

## 4장

### 1. 분류의 개요

### 2. 결정 트리

- ML 알고리즘 중 직관적으로 이해하기 쉬운 알고리즘
- 데이터에 있는 규칙을 학습을 통해 자동으로 찾아내 트리 기반의 분류 규칙을 만드는 것
- 프로그램에 적용될 if, else를 자동으로 찾아내 예측을 위한 규칙을 만드는 알고리즘
- 규칙노드 → 규칙 조건, 리프 노드 → 결정된 클래스 값
- 새로운 규칙 조건마다 서브 트리 생성
- 트리가 깊어질수록 결정 트리의 예측 성능이 저하될 가능성이 높음
- 가장 적은 결정 노드로 높은 예측 정확도를 가지려면 데이터를 분류할 때 최대한 많은 데이터 세트가 해당 분류에 속할 수 있도록 규칙이 정해져야 함

### 3. 앙상블 학습

- 여러 개의 분류기를 생성하고 그 예측을 결합함으로써 보던 정확한 최종 예측을 도출하는 기법
- 다양한 분류기의 예측 결과를 결합함으로써 단일 분류기보다 신뢰성이 높은 예측값을 얻는 것이 목표
- 보팅: 서로 다른 알고리즘을 가진 분류기를 결합
- 배깅: 각각의 분류기가 모두 같은 기반이지만 데이터 샘플링을 서로 다르게 가져가면서 학습을 수행해 보팅을 하는 것 → 대표적인 방식이 랜덤 포레스트
- 부스팅: 여러 개의 분류기가 순차적으로 학습을 수행하되, 앞에서 학습한 분류기가 예측이 틀린 데이터에 대해서는 올바르게 예측할 수 있도록 다음 분류기에게는 가중치를 부여하면서 학습과 예측을 진행하는 것 → 그래디언트 부스트, XGBoost, LightGBM

### 4. 랜덤 포레스트

- 앙상블 알고리즘 중 비교적 빠른 수행 속도를 가지고 있으며, 다양한 영역에서 높은 예측 성능을 보임
-