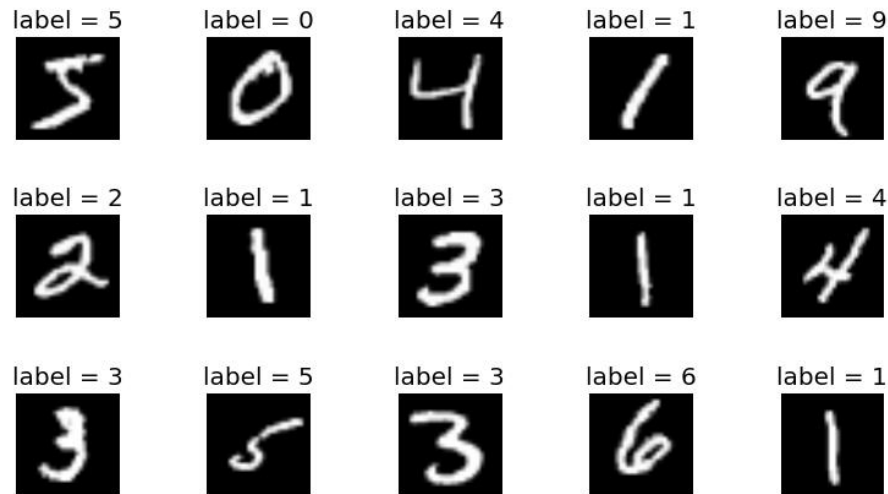


PCA Homework

Deadline: Wednesday April 13, 11:59 PM

The MNIST Dataset is a famous handwritten digit classification dataset that contains 60,000 images of handwritten 0 – 9 digits and 10,000 testing images. Each image has a label 0-9. Here are some sample images from the dataset:

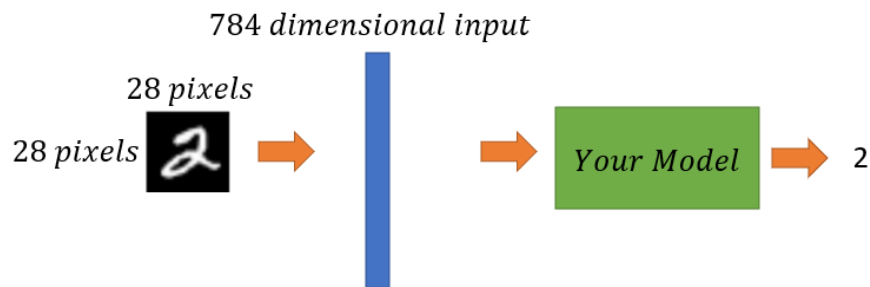


The images are in black and white (so each pixel has 1 value), and have a $28 * 28$ pixel dimension; thus, each image can be represented by a vector of size $28 * 28 = 784$ where each entry is a real number that shows how dark that pixel is (the value of the pixel).

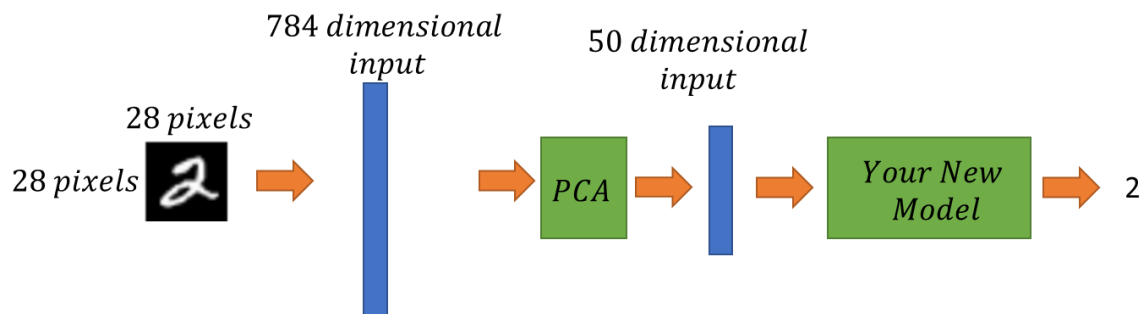
You can find the MNIST Dataset here:

<https://keras.io/api/datasets/mnist/>

Your objective is to develop one classification model (it can be a Decision Tree or Naïve Bayes model or logistic regression, etc.) to map images (X) to their labels (Y). The model is shown here:



Please train your model on the MNIST training set and test it on the MNIST test set and report your test accuracy. Then, use PCA to reduce the dimension of the input (784 to a smaller dimension, for instance, 150 or 100 or 50) and train and test your new model that receives the result of PCA as its input and outputs the label (class) of the image:



To implement the PCA, you can use online packages (e.g., in MATLAB, or Python, ...). For instance, in Python, you can use:

<https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>

and in MATLAB you can use:

<https://www.mathworks.com/help/stats/pca.html>