Case Study 2: Bike Rental Forecast

Ling-Chieh Kung*

In the first lecture, we considered bike rental forecast as our motivating example. Now it is time to propose a solution!

1 Tasks

In the MS Excel file "SDA_Fa-16_case2_data.xlsx" provided to you. The spreadsheet "2012" contains the daily rental records of a public bike rental system in a city in 2011 and 2012. We now want to use the historical data in 2011 and 2012 to build a model to forecast bike rentals in the first six months in 2013. The spreadsheet "2013" contains the daily information for the same bike rental system from 2013/1/1 to 2013/6/30. Note that while we have the predicted *temp* and *windspeed*, *atemp* and *humidity* are not available.

You need to design a way to forecast future daily bike rentals. The first task is to construct a forecasting method. It may be a regression formula, a combination of regression and other methods, or have nothing to do with regression. Write down the rationale behind your forecasting method (e.g., you put a squared term of an independent variable after looking at a scatter plot) and some evidence supporting your method (e.g., small p-values, large adjusted R^2, etc.). If your method brings in any managerial implications (e.g., showing some nontrivial relationship between two variables), highlight them.

Then use your method to generate forecast values and put them into the *cnt* column in the sheet "2013." Having the actual rental numbers, the instructing team will check your forecasts according to the following formula:

$$\text{MAE} = \frac{\sum_{i=1}^{181} |f_i - r_i|}{181},$$

*Department of Information Management, National Taiwan University; lckung@ntu.edu.tw.

where f_i and r_i are the forecast value and real value of day i, $i = 1, ..., 181$. Smaller MAE means more a accurate forecast.

Once you construct a way to forecast and obtain a set of values, how may you judge the quality of the forecast? As you do not have the actual values in 2013, you may not compare your forecast values with the actual ones. A typical way is to use the same forecasting method to generate "forecast" values for 2011 and 2012. You then may calculate the MAE for the days in 2011 and 2012. This MAE should not be too high.[1]

2 Submission rules

Each team must submit a written report to address the above tasks. You do not need to include the details of your solution process, but you may want to summarize your procedure. Focus more on the presenting your results. The report, including the cover page (if there is one), cannot be longer than ***12 pages*** (i.e., six double-sided pieces of paper). A report is considered good if it addresses the tasks correctly, precisely, and concisely. Including managerial implications found by analyzing the data is always a plus. Obviously, a well-formatted report is expected.

At the end of the report, a team may specify the amount of works done by each team member. Each team member will be graded by considering the work split. Please note that this is optional and if you think all the team members do the same amount of works and should receive the same grade, you do not need to specify this.

A hard copy of the written report must be submitted in class by ***3:00 pm, December 7, 2016***. Electronic copies of the report (for all teams), slides (for the presenting teams), and forecasting outcome (put in the MS Excel file, for all team) should be uploaded to CEIBA by the same due time. For each team, only one member needs to upload the files. Submissions between 3:00 pm and 4:00 pm on the due date will get 20% off as a penalty. Submissions later than 4:00 pm are not accepted.

The report (for all teams) counts for 8% of the semester grades. The forecasting accuracy counts for 2% (the best performing team gets 2%, the second best gets 1.95%, the third gets 1.9%, etc.). The oral presentation (for presenting teams) counts for 10%.

[1]Hint: Though not easy, it is reasonable to pursue a less-than-1000 MAE for this case.