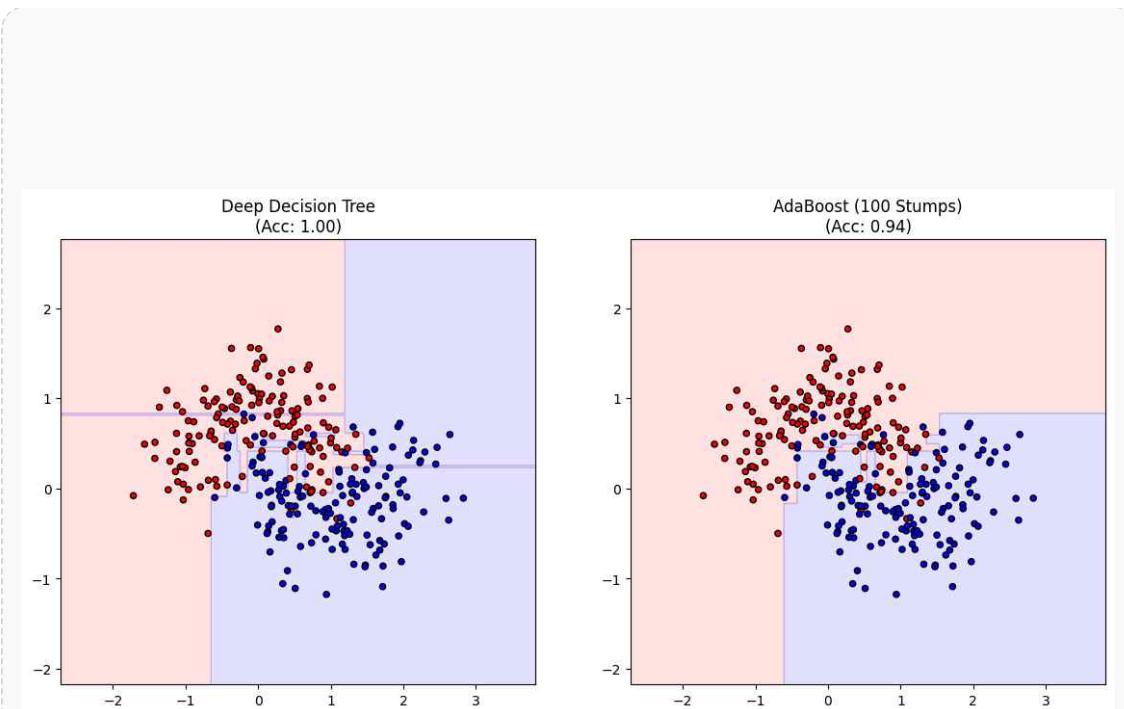


그림 2: Deep Tree vs Stumps 구조 비교

항목	Decision Tree	AdaBoost
모델 형태	깊은 트리 (Deep Tree) 혼자서 복잡한 규칙 생성	스텀프 (Stumps) 깊이가 1인 단순한 트리의 모임
학습 방식	독립적 분할 학습	순차적 종속 학습
과적합	위험 높음 (High Variance)	상대적으로 낮음 (Better Generalization)

4. 장단점 심층 분석

⊕ 장점 (Pros)

- ✓ **강력한 예측 성능:** 여러 모델의 지혜를 모아 단일 모델보다 뛰어난 정확도를 제공합니다.
- ✓ **사용 용이성:** 복잡한 파라미터 튜닝 없이도 기본 성능이 우수합니다.
- ✓ **특성 중요도 파악:** 어떤 변수(Feature)가 중요한지 쉽게 알 수 있습니다.

⚠ 단점 및 주의사항

✗ **이상치(Outlier)에 매우 취약:**

오답에 가중치를 주는 방식 때문에, 노이즈 데이터에 과도하게 집착하여 모델 성능이 떨어질 수 있습니다.

✗ **속도 문제:**

순차적으로 학습해야 하므로 병렬 처리가 불가능해 학습 시간이 오래 걸릴 수 있습니다.

 5. 결론 (Conclusion)

“

AdaBoost는
"실패를 통해 성장하는"
강력한 알고리즘입니다.

”

최종 요약

- ✓ 데이터 품질이 좋고(노이즈 적음), 높은 정확도가 필요할 때 추천합니다.
- ✓ Random Forest와 비교하여 실험해보는 것이 좋습니다.
- ✓ 이상치(Outlier) 제거는 필수적인 전처리 과정입니다.

감사합니다.

Q & A