# DINH HOANG VIET PHUONG – 301123263

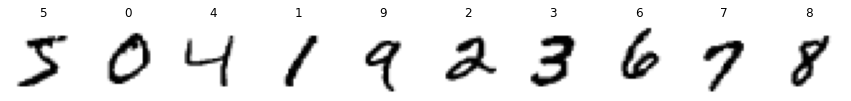
# LAB 02 REPORT

4-9. This section imports necessary libraries.

12-17. This retrieves the mnist\_784 dataset from OpenML, a dataset of hand-written digits. X holds the pixel values for each image, and y contains the corresponding labels (digit values)

20-31. Here, the code creates a dictionary (displayed\_digits) to store one image for each unique digit (0 to 9). It iterates through the labels (y) and saves the first occurrence of each digit.

33-40. This part displays the digits. It creates a figure with 10 subplots, one for each digit, and displays each saved image with its corresponding digit label.



43-52. This section applies PCA to get the first two principal components. The explained variance ratio indicates how much variance in the data each component explains.

Explained variance ratio of 1st principal component: 0.0975

Explained variance ratio of 2nd principal component: 0.0716

55-73. This section plots projections of the data onto the first and second principal components.

A red and blue lines on white paper

Description automatically generated

76-88. Here, the code uses IncrementalPCA to reduce the dataset's dimensionality to 154 principal components. The data is processed in batches, allowing for scalability with larger datasets.

Reduced dataset shape: (70000, 154)

91-104. This section is about visualizing the original and reconstructed (after dimensionality reduction) images of a digit to demonstrate the effectiveness of IncrementalPCA.

106-120. Finally, this section plots the original image next to the reconstructed image to visualize how much detail is preserved after dimensionality reduction.

A close-up of a logo

Description automatically generated