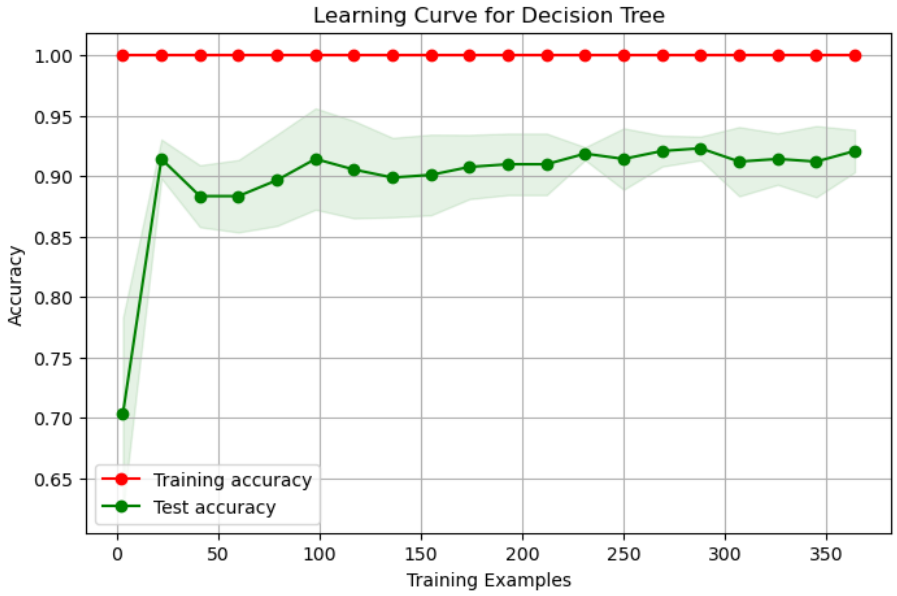
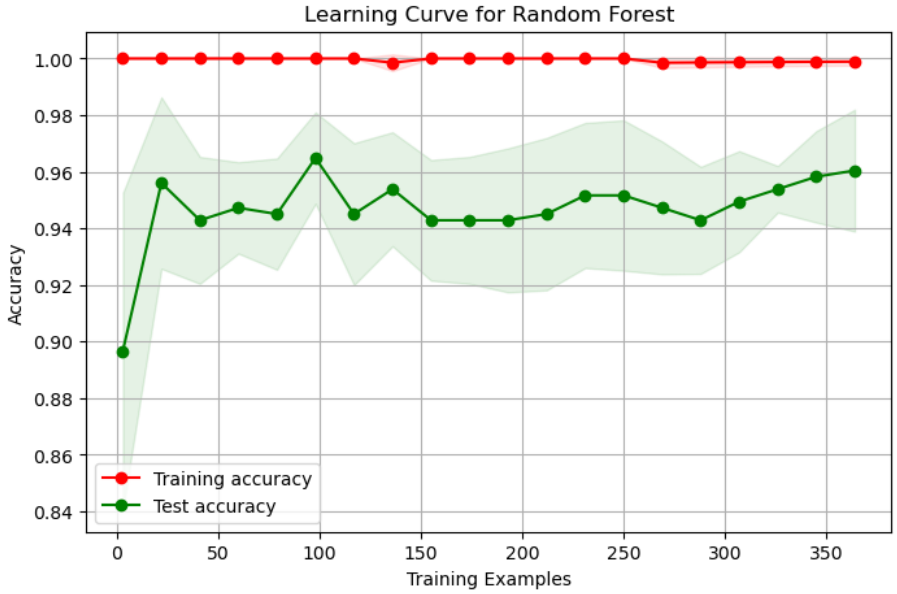
1.1)

The learning curve for two models (Decision tree, Random forest) is shown below:





And the accuracy of each model is as follows:

