Report Assignment - 4 SML

The MNIST dataset is loaded and filtered to include only classes 0, 1, and 2. Each image is flattened from a 2D array into a 1D array.

PCA is implemented to reduce the dimensionality of the dataset to 10 principal components. This step involves standardizing the data, computing the covariance matrix, extracting eigenvalues and eigenvectors, and projecting the data onto the new feature space.

The decision tree is built by recursively splitting the training data based on the feature and split point that result in the lowest Gini index.

1. We iterate over each feature and potential split point within the feature.
2. We divide the dataset into two groups based on the split point and calculate the Gini index for the split.
3. We select the lowest Gini index split as the node's split criterion.
4. We recursively apply the process to each child node until the tree reaches a predetermined maximum depth, resulting in 3 terminal nodes.

To enhance the decision tree model, an ensemble method is implemented :-

1. We create multiple bootstrapped datasets from the original training data.
2. We train a separate decision tree on each bootstrapped dataset.
3. We use majority voting among the ensemble of trees to make final predictions for each test sample.