Recitation 4

Data, Inference and Applied Machine Learning

Friday 7 October 2022

Assignment Objectives

- Fit linear regression models
- Make predictions or forecasts
- Evaluate the performance of linear models
- Conduct independent research and report writing
- A sneak peak into time series analysis

Question 1

- Download the two datasets as instructed - FTSE and Housing
- Identify the dependent and independent variables
- Calculate the monthly returns for each variable --***hint r(t) = p(t)/p(t-1) - 1***
- Create the regression model -- ***hint fitlm (MATLAB) or linregress (SciPy)***
- Plot the actual and predicted on a scatter plot
- Calculate the correlation coefficients
- Interpret your results
- Finally, conduct a hypothesis test between the dependent and independent variables

Question 2

- Download the *college.csv* file and extract the necessary columns
- Calculate the correlation coefficients
- Perform ***stepwise linear regression*** on the independent variables
- Identify the predictor variables that are useful in the prediction. Explain why?
- Use ***BIC*** to select the model based on the given independent variables
- Calculate the accuracy of the BIC model versus the stepwise model using only useful variables

Question 2 - cont'd

- Compute the accuracy of the chosen model using the five predictor variables and another one for only the useful variables
- Calculate the graduation rate for CMU with the most accurate model.
- Analyze your result.

Question 3

- Identify your problem statement(the trend you intend to study)
- Provide the source of your data
- List down your assumptions
- Outline your methodology
- Perform the required statistical analysis
- Explain the results from your study
- Predict the situation in 2021

Question 4

- Download data from Quandl(code:”ODA/ISR_LUR”)
- Process the data such as deleting unused rows. You may need to use functions such as datenum to change date formats, other commands are table2cell and table2array.
- Fit linear regression model by using date as the independent variable and unemployment rates as the dependent variable.
- Predict the rate of unemployment by 2020
- Evaluate performance of the model. You may use MAPE

Submission Files (Python)

- **Single** Python code file(.ipynb) - **andrewID_DIAML_AssignmentNo.ipynb**
- Assignment report(.pdf) - **andrewID_DIAML_AssignmentNo.pdf**
  - Indicate the libraries you have used in your code at the beginning of the report (after the title page)
- Data files (as given)

Submission process:

- Put all data files and the source code in a **single folder** named with your **andrewID**
- Zip this folder and submit the zipped **(.zip)** with your report (**.pdf**)to CANVAS

Submission Files (MATLAB)

- **Single** MATLAB code file (.m) - **andrewID_DIAML_AssignmentNo.m**
- Assignment report(.pdf) - **andrewID_DIAML_AssignmentNo.pdf**
- Data files (as given)

Submission process:

- Put all data files and the source code in a **single folder** named with your **andrewID**
- Zip this folder and submit the zipped **(.zip)** with your report (**.pdf**)to CANVAS

Q&A