

Pair Programing

9주차

1조
201984014
성도윤
201904126
허준혁

Q1. 앙상블이란?

A1. 프랑스어의 느낌이 난다

- > 여러개의 예측기로 이루어진 그룹
- > 앙상블 학습은 여러개의 예측기로 이루어진 그룹으로 하니 정확도가 높을 것 같다

Q2. 앙상블 기술의 종류는?

A2. 투표, 배깅, 페이스팅 등등이 있다

Q3. 투표는 무엇일까?

A3. 직접/간접 투표가 있다

- > 직접 투표는 우리가 하는 것처럼 다수결로 정해진다
- > 간접 투표는 예측값으로 정해진다
- > 간접투표를 하는 이유가 뭘까?
- > 직접투표와 비교하면 어떨까?
- > 성능이 더 좋다?
- > 왜?
- > 확률에 비중을 두고, 모델의 예측에 대한 고려와 모델의 신뢰도를 반영할 수 있어서

Q4. 투표식 분류기의 특징은?

A4. 개별 분류기보다 정확한 예측이 가능하다

- > 하지만 무조건 좋을까?
- > 독립성이 매우 중요한 것 같다
- > 파라미터 값을 어떻게 해야 직접/간접 투표 방식을 지정할 수 있지?
- > hard : 직접 투표, soft : 간접 투표

Q5. 배깅, 페이스팅 이란?

A5. 애원(begging) 영단어??

- > 그런줄 알았지만.. bootstrap aggregation
- > 부분집합을 대상으로 학습하는 방식
- > 예측기의 결과를 종합해서 결과 지정

Q6. 분류 모델과 회귀 모델?

A6. 분류 모델은 직접 투표 방식 사용

- > 회귀 모델은 간접 투표 방식을 사용하나?
- > 수집된 예측값들의 평균값 선택이다..

Q7. 랜덤 포레스트란?

A7. 무작위 숲!

-> 앙상블을 최적화한 모델?

-> 편향을 손해보는 대신 분산을 낮추는 것 같다

Q8. 특성 중요도는?

A8. 해당 특성을 사용한 노드가 불순도를 얼마나 감소시키는지 측정

-> 불순도를 많이 줄이면 그만큼 중요도가 클까?

-> 맞다

Q9. 부스팅이란?

A9. 앙상블 기법

-> 성능이 약한 학습기를 여러 개 연결하여 강한 성능의 학습기를 만든다

-> 순차적으로 성능을 높여간다

-> 순차적 학습이라 배깅/페이스팅에 비해 확장성이 떨어질 것 같다

-> 에이다부스트, 그레이디언트 부스팅이 있다