

Project 1: Forecasting Appliance Energy Usage

Chen Nan

Due date: September 30, 2018

1 Introduction

This project studies predictive models for the energy use of appliances. Data used include measurements of temperature and humidity sensors from a wireless network, weather from a nearby airport station and recorded energy use of lighting fixtures.

The energy (Wh) data is logged every 10 min for the appliances. The 10 min reporting interval was chosen to be able to capture quick changes in energy consumption. Another sub-metered load (lights) is included in the analysis since it has been shown to be a good predictor of room occupancy when combined with relative humidity measurements. The wireless sensor network's temperature and humidity recordings were averaged for the corresponding 10 min periods and merged with the energy data set by date and time. The time span of the data set is 137 days (4.5 months). The energy consumption profile shows a high variability. Although there is no weather station outside the house, weather data for the nearest airport weather station, which is located about 12 km from the house, is merged by date and time in this study to evaluate its impact on the prediction of the energy consumption of appliances. The weather data is at hourly intervals, linear interpolation is used to have a complete data set (at 10 min intervals). The following list presents all the variables or features. From the date/time variable other extra features are generated: the number of seconds from midnight for each day (NSM), the week status (weekend or workday) and the day of the week.

1. date time: year-month-day minute:second
2. Appliances: energy use in Wh (**Response to be predicted**)
3. lights: energy use of light fixtures in the house in Wh
4. T1: Temperature in kitchen area, in Celsius
5. RH1: Humidity in kitchen area, in %
6. T2: Temperature in living room area, in Celsius
7. RH2: Humidity in living room area, in %

8. T3: Temperature in laundry room area
9. RH3, Humidity in laundry room area, in %
10. T4: Temperature in office room, in Celsius
11. RH4: Humidity in office room, in %
12. T5: Temperature in bathroom, in Celsius
13. RH5: Humidity in bathroom, in %
14. T6: Temperature outside the building (north side), in Celsius
15. RH6: Humidity outside the building (north side), in %
16. T7: Temperature in ironing room , in Celsius
17. RH7: Humidity in ironing room, in %
18. T8: Temperature in teenager room 2, in Celsius
19. RH8: Humidity in teenager room 2, in %
20. T9: Temperature in parents room, in Celsius
21. RH9: Humidity in parents room, in %
22. To: Temperature outside (from weather station), in Celsius
23. Pressure (from weather station), in mm Hg
24. RH_{out}: Humidity outside (from weather station), in %
25. Windspeed (from weather station): in m/s
26. Visibility (from weather station): in km
27. Tdewpoint (from weather station): Celsius

Since the dataset contains several features or parameters and considering that the airport weather station is not at the same location as the house, it is also desirable to find out which parameters are the most important and which ones do not improve the prediction of the appliances' energy consumption.

2 Dataset

The project has two datasets. The training dataset is for model estimation and model selection. It includes 14803 samples. Each sample has one output value (i.e., energy use in Wh), and 26 features defined in the previous section.

The testing dataset has the same format, except that the output values are not provided. You are expected to predict the output values for each sample in the test dataset. The accuracies of your prediction will be evaluated based on the numbers you provided.

3 Project Assignment

Your task is to build a regression model to predict the appliance energy use in Wh for the given 10 mins interval.

3.1 Step 1: Simple Regression Model

In this part, you are required to develop a simple model that can be used for predicting the energy usage. To reduce the difficulty, you are allowed only limited manipulations of the original data set. You are allowed to take power transformations of the original variables (square roots, logs, inverses, squares, etc), but you are *NOT* allowed to create interaction variables. Your model should include *NO* more than 5 predictors/covariates, but should explain as much variability as possible.

After obtaining the model with aforementioned features, you are required to analyze the model and provide meaningful interpretations. Please focus your attention on the interpretation of the model. A strong analysis should include the interpretation of various coefficients, statistics, and plots associated with their model and the verification of any necessary assumptions.

3.2 Step 2: Complex Regression Model

In this part, you are free to construct the “best” regression model for predicting energy consumption. You are encouraged to experiment with any of the methods that were discussed during the semester for finding a suitable model. You are allowed to create any new variables you desire (such as quadratic, interaction, or indicator variables). Your model needs to be estimated based on the training data, and provides prediction on the testing data. Forecast errors will be evaluated as a component of your project score.

Note: You are allowed to construct multiple regression models to make the forecasting. Only the final forecasting results should be submitted for evaluation.

3.3 Step 3: (Optional) Free Form Model

You can choose any arbitrary model, including but not limited to regression models, for prediction purpose. If you choose to do this part, you need to summarize the method you choose, report the results, and compare the results with regression models in your report. The forecasting accuracy from this model will be evaluated. If your accuracy is better than that of the best regression model in the class, you will be awarded *3 bonus points*.

Attention: Make sure your results are replicable by the codes you submitted. Unreproducible results are considered cheating/plagiarism.