Carl Morgenstern

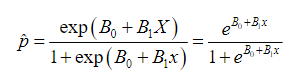
Project Phase 1: Final Project Abstract for Machine Learning

EV identification and forecasting

The goal of this project is to be able to recognize whether an (or multiple) Electric vehicles (EV) are at a customers’ home, based on the customer load data. Once an EV is identified, the charging pattern of the customer can be found, to be able to forecast when the customer load is going to spike. This is important to know, because if utilities know where all the EVs are and their normal charging pattern, they can better forecast the load, and run analysis to see if a transformer may become overloaded. Once the car charging becomes “smarter” this knowledge could also be used for utilities to schedule cars to charge at different times, and maybe even discharge to pick up load. In later research I am planning on extracting locations and parameters from battery systems, so this project will give me a good introduction to a lot of the techniques I will be using in the future. The data required for this project is available here: <https://data.nrel.gov/submissions/69>.

Luckily, the data is labeled, so it allows for supervised learning. When an EV is charging, the load spikes for 15-30 minutes and then drops back to normal, so this should be a pretty linear relationship when using the 5-minute intervals available in the data. Once the EV has been confidently identified, I could use another supervised machine learning algorithm, trained by the habits of the customer to forecast the following move. Supervised learning is useful here because the data is all labeled.

Because I will be attempting to classify between two states of, charging and not charging, logistic regression will be the best technique to use. Logistic regression is defined by:

 for Y =1.

We are trying to determine the best weights (B) of the input variables to fit the data to determine if the EV is charging or not charging. Identifying if there is an EV should be straightforward using inputs such as total consumption and change of consumption. A different logistic regression model can also be used to predict the charging patterns per customer, using inputs such as time from last charge, weekday/ weekend, time of day.

I expect the EV identification to be very accurate, so my success metric is how many charging cycles will I need to determine if an EV is charging at a customer’s location. My guess that only 1 or 2 cycle’s will be needed for the algorithm to determine if an EV is located at the customer residence with greater than 99% accuracy. My metric for success for forecasting is having a minimal window where I expect the next charge to be, and how far in advance I can guess that window. I expect the accuracy will increase as the time from the last charge increases.

There is no need to do feature extraction since the data only has a few inputs, I will generate new ones. The inputs I will be creating will be based on my domain knowledge, such as time from last charge and change in change of consumption. If the logistic regression predictor does not work well, I could try a naïve bayes classifier to create a generative model, to determine the pattern based on the biases of the customer preferences, rather than the differences between EV charging states.