MSDS 413, Assignment 9
Non-linear Modeling and Model Monitoring (TS9)

Introduction

For this assignment, you will use the datasets and R script (TS9.R) included in the zip file (TS9.zip), posted to the Module 9 Overview page of Canvas. You will read the data files into R and conduct the requested analyses.

The instructions for submitting your assignment follow the Procedure section below.

The following list defines the data sets and their respective variables.

- AustralianWine.csv (volumes are 1000× Kiloliters
  - Month
  - Fortified
  - Red
  - Rose
  - sparkling
  - Sweet.white
  - Dry.white
- Bike data (Washington, D.C.): `https://www.kaggle.com/c/bike-sharing-demand/dataselect=train.csv`
  Filename: bikes.csv
  - date-time: YYYY-MM-DD HH:MM:SS
  - season: 1, 2, 3, or 4
  - holiday: 0 = not a holiday, 1 = holiday
  - workingday: 0 = not a workday, 1 = workday
  - weather: 4 classes, 1, 2, 3, 4
  - temp: temperature, degrees Celsius
  - atemp: apparent temperature, degrees Celsius
  - humidity: mm Hg
  - windspeed: km/hr
  - casual: 0 = not casual, 1 = casual
  - registered: 0 = not registered, 1 = registered
  - count: total number of bike rentals

Your objective is to explore the time series behavior of these data sets including EDA, modeling, model diagnostics, and interpretation.

Procedure

The following steps are necessary to complete this assignment. Address each and every part and ensure that you cover all the details specified in the questions. Be sure to use all your analytical skills.

1. **Exponential Smoothing** (50 points). For this exercise, use the monthly Australian wine sales *Fortified* data. (Data set: AustralianWine.csv.)
   - 1.1. Perform EDA. Why is a multiplicative seasonal model necessary?
   - 1.2. Forecast the next two years using Holt-Winters multiplicative method.
   - 1.3. Experiment with making the trend exponential and/or damped.
   - 1.4. Compare the RMSE of the one-step forecasts from the various methods. Which do you prefer?
   - 1.5. Now fit each of the following models to the same data using ets:
     - 1.5.1. a multiplicative Holt-Winters method;
     - 1.5.2. an ETS model;
     - 1.5.3. an additive ETS model applied to a Box-Cox transformed series;
     - 1.5.4. a seasonal naive method applied to the Box-Cox transformed series;
     - 1.5.5. a STL decomposition applied to the Box-Cox transformed data plus ETS model applied to the seasonally adjusted (transformed) data.
   - 1.6. For each model, look at the residual diagnostics and compare the forecasts for the next two years. Which do you prefer?
2. **Prophet 1** (25 points) Use the bikes data to forecast bike rentals using Prophet.
   - 2.1. EDA. Then partition the data using the 2011 data as a training data. Partition the first 6 months of the 2012 data as a validation data set. Use the last 6 months of 2012 as a test data set.
   - 2.2. Define the marked holidays by name.
   - 2.3. Construct a Prophet model using the default values for seasonality and changepoints using the training data. Use the holidays as defined above. Describe the model components.
   - 2.4. Forecast the next 6 months from the default model. Plot the training data followed by the forecasted data. Describe the plot.
   - 2.5. Produce forecast accuracy measures based on the validation data, RMSE and MAPE. What do they tell us?
3. **Prophet 2** (25 points) Use the bike data to refine the Prophet model as follows:

3.1. Create a matrix of changepoint flexibiity values, seasonal strength values, holiday strength values, set growth as logistic, and set capacity levels. Describe how the column values in the matrix change relative to each other.

3.2. Using MAPE as the selection criterion, find and list the optimum parameter set from the matrix constructed above.

3.3. Retrain the Prophet model using the training data, the validation data, and the optimum parameters from above. Compare the optimum model components with the default model components.

3.4. Forecast the next 6 months from the optimum model. Plot the data followed by the forecasted data. Describe the plot and compare it to the default model plot.

3.5. Produce, describe, and compare forecast accuracy measures based on the test data, RMSE and MAPE with those of the default model.

Deliverables

Your instructor may modify these and all the following directions. See Section Submission Directions below. The assignment deliverables, each in pdf format, are as follows:

- *Only if requested by instructor*
  - The program or script
  - Logs
  - Outputs
- Mandatory
  Data analysis write-up: no programs, logs, or just code outputs; **complete EDA and model diagnostics are expected unless otherwise instructed; I will be looking for innovative interpretations in the assignments over and above the rote adherence to assignment requirements. Only partial credit will be awarded for rote adherence to assignment requirements.**.

The data analysis must follow and use the item numbering of each assignment, i.e., use the numbers, say, 1 - 5, with the sub-lettering if used. These deliverables are provided according to the instructions in the Submission Directions section below.

Submission Directions

Title Page

Include a title page with your name and the assignment designation. Leave room for instructor comments.

File Names

The assignment write-up file shall be submitted to Canvas according to the schedule in the syllabus using the item (1) naming convention below. The naming convention is case sensitive. Use letters and numbers as given. **The file name parts have no spaces or other separator characters**. TS9Lastname.pdf (submit via Canvas)

The parts are the assignment code, TS9; your last name with only the first letter capitalized; a period, and lastly, the extension "pdf". Generically,

TS9Lastname.pdf

For example: Suppose your name is Student McStats. Your filename then is:

TS9Mcstats.pdf

The analysis write-up file must be submitted for grading. Each write-up requires a title page for instructor comments. The analysis may use either R or any other statistics package you wish, or if you use more than one package, you must use the germane tables, plots, etc., in a single report. If you use more than one package, differences and similarities should be indicated.

email

ONLY IF REQUESTED email your instructor the program (script or code), log and output as separate pdf files. The R log and output may be combined. The file names shall be as follows:

- The program or script file names
  - TS9LastnameRprog.pdf
- The log file names
  - TS9LastnameRlog.pdf