DATA 255 | Homework -3
Deadline: 4/23/2025 (Midnight 11.59 PM)
20 points

Problem 1 (10 pts): Implement and compare different Autoencoder architectures for detecting fraudulent transactions (anomalies) in the Kaggle Credit Card Fraud Detection Dataset. You can drop the 'time' feature. Use **only non-fraud (label = 0)** for training; both fraud and non-fraud for testing. Minimum number of training epochs is 50. You can implement early stopping (callbacks) criteria.

Implement the following Autoencoder variants:

Vanilla Autoencoder: 1 hidden layer (encoder + decoder).

Deep Autoencoder: Minimum 3 layers in encoder and 3 in decoder. Use ReLU activations and dropout.

Sparse Autoencoder: Same architecture as Deep AE. Add L1 regularization to enforce sparsity.

Denoising Autoencoder: Add Gaussian noise to input during training.

Plot distribution of reconstruction error of the test data. Use a threshold on reconstruction error on the test set and report Precision, Recall, and F1-Score.

Problem 2 (8+2 = 10 pts): Apply Variational Autoencoder on the **Fashion MNIST** Dataset. Number of epochs should be minimum of 100.

a. Use minimum of 4 convolutional layers in the encoder and 4 deconvolutional layers (Conv2DTranspose/ upscale) in the decoder.
b. Display how the latent space clusters different classes of the training data.

Useful link-

Fmnist: keras- https://keras.io/api/datasets/fashion_mnist/

Keras example - https://keras.io/examples/generative/vae/

Kaggle Credit Card Fraud Detection Dataset: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud/data

You are required to submit:

1. An MS/PDF/Scanned document:
    a. Include all the steps of your calculations.
    b. Attach screenshots of the code output.
    c. Include the summary of the model
    d. Include a Table - Mention all the hyperparameters you selected: activation function in hidden layer and output layer, weight initializer, number of hidden layers, neurons in hidden layers, loss function, optimizer, number of epochs, batch size, learning rate, evaluation metric
2. Source code:
    a. Python (Jupyter Notebook)
    b. Ensure it is well-organized with comments and proper indentation.

- Failure to submit the source code will result in a deduction of 5 points.
- Format your filenames as follows: "your_last_name_HW2.pdf" for the document and "your_last_name_HW2_source_code.ipynb" for the source code.
- Before submitting the source code, please double-check that it runs without any errors.
- Must submit the files separately.
- Do not compress into a zip file.
- HW submitted more than 24 hours late will not be accepted for credit.