(TODO 2-1) learning curve 및 성능 평가 결과를 참고하여 Decision Tree 모델이 오버피팅 되었는지 판단해주세요. 판단의 근거를 제시하고, ML 모델에서 오버피팅을 완화할 수 있는 방안을 찾아 함께 작성해주세요.

- functions.py 파일에 구현된 `plot_learning_curve` 의 코드를 바탕으로 learning curve가 의미하는 바가 무엇인지 생각해보세요.
- 오버피팅인지 아닌지의 판단은 성능 평가 결과를 바탕으로 이루어져야 합니다.

(TODO 2-2) 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.

2-1 오버피팅 되었다
학습 데이터 정확도 :
학습 데이터에 대한 정확도가 매운 높음. 학습 정확도가 1.0에 가까운 값에 도달 이는 모델이 학습 데이터의 모든 세부 사항을 지나치게 학습했음을 나타냄. 즉 모델이 학습 데이터에 과적합되었을 가능성이 크다

검증 데이터 정확도 :
검증 데이터에 대한 정확도가 학습 정확도에 비해 상대적으로 낮은 0.8147이다 .
이는 모델이 새로운 데이터에 대해 일반화하지 못함을 의미한다.

2.-1 오버피팅 완화 방안
프루닝 : Decision Tree 모델의 깊이를 제한하거나 가지치기를 통해 트리의 복잡도를 줄인다
앙상블 기법 사용 :여러 개의 트리를 결합하는 랜덤 포레스트나 부스팅 기법을 사용

2-2 일반화 성능이 더 좋은 이유
1. 다양성 : 각 개별 모델이 다른 부분에서 학습하여 다양한 관점을 결합. 이는 모델이 다양한 패턴을 더 잘 학습하게 하여 과적합을 방지함
2. 오류 감소 :여러 모델의 예측을 평균화하거나 투표함으로써 개별 모델의 오류를 상쇄. 이는 예측의 정확도를 높이고 안정성을 증가시킴
3. 강화 학습 :부스팅 기법을 사용하여 각 모델이 이전 모델의 오류를 학습하고 보정. 이는 모델이 점점 더 정확해지도록 함.

XGBoost 가 0.8689로 Decision Tree 0.8147 보다 더 좋은 성능을 보여줬다