Time series Analysis and Modeling

DATS 6450

Homework # 7- Fitting nonlinear dataset into an Autoregressive Model

In this homework, you will use the tute1.csv data set. An autoregressive model will be used to fit the "Sales" data, the dependent variable. The coefficients of the AR model will be estimated using the least square estimate from the following equations:

$$\boxed{\hat{\boldsymbol{a}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}}$$

Where

$$X = \begin{pmatrix} -y(n_a-1) & -y(n_a-2) & \cdots & -y(0) \\ -y(n_a) & -y(n_a-1) & \cdots & -y(1) \\ \vdots & \vdots & \vdots & \vdots \\ -y(n_a+T-1) & -y(n_a+T-2) & \cdots & -y(T) \end{pmatrix}$$

$$\boldsymbol{Y} = \begin{pmatrix} y(n_a) \\ y(n_a+1) \\ \vdots \\ y(n_a+T) \end{pmatrix} \quad \hat{\boldsymbol{a}} = \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_{n_a} \end{pmatrix}$$

Hint: The only challenge here is to know what the order of the AR model is. Since the GPAC concept has not covered yet, you can start the AR order from order 1 and increase it to 5.

Using Pandas library load the "Sales" column of the time series data called "tute1". Pick the order of AR to be 1. Estimate the corresponding coefficient(s). No need for splitting dataset.

1- Display the estimated coefficients on the console.
2- Calculate the Mean Square Error (MSE) and display it with an appropriate message on the console.
3- Calculate Q and display it with an appropriate message on the console. (20 lags)
4- Plot the AFC of residuals. (20 lags)
5- Plot the histogram plot of the residuals.
6- Calculate the mean & variance of residuals and display it with an appropriate message on the console.
7- Plot the estimated values versus the true values with respect to time.
8- Change the order to 2, 3, 4 and 5 and repeat the step 2 through 8.
9- Create a table and put all this information for different order number inside a table.
10- Pick the best AR model order which represents the "Sales" dataset the best. You need to justify why the picked order number makes sense.
11- Write the final AR model with the best order number and the corresponding coefficients as the model that best represents this dataset.

Be ready to upload the **solution report (as a single pdf**) plus **the .py file** through BB by the due date.