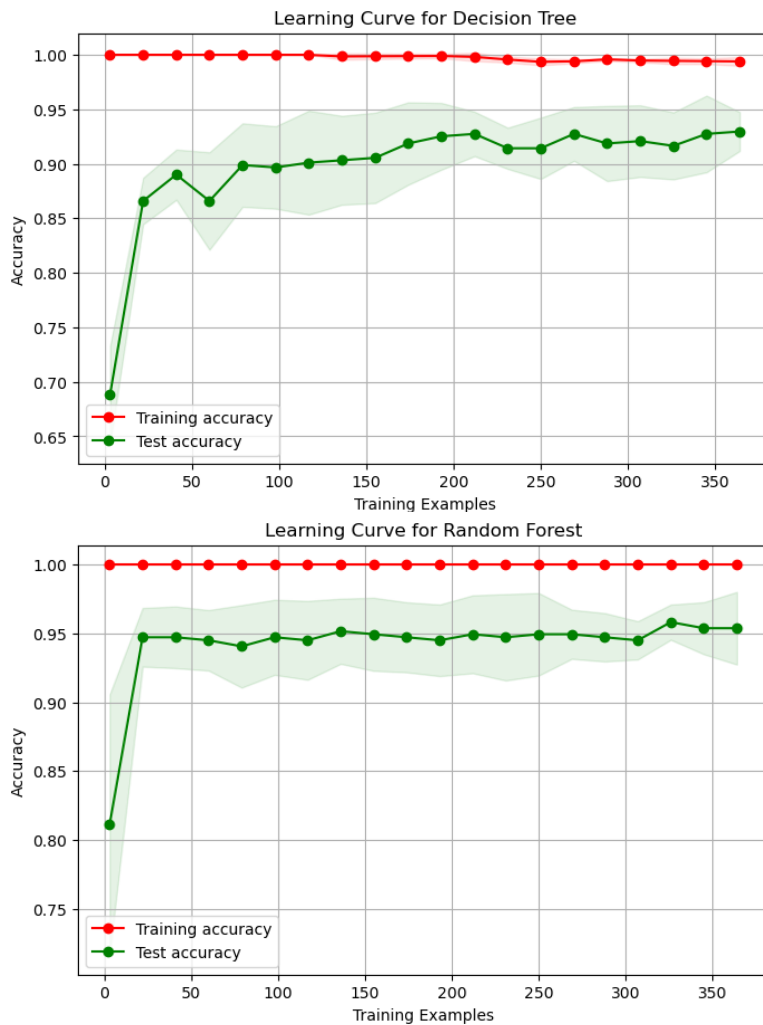


ML report

1.1) Report the graphs as well as the train and test accuracy for both models in the report.



Decision Tree - Training Accuracy: 0.9934, Test Accuracy: 0.9123

Random Forest - Training Accuracy: 1.0000, Test Accuracy: 0.9561

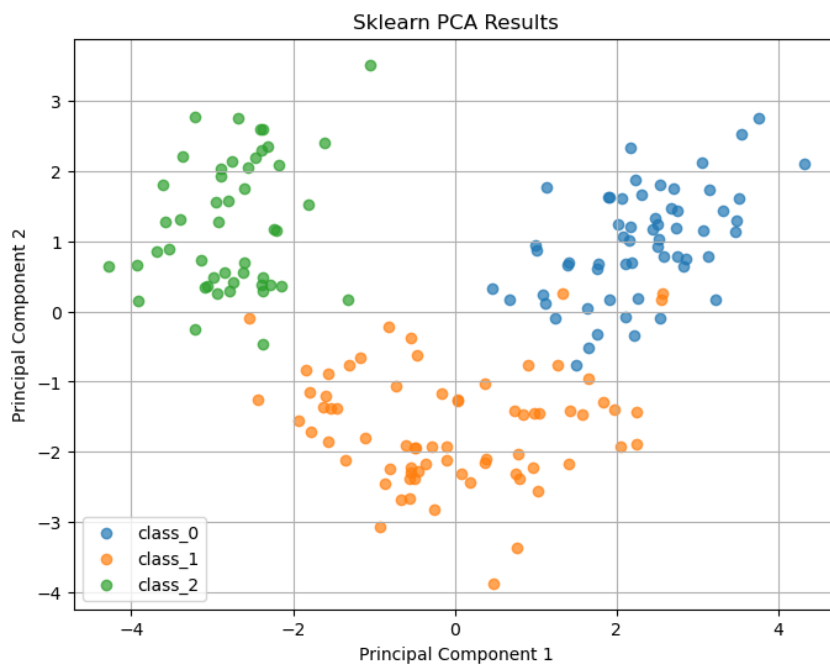
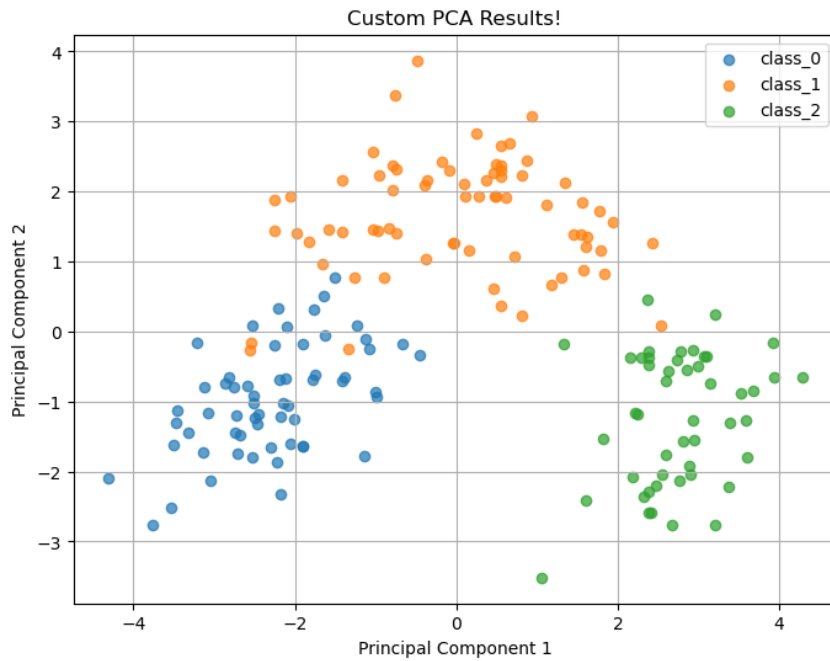
1.2) What is the better model and please provide evidence a supporting arguments that back your decision?

