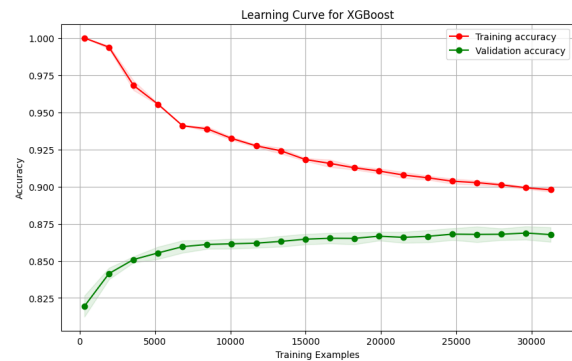
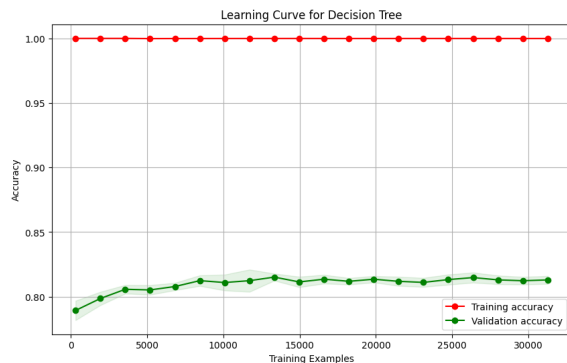


report

(TODO 2-1) Decision Tree 모델의 오버피팅 판단 및 완화 방안

Decision Tree - Training Accuracy:
0.9999, Test Accuracy: 0.8147

XGBoost - Training Accuracy:
0.8933, Test Accuracy: 0.8689



Decision Tree의 learning curve 및 성능 평가 결과를 참고하였을때, training accuracy 가 높는데 test accuracy는 현저히 낮은 것으로 보아 오버피팅 되었다고 판단할 수 있습니다.

오버피팅 완화 방안:

1. 모델 복잡도 제한:

- **트리의 깊이 제한(`max_depth`)**: 트리의 최대 깊이를 제한하여 모델의 복잡도를 줄입니다.
- **리프 노드의 최소 샘플 수(`min_samples_leaf`)**: 리프 노드에 들어갈 최소 샘플 수를 지정하여 모델의 일반화 성능을 향상시킵니다.

2. 데이터 증강 및 샘플링:

- **데이터 증강(Data Augmentation)**: 더 많은 데이터를 수집하거나 기존 데이터를 변형하여 학습 데이터의 양을 증가시킵니다.
- **샘플링 기법**: 학습 데이터에서 일부 샘플을 선택하거나 부스트래핑을 사용하여 모델의 학습을 돕습니다.

3. 앙상블 기법:

- 여러 모델을 결합하여 예측을 수행하는 앙상블 방법(예: 랜덤 포레스트)을 사용하면 단일 Decision Tree의 단점을 보완할 수 있습니다.

(TODO 2-2) 앙상블 모델의 일반화 성능 우수성 및 XGBoost의 평가

앙상블 모델의 일반화 성능이 좋은 이유:

앙상블 모델은 여러 개의 개별 모델을 결합하여 하나의 강력한 모델을 만듭니다. 이는 다음과 같은 이유로 일반화 성능이 뛰어납니다:

1. **다양성**: 서로 다른 모델의 결합은 다양한 관점에서 데이터를 해석하게 하여, 단일 모델의 편향된 해석을 줄입니다.
2. **노이즈 감소**: 개별 모델의 예측 오류를 상쇄시켜 결과적으로 예측의 정확성을 높입니다.
3. **오버피팅 방지**: 개별 모델의 오버피팅을 방지하고 더 일반화된 모델을 생성합니다.

XGBoost의 성능 평가:

학습 샘플이 추가 될수록 training accuracy와 test accuracy의 차이가 줄어들고 최종 성능의 차이도 크지 않은 것으로 보아 XGBoost가 Decision Tree보다 나은 일반화 성능을 보인다고 할 수 있습니다.