

ml report

learning curve 및 성능 평가 결과를 참고하여 Decision Tree 모델이 오버피팅되었는지 판단해주세요. 판단의 근거를 제시하고, ML 모델에서 오버피팅을 완화할 수 있는 방안을 찾아 함께 작성해주세요.

오버피팅이 되어있는 것 같습니다. training data에서는 거의 100% 정확도를 보여주지만 반대로 validation data에서는 이보다 낮은 80%정도의 정확도를 보여주고 있습니다. 오버피팅이 발생한 경우 traing accuracy와 validateion accuracy가 비슷한 수준으로 근접할때까지 데이터를 추가 하면 오버피팅을 완화할 수 있습니다.

일반적으로 앙상블 모델은 다른 모델에 비해 일반화 성능이 좋습니다. 그 이유가 무엇인지 설명하고, 우리의 성능 평가 결과에서도 XGBoost가 Decision Tree보다 나은 일반화 성능을 보이는지 판단해주세요.

우선 앙상블 모델은 여러 예측 모델들을 결합함에 따라 오버피팅이 일어날 가능성을 줄여줍니다. 특히 데이터가 적거나 복잡한 경우 단일모델에 비해 오버피팅이 일어날 가능성이 적다고 합니다. 또한 앙상블 모델에서는 예측을 결합하는 식으로 작동하기 때문에 각 모델의 개별 오류가 상쇄될 수 있어 일반화 성능이 좋다고 볼수 있습니다. 실제로 과제의 그래프를 보면 XGBoost가 Decision Tree보다 오버피팅이 덜 된것을 확인 할 수 있습니다.

참고 스크린샷