- Decision tree

  Train accuracy: 1.0000

  Test accuracy: 0.9123

- Random forest

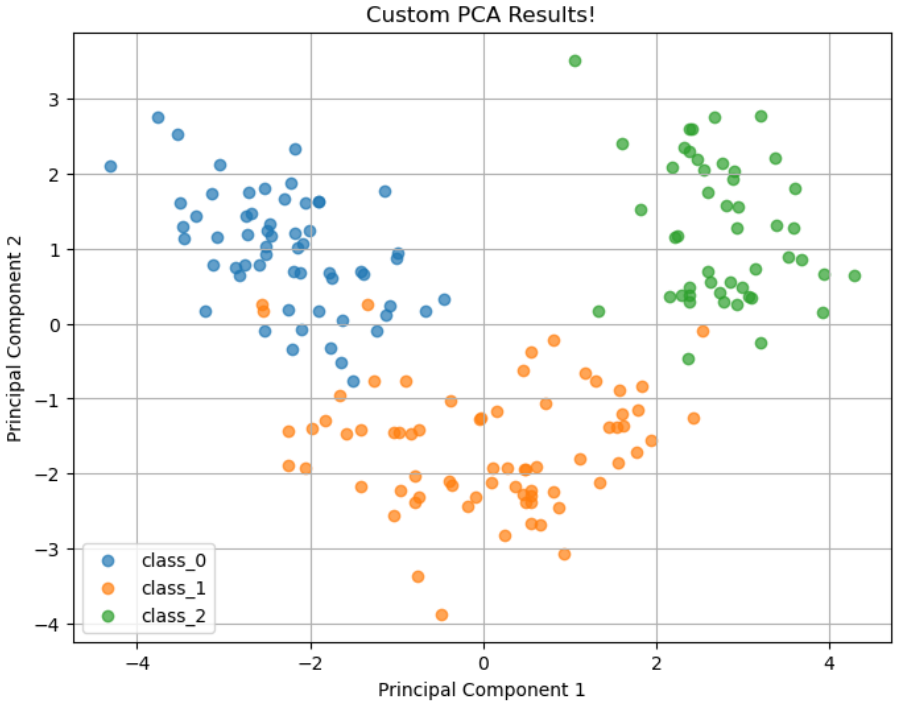
  Train accuracy: 0.9978

  Test accuracy: 0.9649

1.2)

Based on 1.1), random forest is the better model. It achieves a higher test accuracy (0.9649 vs. 0.9123) and is less prone to overfitting. The ensemble mechanism of the random forest helps reduce the variance compared to the single decision tree, thereby improving overall performance and robustness when making predictions on unseen data.

2.1)



2.2)

Benefits of PCA

1) Dimensionality Reduction: PCA projects high-dimensional data onto a lower-dimensional subspace while retaining as much variance as possible. This can improve both model performance and computational efficiency.
   It is based on Eckart-Young theorem, which suggests the closest rank k matrix to A is $A_k = \sigma_1 u_1 v^T + \cdots + \sigma_k u_k v^T$ with rank k.
2) Uncorrelated Features
   The principal components are orthogonal (uncorrelated), which can simplify certain modeling tasks and improve interpretability of variance explained by each component.

Disadvantages of PCA

1) Linear Assumption
   PCA relies on linear transformations. It may fail to capture complex, non-linear relationships in the data.
2) Weak Interpretability
   Principal components are linear combinations of original features and can be difficult to interpret in the context of the original variables.

Alternative Dimensionality Reduction Methods

Autoencoder: Using neural network, autoencoder enables non-linear dimensionality reduction.

Isomap or UMAP: Manifold learning that preserves global or local data structure in non-linear embeddings

3.1)

Hard margin SVM assumes the data is perfectly linearly separable. Hence it enforces zero misclassification during training, in other words, it relies on finding the maximum margin without allowing any margin violations.

Soft margin SVM allows margin violations. It introduces a penalty parameter to control the trade-off between maximizing the margin and minimizing the margin violations.

	Hard Margin SVM	Soft Margin SVM
Advantages	If the data is perfectly separable, the decision boundary is very clear and typically straightforward to interpret.	Robust to noise: allowing some misclassifications prevents fitting to outliers. Flexible adjustment via the penalty parameter C to balance margin width vs. misclassification cost.
Disadvantages	Very sensitive to outliers: even one outlier can prevent the margin from being found	More complex model: need to tune additional hyperparameters (e.g., CCC), which can increase computational cost.

3.2)