SOON CHUN HYANG UNIVERSITY

2024 Probabilistic Model Class

Chapter 7. Application

7.4. Wine Quality Classification

SAIL
Senseable Artificial Intelligence Laboratory

순천향대학교 미래융합기술학과

Senseable AI Lab

석사과정 김병훈

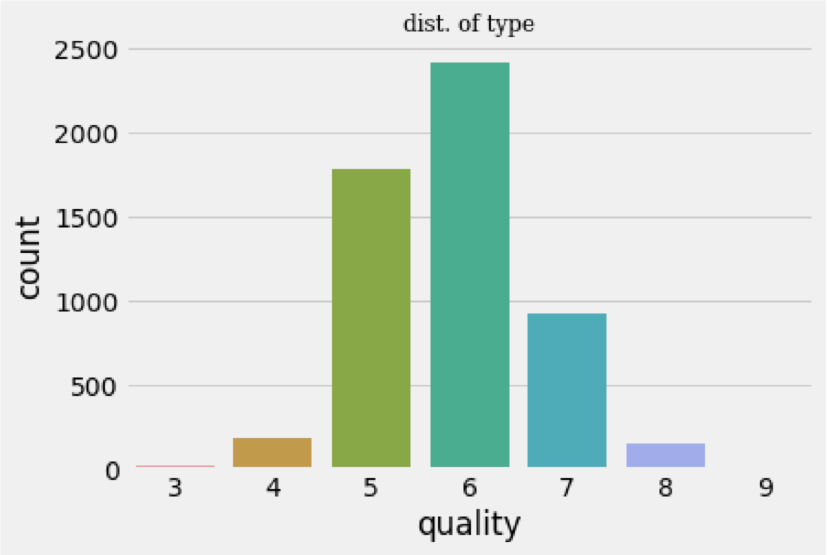
1. Introduction

Wine Quality Classification

- UCI machine-learning repository에서 제공하는 Basic Dataset
- 3~9의 값을 가지는 Quality에 대한 분류
- 두 개의 이진 분류로 단순화(3,4,5,6: Low / 7,8,9: High)

	fixed acidity	volatile acidity	citric acid	...	sulphates	alcohol	type	quality
0	5.6	0.695	0.06	...	0.44	10.2	white	5
1	8.8	0.610	0.14	...	0.59	9.5	red	5
2	7.9	0.210	0.39	...	0.52	10.9	white	5
3	7.0	0.210	0.31	...	0.50	10.8	white	6
4	7.8	0.400	0.26	...	0.43	10.9	white	6

와인 품질 분류 데이터 예시

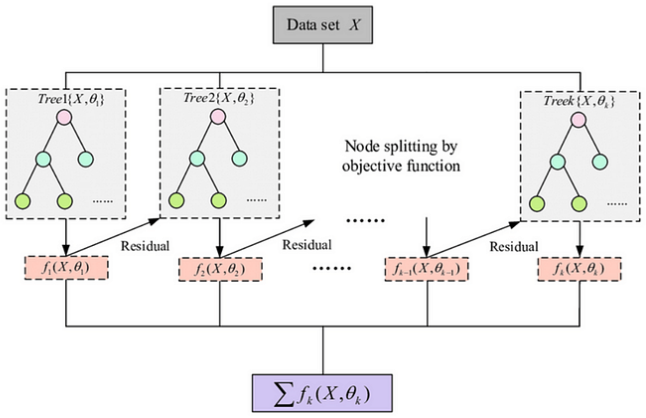


Target 변수 분포 시각화

1. Introduction

XGBoost Model

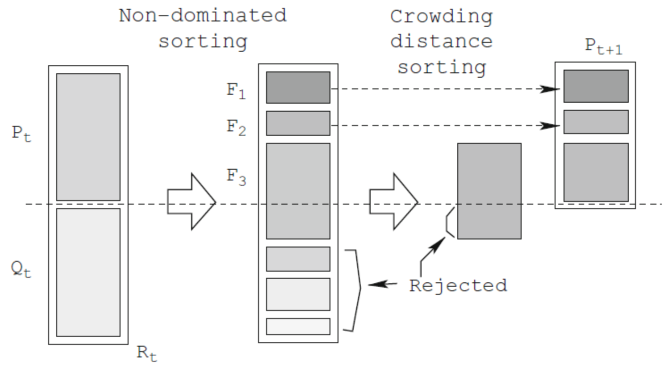
- 그레디언트 부스팅 기반의 의사결정 트리 알고리즘
- 최적의 성능을 위해 하이퍼파라미터에 대한 튜닝 필요



Flow chart of XGBoost

NSGA-II(Non-dominated Sorting Genetic Algorithm II)

- 비지배 정렬과 군집 거리 계산을 통해 랭킹을 매겨 다목적 최적화를 수행
- 하이퍼파라미터 튜닝(성능 최적화)과 Feature selection 두개의 다중 목적을 최적화



