UNIVERSITY OF THE WEST OF ENGLAND

Lab 6: Regression and Time Series Analysis

Lesson Intended Learning Outcome:

On completion of this lab, students are expected to be able to:

- Depict and implement the various algorithms of regression; *and*
- Depict and implement the various algorithms of time series analysis.

Lab Exercises

Complete the following exercises using Jupyter Notebook with Python 3.x and Scikit-Learn.

1. Given the dataset "`Diabetes.csv`".
    a) Create a ***correlation matrix*** to show the correlations among all the features. **Hint:** See `LinearRegression.ipynb` .
    b) Divide the given dataset into two sub-groups by people's age: one group with 50 or below and the other above 50. Create a ***Logistic Regression*** model for each of the two groups to predict who has diabetes and to evaluate the two models. You should use 80% of the given dataset as the training set and other as the testing set.

2. Given the dataset "`0388.HK.csv`".
    a) Set the column 'Date' of the dataset as index and extract the columns 'High' and 'Close' for time series modelling.
    b) Create a moving average model on a time window of 100 days on each of the columns 'High' and 'Close' by plotting the result smoothed curve with a lower bound and an upper bound.

**** End of Lab 6 ****