Written report

Question 1

A challenge is the size of the MNIST dataset, which caused memory issues when applying PCA. To fight this, I explored IncrementalPCA, which lets you process the dataset in smaller batches, reducing memory usage. Lastly, ensuring accuracy took some trial and error. Through these challenges, I learned how important it is to strike a balance between computational efficiency and preserving essential data features.

Question 2

There was a memory and processing time challenge. The kernel PCA can be computationally intensive, especially for large datasets. Using GridSearchCV adds to the complexity since it involves fitting the model multiple times with different hyperparameter. To solve this issue, I used StandardScaler to normalize the data before applying Kernel PCA. This step helps ensure that all features contribute equally to the kernel matrix, which can improve performance and reduce computational burden.