2. Methods

Learning Process

1. Initialization : 임의의 초기 Population 생성
2. Pareto Front 생성 : 두 개의 목표에 대해 모든 측면에서 우수하거나 동일한 성능을 보이는 해를 찾음
3. Pareto Front의 Ranking 산출 : 각 솔루션의 랭크를 결정
4. 군집 거리(Crowding Distance) 계산 : 인접한 개체 사이의 차이를 계산
5. Crossover & Mutation : 랭크와 군집 거리에 기반한 선택 / 유전자를 조합하거나 변형

```
NSGAII optimization:
1 evaluations (AUC: 0.78 nr.features:7)
20 evaluations (AUC: 0.76 nr.features:4)
40 evaluations (AUC: 0.8 nr.features:6)
60 evaluations (AUC: 0.78 nr.features:6)
80 evaluations (AUC: 0.63 nr.features:3)
100 evaluations (AUC: 0.77 nr.features:3)
120 evaluations (AUC: 0.77 nr.features:2)
140 evaluations (AUC: 0.75 nr.features:4)
160 evaluations (AUC: 0.5 nr.features:1)
180 evaluations (AUC: 0.8 nr.features:7)
200 evaluations (AUC: 0.77 nr.features:2)
220 evaluations (AUC: 0.79 nr.features:3)
time elapsed: 43.27
```

2. Methods

R Shiny

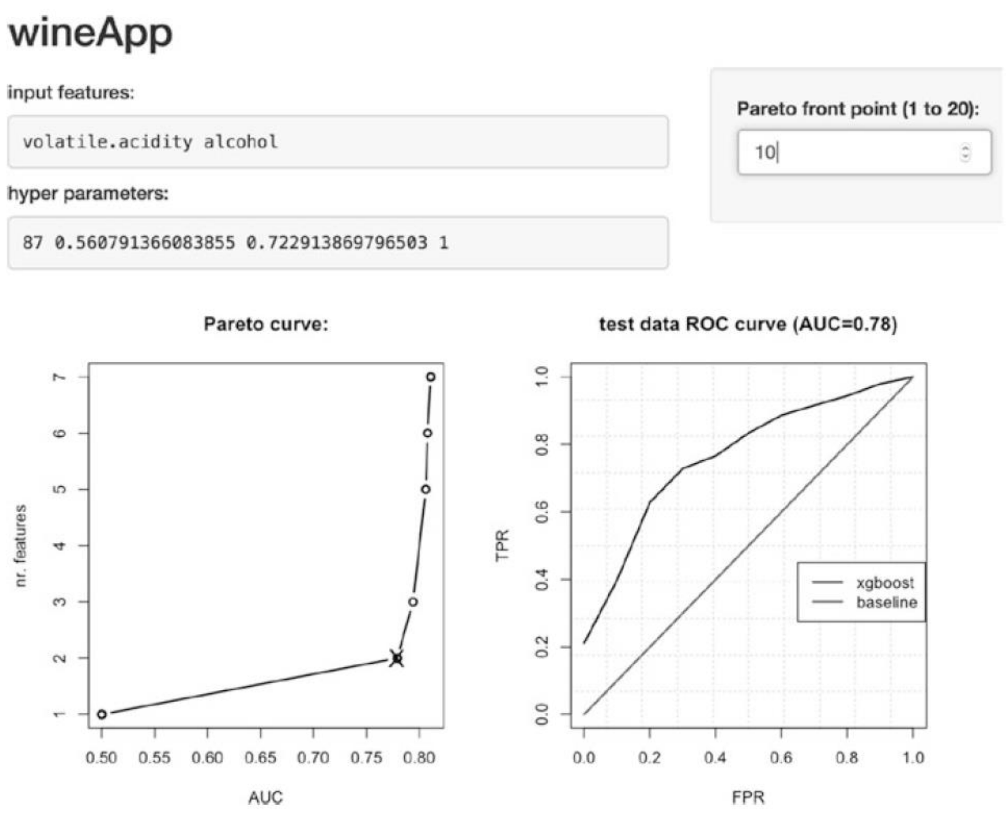


Fig. 7.7 Example of the wineApp execution

3. Comparison of algorithm

```
NSGAII optimization:
1 evaluations (AUC: 0.78 nr.features:7)
20 evaluations (AUC: 0.76 nr.features:4)
40 evaluations (AUC: 0.8 nr.features:6)
60 evaluations (AUC: 0.78 nr.features:6)
80 evaluations (AUC: 0.63 nr.features:3)
100 evaluations (AUC: 0.77 nr.features:3)
120 evaluations (AUC: 0.77 nr.features:2)
140 evaluations (AUC: 0.75 nr.features:4)
160 evaluations (AUC: 0.5 nr.features:1)
180 evaluations (AUC: 0.8 nr.features:7)
200 evaluations (AUC: 0.77 nr.features:2)
220 evaluations (AUC: 0.79 nr.features:3)
time elapsed: 43.27
```

SMS-EMOA optimization:

```
1 evaluations (AUC: 0.78 nr.features:7)
20 evaluations (AUC: 0.76 nr.features:4)
40 evaluations (AUC: 0.61 nr.features:2)
60 evaluations (AUC: 0.81 nr.features:8)
80 evaluations (AUC: 0.62 nr.features:2)
100 evaluations (AUC: 0.8 nr.features:4)
```

Thanks