Homework 5

EEE 4774 & 6777 Data Analytics

1 Recurrent Neural Networks

Download the dataset "airline-passengers.csv", which gives the monthly number of airline passengers in units of thousands. We are not interested in the date, given that each observation is separated by the same interval of one month. Hence, you can exclude the first column.

a) [10 pts] Plot the time series data.

b) [10 pts] Explain the short-term and long-term temporal patterns that you see in the data considering that this is the monthly number of airline passengers.

c) [40 pts] Follow the steps below to build an LSTM network using Keras to predict the number of passengers next month given the number this month.

- Fix the random number seed to 0 in numpy for reproducibility.
- Convert the integer values in the data to floating point values, which are more suitable for modeling with a neural network.
- Normalize the data to range $(0, 1)$ since LSTMs are sensitive to the scale of the input data, specifically when the sigmoid or tanh activation functions are used. You can use the MinMaxScaler function in sklearn.
- Split the first 67% of the data as training data and the remaining 33% as testing data.
- Obtain the two-column input-output dataset to be used in LSTM, where the first column is the input (number of passengers in the current month) for each instance (i.e., each row), and the second column is the output (number of passengers next month). In this dataset, the values in the first column will be the same as the values in the second column of previous row.
- Reshape the input data to be in the [samples, time steps, features] format, which LSTM network expects. The time step will be 1 with this dataset since each monthly sample represents a unit time step. You can use the reshape function in numpy with (Xtrain.shape[0], 1, Xtrain.shape[1]) as the new size, where Xtrain is the training input data. Apply the same reshaping to the test data Xtest.

- Build the LSTM network with a hidden layer of 4 neurons, and an output layer with a single neuron. Use sigmoid activation function in the hidden layer, 100 epochs, batch size of 1, "adam" optimizer, and mean-squared-error loss.
- Predict the next month's number of passengers and invert the predictions using the "inverse-transform" attribute of MinMaxScaler to go back to the original unit (thousands of passengers per month).
- Calculate and print the root-mean-squared-error (RMSE) in training and test data.
- Finally, align and plot the actual and predicted values for the entire dataset including training and testing.

d) [40 points] Follow the above steps to build an LSTM network using Keras to predict the next month's value given this month and the previous two months, i.e., input size of 3.