Recitation 4

Data, Inference and Applied Machine Learning

Friday 03 October 2024

Assignment Objectives

- Fit linear regression models
- Make predictions or forecasts
- Evaluate the performance of linear models
- Conduct independent research and report writing

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 predictions 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 canvas

- Process the data. You may need to convert date column to datetime format, then to numeric afterwards (Hint: use toordinal() in python).

- 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 (make sure the year 2020 is in the same format as x, numeric/toordinal())

- Evaluate performance of the model. You may use MAPE

Resources

Under Files > Resources;

HW4: [Python resources](#)

Submission Process

- 1. Put the source **code file** and **data files** in a single folder
- 2. Name of the folder should be the same as your andrew ID
- 3. **Zip this folder and attach the zipped file on assignment submission page (CANVAS)**
- 4. After attaching zipped file, click on "Add Another File" from assignment submission page
- and **attach your report**
- 5. Submit your assignment

- **N.B. This new process will allow us to compile your reports in Turnitin to check for plagiarism.**

Submission Process

Specific reasons for a submission being classified as incomplete include:

- Failure to correctly name your folder with your Andrew ID,
- Failure to correctly name your report, and code file with andrewID_DIAML_AssignmentNo. For example, mcsharry_DIAML_Assignment1, mcsharry_DIAML_Assignment2 and mcsharry_DIAML_Assignment3.
- A missing report describing the steps, results, and insights
- A missing dataset required for running the code
- A missing code file such as .ipynb or .m file
- An error in the file path needed to run the code

Q&A