Airline Passenger Prediction using LSTM

Introduction

This project aims to forecast the number of airline passengers using a Long Short-Term Memory (LSTM) neural network. LSTM, a type of Recurrent Neural Network (RNN), is suitable for time series forecasting due to its ability to capture long-term dependencies in sequential data. The dataset used is the "Airline Passengers" dataset, which contains monthly totals of international airline passengers from 1949 to 1960.

Data Description

- **Dataset:** `airline-passengers.csv`
- **Description:** The dataset consists of one column containing the number of airline passengers recorded monthly.
- **File Format:** CSV
- **Data Preprocessing:**
  - **Loading Data:** The dataset is loaded into a DataFrame using pandas.
  - **Data Conversion:** The data is converted to a NumPy array and its type is changed to `float32` for consistency.
  - **Normalization:** Data is scaled using MinMaxScaler to transform the values into the range [0, 1], which helps in faster convergence of the model.

Steps Involved

1. **Load and Preprocess Data:**
   - Read the dataset from the CSV file.
   - Convert the data into a NumPy array and scale it using MinMaxScaler.
2. **Create Dataset Matrix:**
   - Split the dataset into training and test sets.
   - Define the `create_dataset` function to transform the dataset into a format suitable for time series forecasting.
3. **Model Building:**
   - Define and compile an LSTM model using Keras.
   - Train the model using the training data.
4. **Make Predictions:**
   - Predict the values for both training and test datasets.
   - Inverse transform the predictions to their original scale.
5. **Evaluate Model:**
   - Calculate the Root Mean Squared Error (RMSE) for both training and test predictions.
6. **Visualize Results:**
   - Plot the original data, training predictions, and test predictions to visualize the model's performance.

Methodology

1. **Data Preparation:**
    - **Scaling:** Normalize the data to improve the model's performance.
    - **Dataset Creation:** Convert the scaled data into a suitable format for the LSTM model by defining a look-back period.
2. **Model Design:**
    - **LSTM Architecture:** Use an LSTM layer with 4 units followed by a Dense layer to output the predicted value.
    - **Compilation:** Use Mean Squared Error (MSE) as the loss function and Adam optimizer for training.
3. **Model Training:**
    - Train the model for 100 epochs with a batch size of 1, and set verbosity to 2 for detailed logging.
4. **Prediction and Evaluation:**
    - **Inverse Transform:** Convert predictions back to the original scale for comparison.
    - **Error Calculation:** Compute RMSE to evaluate the model’s performance.
5. **Visualization:**
    - Plot the original data along with the model’s predictions to visually assess how well the model has learned the temporal patterns.

Future Work

- **Hyperparameter Tuning:** Experiment with different numbers of LSTM units, look-back periods, and epochs to optimize model performance.
- **Additional Features:** Incorporate external factors such as holidays or economic indicators to improve predictions.
- **Model Enhancement:** Explore more advanced architectures like Bidirectional LSTM or Attention mechanisms to capture more complex patterns in the data.
- **Cross-validation:** Implement k-fold cross-validation to ensure the model's robustness and generalizability.

Conclusion

The LSTM model has been successfully implemented to predict airline passenger numbers. The model was able to capture the temporal patterns in the data, as visualized through the comparison of predictions with actual values. The RMSE values indicate the model's prediction accuracy. Further refinement and additional features could potentially improve the model's performance.

Results

- **Training RMSE:** [Calculated RMSE Value]
- **Testing RMSE:** [Calculated RMSE Value]

Summary

This project demonstrates the use of LSTM networks for time series forecasting with airline passenger data. The preprocessing steps, model design, and evaluation metrics provided a comprehensive approach to predicting passenger counts. Future work aims to enhance the model’s accuracy and robustness through various improvements and additional data considerations.