의사결정 트리 모델: 학습 정확도와 테스트 정확도에서 높은 성능을 보였지만, 과적합이 발생하는 경향이 있었습니다.

랜덤 포레스트 모델: 여러 개의 트리로 학습하여 더 안정적인 예측을 제공하고 과적합을 방지할 수 있습니다. 또한 테스트 데이터에 대한 성능이 더 우수했습니다.

랜덤 포레스트 모델이 의사결정 트리 모델보다 더 나은 높은 accuracy를 내기 때문에 더 낫다고 할 수 있습니다. 랜덤 포레스트는 여러 트리의 결과를 종합하여 더 일반화된 성능을 제공합니다.

2.1) Finish the #TODO's in section 2.1) to implement a custom version of the PCA algorithm! Report the plots in the report.pdf!



2.2) [10 Points] What are the benefits and the disadvantages of PCA? Could y provide another dimensionality reduction method that can be used apart from PCA? State your answer in the report.pdf.

- PCA 의 장점:

데이터의 중요한 특성만을 유지하며 차원을 축소할 수 있어 계산 효율성을 높일 수 있습니다.

고차원 데이터에서 노이즈를 줄이고 데이터의 구조를 더 잘 이해할 수 있게 됩니다.

- PCA 의 단점:

선형적인 차원 축소 기법이기에 때문에 비선형 구조를 가진 데이터에 대해서는 성능이 떨어질 수 있습니다.

데이터의 해석이 어려워질 수 있습니다.

- 다른 차원 축소 방법:

t-SNE: 비선형 차원 축소 기법으로 시각화에 유용합니다.

LDA(Linear Discriminant Analysis): 클래스 간 분산을 최대화하여 차원을 축소하는 방법으로 분류 작업에 효과적입니다.

3.1) [10 Points] What is the difference between a soft margin SVM and hard margin SVM? Furthermore, can you provide advantages and disadvantages of both methods? State your answer in the report.pdf.

하드 마진 SVM:

- 데이터가 완벽하게 선형 분리가 가능할 때 사용됩니다. 모든 데이터 포인트가 결정 경계에서 정확히 분류됩니다.
- 장점: 마진을 최대화하여 최고의 분류 경계를 찾습니다.
- 단점: 데이터에 노이즈가 있거나 선형적으로 분리되지 않으면 성능이 저하됩니다. 과적합이 발생할 수 있습니다.

소프트 마진 SVM:

- 데이터에 약간의 오류를 허용하며 마진을 최대화하려는 방식으로 최적화됩니다. 일부 데이터 포인트는 잘못 분류될 수 있습니다.
- 장점: 노이즈에 강하고 실제 데이터에서 더 유연하게 작동합니다.
- 단점: 잘못 분류된 데이터가 있을 수 있으며, 이를 제어하기 위해 하이퍼파라미터(예: C)를 조정해야 합니다.

3.2) [30 Points] Finish the #TODO's in section 3.1) to implement a binary SVM using sub-gradient methods! You only need to fill in the fit function, which is the main function that updates the weights! Please provide the plot in your report.pdf.

