7주차 숙제

앙상블 학습

정의: 여러 개의 모델을 결합해 더 강력한 모델을 만드는 기법(편향 또는 분산을 줄이는게 목적)

주요 기법: 배깅: 여러 모델을 독립적으로 학습 후 결과를 평균냄

부스팅: 모델을 순차적으로 학습, 오차 보정

투표식 분류기: 직접, 간접 투표: 전자는 다수결, 후자는 예측 확률을 평균냄

배깅과 페이스팅:

| **항목** | **배깅(Bagging)** | **페이스팅(Pasting)** |
| --- | --- | --- |
| 이름 의미 | Bootstrap Aggregating | Pasting (붙여넣기) |
| 샘플링 방식 | **중복 허용** (bootstrap) | **중복 미허용** |
| 목표 | 분산 감소, 과대적합 완화 | 분산 감소 (배깅보다 적음) |
| 대표 모델 | **랜덤 포레스트** | (일반적인 앙상블 기법에서 사용) |
| 구현 난이도 | 매우 널리 사용  구현 쉬움 | 상대적으로 덜 일반적 |

랜덤 패치와 랜덤 서브스페이스:

| **구분** | **랜덤 패치** | **랜덤 서브스페이스** |
| --- | --- | --- |
| **샘플** | **일부만 사용** (샘플도 샘플링) | 전체 사용 |
| **특성(Feature)** | **일부만 사용** (특성도 샘플링) | **일부만 사용** |
| 사용 대상 | 고차원 데이터, 이미지 등 | 고차원 데이터, 텍스트, 이미지 |
| 효과 | **더 높은 다양성**  분산 감소 | 특성간 과적합 방지  일반화 향상 |

두 기법 모두 배깅 방식에서 사용됨.

랜덤 포레스트:

결정트리 + 배깅 + 랜덤 특성 선택

부스팅:

첫 번째 약한 모델 학습 -> 오차에 집중하여 두 번째 모델 학습 -> 애측값 업데이트하고 전과정 다시 반복 -> 모든 모델의 예측을 가중 평균하여 최종 예측