


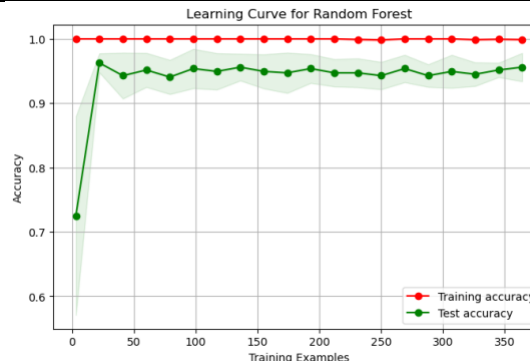
6회차 ML 과제 보고서

26기 김이지

1. 앙상블 기법

(1) 모델 학습 및 성능 평가 [10점]

다음 두 가지 앙상블 기법을 사용하여 모델을 학습했습니다.

Decision Tree	Random Forest
	
<code>`max_depth=5`</code> <code>`random_state=2020147527`</code>	<code>`n_estimators=100`</code> <code>`max_depth=7`</code> <code>`random_state=2020147527`</code>
[Training accuracy] : 0.9934 [Test accuracy] : 0.9035	[Training accuracy] : 0.9978 [Test accuracy] : 0.9561

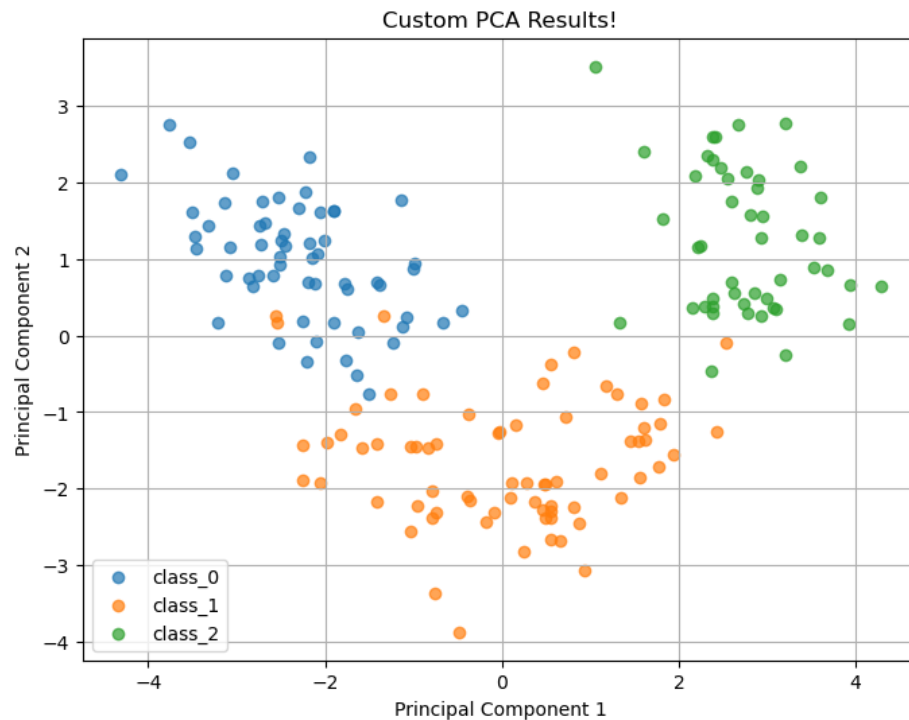
(2) 모델 비교 및 분석 [10점]

더 나은 모델 : **Random Forest**

근거 : 두 모델에서 학습 성능은 비슷한 반면 Decision Tree 보다 Random Forest에서 학습 성능과 테스트 성능의 차이가 작게 나타남. 즉, Random Forest에서 과적합 가능성이 더 낮고, 즉, 일반화 가능성이 더 높음. Random Forest 여러 개의 Decision Tree를 조합해서 사용하기 때문으로 보임.

2. PCA

(1) 시각화 및 결과 [30점]



세 개의 클래스로 비교적 잘 구분되고 있음.

(2) 장점과 단점 및 대체방안 [10점]

장점:

- 차원 축소를 통해 연산 속도 향상
- 데이터 시각화 및 노이즈 감소 효과

단점:

- 일부 정보 손실 가능
- 해석의 어려움 (주성분이 원래 변수와 무관할 수 있음)

대체 방법:

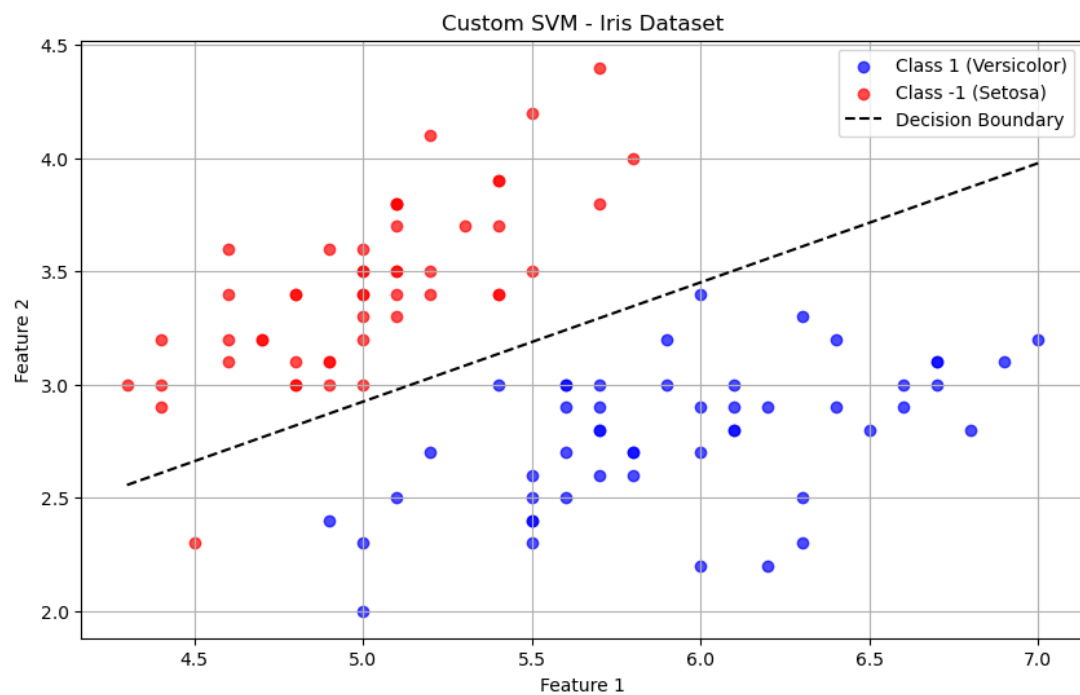
PCA 이외에도 t-SNE, LDA 등의 차원 축소 기법을 사용할 수 있습니다.

3. SVM

(1) 소프트 마진 vs 하드 마진 SVM 비교 [10 점]

	Soft margin SVM	Hard margin SVM
설명	일부 데이터 오차 허용	모든 데이터가 완벽히 분리되어야 함
장점	일반화 성능 우수, 이상치 영향 적음	결정 경계가 명확, 데이터가 완벽히 분리 가능
단점	적절한 매개변수 선택이 중요	이상치에 민감, 과적합 위험

(2) 결과 [30 점]



결정 경계가 두개의 클래스를 비교적 잘 구분하고 있음. 하지만 일부 데이터 포인트가 오분류된 것을 확인할 수 있음.