MSDS 604 Time Series Analysis

Shan Wang

Fall 2023 Module 2

Homework 3 Due Tuesday Nov 14 11:59pm

Introduction: submit one .pdf file containing all you answers for the homework, including screenshot of python code, output or plot if asked. The .pdf can be converted from Latex file, pictures of your handwriting solutions, word files, markdown files and .etc (anything that can be converted into .pdf). If there are coding problems, only include the answer of the question in the .pdf, and upload a separate notebook for Python code. This homework requires the submission of both pdf file and python notebook.

In time series analysis, to forecast multi-steps in the future, there are two major ways. Assume we have the given data x_0, ..., x_n, and the goal is to forecast the next two steps.

- Method 1: direct multistep forecastin (same as in lecture):
  - select and fit a model using x_0, ..., x_n: $model_1 = arima.fit(x_0, ..., x_n)$
  - the model forecast directly gives the multi-step forecast:
  - $[\hat{x}_{n+1}, \hat{x}_{n+2}] = model_1.forecast(2)$
- Method 2: recursive forecasting
  - select and fit a model using x_0, ..., x_n: $model_1 = arima.fit(x_0, ..., x_n)$
  - predict the first step $\hat{x}_{n+1} = model_1.forecast(1)$
  - fit $model_1$ one more time using $x_0, ..., x_n, \hat{x}_{n+1}$ to get $model_2$
  - predict the second step $\hat{x}_{n+2} = model_2.forecast(1)$

Let's evaluate the performance of these two approaches using an example. The dataset $profit.csv$ recorded the profits (in $\$k$) of an investment product in 200 days (positive number shows increased price compared to original price, negative number shows dropped price from original price). In homework 2, you were asked to identify this data is stationary. Now our goal is to give 3-step forecast $\hat{x}_{n+1}$, $\hat{x}_{n+2}$ and $\hat{x}_{n+3}$. For the given data, leave out the last 3 observations (index 197-199) as your test set (don't touch it until you are told to do so), and use the rest data (index 0-196) as history for ARMA model selection. For all questions, use max p, q=5.

1. We are going to practice method 1 using history dataset.

- Based on the forecast goal, design a 5-fold cross-validation to select the order of ARMA model using the history dataset. Different from the lecture, calculate the RMSE of each forecasting step separately. In the answer, attach a screenshot of the definition of the python function for cross-validation, and fill out the table below:

	1-step	2-step	3-step	avg of 3 steps
Best-RMSE				
Best-Model (p,q)				

2. We are going to practice method 2 using history dataset.

   - Based on the forecast goal, design a 5-fold cross-validation to select the order of ARMA model using the history dataset. In the answer, attach a screenshot of the definition of the python function for cross-validation, and fill out the table below. Hint: in method 2, you need to update the data to fit the model at each step to generate the forecast in the next step.

	1-step	2-step	3-step	avg of 3 steps
Best-RMSE				
Best-Model (p,q)				

3. Take model with the best-RMSE for avg of 3 steps forecasting from each method respectively. Refit each model with all the history data and forecast the next 3 steps (again, follow method 1 and method 2). Report the test RMSE from each method. Which one is better?

4. Draw a plot with the last 20 values of the history data, test data, and test forecasts from two approaches.