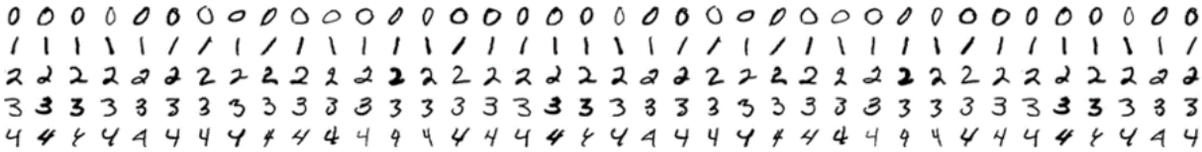
Autoencoder and Clustering

DeadLine: 31 khordad 1402 Assignment 3

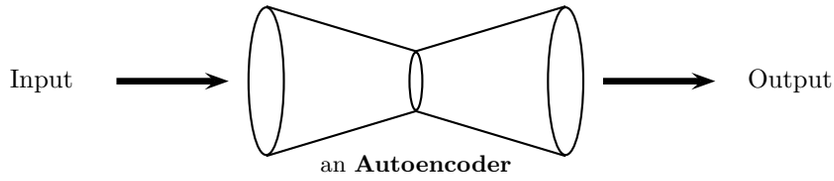
MNIST Images Data Set

The MNIST dataset is consists of a collection of handwritten digit images, where each image is a grayscale 28x28 pixel representation of a single digit ranging from 0 to 9. The dataset is split into two main parts: a training set with 60,000 examples (which splitet to parts itself, 18,000 labeled 42,000 unlabeled) and a test set with 10,000 examples. The dataset's ten classes correspond to the digits 0 to 9.



1 Autoencoder as Feature Extraction

An autoencoder is an unsupervised learning technique for neural networks that learns efficient data representations (encoding) by training the network to ignore signal noise Autoencoders can be used for image denoising, image compression, and in some cases, even generation of image data. The schematic diagram of the Autoencoder is givern in the figure below.



an **Autoencoder**

1. Implementation:
   - Train an autoencoder using the labeled training set from the MNIST dataset.
   - Describe the architecture and hyperparameters used in the autoencoder.
   - Visualize and describe the learned representations or encoded features.
2. Feature Extraction:
   - Apply the trained autoencoder to the labeled data.
   - Extract the encoded features from the labeled data.
   - Discuss the advantages of using autoencoder features for classification tasks.

2 MLP Classifier with Cross-Validation

1. Perform 5-fold cross-validation on the labeled data using an MLP classifier.
2. Evaluation:
   - Train the MLP classifier with the best parameters obtained from cross-validation.
   - Calculate the accuracy of the classifier on the labeled data.
   - Generate a confusion matrix to evaluate the performance of the classifier.

3 Clustering and Combined Classification

1. Clustering:
    - Choose a clustering algorithm (e.g., K-means, DBSCAN) from those covered in the course.
    - Apply the clustering algorithm to the unlabeled trainig set from the MNIST dataset and label them.
    - Discuss the choice of the clustering algorithm and its suitability for the data.
2. Combined Classification:
    - Combine the labeled data and labeled clusters to create a new training dataset.
    - Apply the same MLP classifier used in Part 2 to the new training data.
    - Calculate the accuracy and generate a confusion matrix for the combined classification.

4 Additional Guidance

- Make sure your code is in .ipynb format.
- Along with your code, please include a report file that thoroughly analyzes your results.
- Use appropriate visualizations and statistics to support your analysis and conclusions.