A low-angle, upward-looking photograph of several modern skyscrapers reaching towards a bright blue sky with scattered white clouds. A small white airplane is visible in the sky, flying towards the center. The buildings have various architectural styles, including glass facades and grid-like patterns.

# Chapter 6 Ensemble Methods

---



## CONTENTS

---

1. Motivation
2. Objective
3. Literature Review
4. Proposed Model
5. Empirical Study
6. Conclusion



# 01.

---

## Motivation

# 오류의 세 가지 원인

- **편향(Bias)**
  - 비현실적 가정에서 비롯
  - 특징과 결과 간의 중요 관계 인식에 큰 영향
  - 과소적합
- **분산(Variance)**
  - 훈련 데이터에 대한 민감도
  - 과적합
- **잡음(Noise)**
  - 예측하지 못한 변화와 측정 오류와 같은 관측값의 분산
  - 설명 불가능

## 오류의 세 가지 원인

- 예측오차
- $E[(y_i - \hat{f}[x_i])^2] = (E[\hat{f}[x_i] - f[x_i]])^2 + V[\hat{f}[x_i]] + \sigma_\epsilon^2$ 
  - $E[(y_i - \hat{f}[x_i])^2]$  : 제곱오차 평균
  - $(E[\hat{f}[x_i] - f[x_i]])$  : 편향
  - $V[\hat{f}[x_i]]$  : 분산
  - $\sigma_\epsilon^2$  : 잡음

# 02.

---

## 부트스트랩 종합

## 부트스트랩 종합

- 정의
  - 부트스트랩
    - Resampling
    - Resampling을 통하여 표본의 성질에 대해서도 추정 가능하다는 것
- 배경
  - Bootstrap aggregating의 줄임말
  - 여러 개의 Bootstrap 자료를 생성하고, 각 자료를 모델링한 후 결합하여 최종 예측 모형 산출
- 부스팅
  - 잘못 분류된 개체들에 집중하여 새로운 분류규칙을 만드는 단계 반복 방법

- 배경

- 방법

- 1. 복원 추출을 통해 N개의 훈련 데이터 세트 생성
    - 2. N 추정기를 통해 각각의 훈련 집합에 적합
      - 각각 독립적으로 적합
    - 3. N 모델로부터 각 개별 예측의 단순 평균(다수결 투표)

- 특징

- 분산 축소
      - 과적합 해결에 도움
      - 최대한 독립적인 표본을 추출할수록 효율적
    - 개선된 정확도
      - N(모델의 수)가 충분히 크다면 배깅을 하는 것이 더 좋다
      - 다만 배깅은 편향을 줄이기보다 분산을 줄이는데 더 탁월함



# 03.

---

## 랜덤 포레스트

- 정의
  - 결정 트리의 합
    - 개별 결정트리는 오버핏될 확률이 높음
  - 랜덤 포레스트는 더 낮은 분산으로 앙상블 예측 시도
- 장점
  - 예측의 분산을 과적합 없이 줄일 수 있음
  - 특징 중요도 계산 가능
- 한계
  - 개별 결정 트리보다 더 분산은 낮으나 편향은 낮지 않을 수 있음

# 04.

---

부스팅

## • ADA 부스트

### • 작동 방식

- 1. 표본 가중값에 따라 복원 추출을 통해 훈련 집합 생성
- 2. 하나의 추정기를 해당 집합에 적합
- 3. 단일 추정기가 허용 임계값보다 높아지면 추정기 유지
- 4. 잘못 분류된 관측값에 더 많은 가중치 부여
- 5. N개 추정기가 생길 때까지 과정 반복

### • 장점

- 예측의 분산과 편향 모두 감소

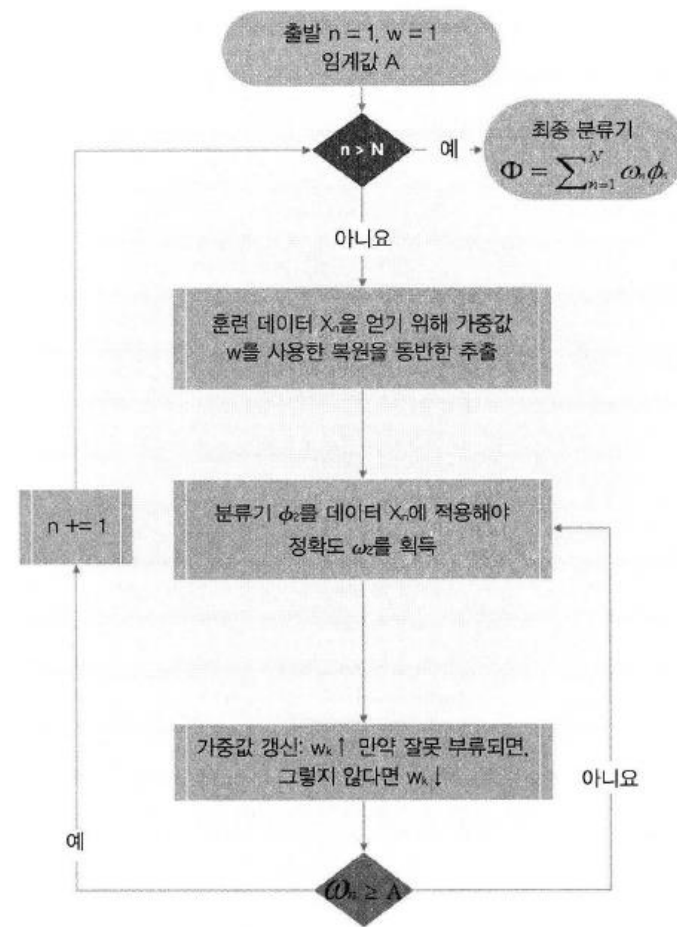


그림 6-3 에이다 부스트 결정 흐름

# 05.

---

## 금융 시장에서 배깅과 부스팅



- 차이
  - 배깅
    - 과적합
  - 부스팅
    - 과소적합
- 금융 시장
  - 일반적으로 과소적합보다 과적합이 더 큰 문제
    - 신호 대비 노이즈가 높기 때문
  - 병렬처리 영역에서 부스팅보다 배깅이 유리

	배깅	부스팅
개별 분류기 적합 방식	동시 적합	순차적 적합
성능 나쁜 분류기	유지	퇴출
가중치 부여 방식	반복별 같은 가중치	반복별 다른 가중치
예측 방식	가중 평균	최종 예측
분산과 편향	분산 감소	분산과 편향 감소

# 06.

---

## 배깅의 확장성

## 배깅의 확장성

- 유명 머신 러닝 알고리즘은 표본 크기로부터 잘 확장되지 않음
  - ex) SVM
    - 알고리즘 수렴 시간이 오래 걸림
    - 전역 최적이라는 보장 없음
    - 과적합에 대한 대안 없음
- 실용적 접근
  - 배깅 알고리즘을 구축할 때, 그 기저 추정기로 SVM처럼 확장성이 높지 않은 것 사용

A low-angle, upward-looking photograph of several modern skyscrapers reaching towards a bright blue sky with scattered white clouds. In the center of the frame, a white commercial airplane is seen flying upwards. The perspective creates a sense of height and grandeur.

Thank you