MA3832-Assessment 2

Weighting: 45% *Total marks: 40*

Overview

The assessment covers the content of Week 3-6. It addresses the following learning outcome(s):

- Analyse real world tasks using multi-layer perceptron neural network, ARMA/ARIMA and LSTM for classification and time-series prediction.
- Develop and deploy multi-layer perceptron neural network, ARMA/ARIMA and LSTM in Python
- Tune hyperparameters for neural networks using Python
- Communicate the findings of a formal piece of work and meet a deadline.

Submission

You will need to submit the following:

- A PDF file clearly shows the assignment question, the associated answers, any relevant Python outputs, analyses and discussions.
- Submit Jupyter notebook.
- The task cover sheet

You have up to three attempts to submit your assessment, and only the last submission willbegraded.

.

1 Multi-layer percepton network (15 marks)

1.1 Credit Card Dataset

The data, **"CreditCard_Data.csv"**, is a subset dataset from Yeh and hui Lien (2009). The data contains 10,365 observations and 13 explanatory variables. The response variable, Y, is a binary variable. 1 refers to *default payment* and "0" implies *non-default payment*. The description of 13 explanatory variables is as follows:

- X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.
- X2-X7: Amount of bill statement (NT dollar). X2 = amount of bill statement in September, 2005; X3 = amount of bill statement in August, 2005; . . .; X7 = amount of bill statement in April, 2005.
- X8-X13: Amount of previous payment (NT dollar). X8 = amount paid in September, 2005; X9 = amount paid in August, 2005; . . .;X13 = amount paid in April, 2005.

1.2 Tasks

The goal is to propose a MLP to classify the default payment.

(a) Select 70% of the full dataset as the training data, and retain the remaining as the test dataset. **(1 marks)**

(b) Implement any data wrangling before training a MLP using training data. **(3 marks)**

(c) Propose a neural network model for the default credit classification **(5 marks)**

- Describe the structure of the proposed MLP model. Justify your choice.
- Describe an optimiser and any regularisation techniques implemented in the proposed network.

(d) Report the performance of the proposed MLP on the training dataset. Comment on the results**(2 marks)**

(e) Report the performance of the proposed MLP on the test dataset. Comment on the results **(2 marks)**

(f) Discuss the limitation of your approach and any suggestions to improve the model performance?**(2 marks)**

2 Time series modelling (25 marks)

2.1 Background

The data, **OilPrice.csv**, contains daily Brent oil price from January/2020 to August/2022. The data was collected from Federal Reserve Bank of St. Louis. Our aims are to

- **implement both ARMA/ARIMA and LSTM models to predict oil price;**
- **evaluate the performance of the models in predicting oil price**

The returns of Brent oil price, denoted as y_t, is computed as follows $y_t = ln(Price_t) - ln(Price_{t-1})$ where $Price_t$ is a daily oil price at period t.

2.2 Tasks

(a) Plot oil price and its returns. Comment on the dynamic movement of the Brent oil price and its returns. **(2 marks)**

(b) Proposing an approach to handle with missing values. **(1 marks)**

(c) Use the data up to 29th July 2022 as the training dataset. Propose an ARMA(p,q)/ ARIMA (p,d,q) model to fit the training dataset. Justify your choice. **(4 marks)**

(d) Fit the proposed ARMA/ARIMA to the training data, and then evaluate the forecast performance of the proposed model on the test data. Comment on the performance of the model in predicting Brent oil price. **(3 marks)**

(e) Propose a LSTM model to fit the training dataset. Justify your choice. **(4 marks)**

(f) Train the proposed LSTM model using the training dataset, and then evaluate the forecast performance of the proposed model on the test data. Comment on the performance of the model in predicting Brent oil price. **(3 marks)**

(g) Design a backtesting strategy to evaluate the forecast performance of ARMA/ARIMA and LSTM models for 1 and 2-day-forecast-ahead of Brent oil price for the period of 1/8/2022-15/8/2022. **(4 marks)**

(h) Compare and discuss results obtained from ARMA/ARIMA and LSTM models in part (f). Is there any suggestion that you would like to propose to improve the performance of the models? **(4 marks)**

References

Yeh, I.-C. and hui Lien, C. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36(2, Part 1):2473 – 2480.