VAE and GRU Assignment

Assignment: Time Series Forecasting and Anomaly Detection with Weather Data

In this assignment, you'll implement both a Variational Autoencoder (VAE) and a Gated Recurrent Unit (GRU) network to analyze weather data. You'll use the VAE for anomaly detection and the GRU for time series forecasting.

Dataset

You'll work with the "Hourly Weather Surface - Brazil" dataset, which contains hourly weather measurements from Brazil including temperature, humidity, wind speed, and other atmospheric conditions.

Part 1: Data Preparation

Download the dataset from:
https://www.kaggle.com/datasets/PROPPG-PPG/hourly-weather-surface-brazil-southeast-region

1. Perform exploratory data analysis:
    - Visualize distributions of key features
    - Create time series plots for temperature, humidity, and wind speed
    - Check for missing values and outliers
2. Preprocess the data:
    - Handle missing values
    - Normalize/standardize features
    - Create sequences for time series forecasting
    - Split into training, validation, and test sets

Part 2: VAE Implementation for Anomaly Detection

1. Implement a VAE architecture with:
    - An encoder with at least 2 hidden layers
    - A latent space dimension of your choice (justify your selection)
    - A decoder with at least 2 hidden layers
    - Proper sampling from the latent distribution
2. Train the VAE on normal weather patterns:
    - Use appropriate loss functions (reconstruction loss + KL divergence)

    - Implement early stopping
    - Visualize the training and validation loss curves
3. Anomaly detection:
    - Define an anomaly score based on reconstruction error
    - Detect anomalous weather days (e.g., extreme temperature changes)
    - Visualize the anomalies on a time series plot
    - Evaluate your anomaly detection performance using precision, recall, and F1-score
    - Consider different threshold values for anomaly detection and justify your choice
4. Latent space analysis:
    - Visualize the latent space (use dimensionality reduction if needed)
    - Analyze how different weather patterns are represented in the latent space
    - Generate new synthetic weather data by sampling from the latent space

Part 3: GRU Implementation for Time Series Forecasting

1. Implement a GRU-based network:
    - Input features should include multiple weather metrics
    - At least one GRU layer with justified number of units
    - Appropriate output layer for forecasting
2. Configure the model for sequence-to-sequence forecasting:
    - Use past N hours to predict the next M hours (choose appropriate N and M)
    - Implement teacher forcing during training (with a decay schedule)
3. Train the GRU model:
    - Implement batch training
    - Use an appropriate loss function for regression
    - Implement learning rate scheduling
    - Visualize training and validation loss
4. Evaluate forecasting performance:
    - Compute MSE, MAE, and RMSE on the test set
    - Visualize predictions vs. actual values
    - Analyze how far into the future the model can predict accurately
5. Ablation study:
    - Compare GRU performance to a simpler LSTM and RNN
    - Analyze how changing the sequence length affects performance
    - Investigate how different features impact prediction accuracy

Part 4: Model Integration and Analysis

1. Integrate the VAE and GRU models:
    - Use the VAE to detect anomalies in the input data
    - Analyze how the GRU model performs on normal vs. anomalous data