

CS 229 Project: Home Electricity Forecasting

Atinuke Ademola-Idowu, Pawel Kupsc, Sonya Mollinger

December 11, 2014

Introduction

Electricity load forecasting is an important aspect of power systems planning and operation. At the utility scale, load forecasting is important for pricing and determination of the size of the spinning reserve¹. For this problem, forecasting is done on a large scale by aggregating the power consumed by many homes in a single neighborhood.

On a single home scale for grids with real time pricing, forecasting permits using energy storage systems to decrease cost of energy for the consumer. Knowledge of future power consumption along with future electricity prices can make it possible to decide when to engage a battery storage system as opposed to drawing power from the grid. The aim of the project is to carry out a short term forecast on electricity consumption of a single home.

Background

Load forecasting can be divided into three categories: short-, medium-, and long term. Short-term forecasting corresponds to prediction of one hour to one week ahead. Medium-term refers to one week ahead to one year ahead while long-term forecasts are for more than a year². Here, we focus on developing regression models for short-term forecasting.

There is not a lot of previous work on single-home forecasting. The Smart* project at University of Massachusetts, Amherst³ has studied several homes over a two-three month period. They outfitted each home with sensor data and were able to model the power for five intervals throughout the day. The error of their predictions for some intervals was under 5 %.

Additionally a group from Technische Universität München⁴ collected their own data set in addition to analyzing the REDD⁵ data set. Their errors ranged from 5 to 150 %. The group utilized different sampling strategies including dividing the data by day and by time-series.

Data/Features

The data set was obtained from the UCI data set repository⁶. It represents power consumption per minute measured over 4 years for a single home near Paris, France (approximately 2 million data points). We processed the data into hourly averages for prediction (to give a total of 33090 points).

To better predict the power consumption, we used the Wunderground API⁷ interfaced with Python to find the temperature, humidity, and precipitation near the home's location for each hour over the four years. We combined these with the power data for the raw feature matrix. These were surprisingly few data points missing from either source over the entire data set, so we simply removed these points. Since the data set was very large, we chose to process it in Python using the PANDAS package.

The raw input data features are the power (from the home data set) and the temperature, humidity, and precipitation (from the weather data set). However, we also created a number of features while processing the data. We calculated the day of the week, the past hour's power and past day's power. We tested adding some classification-like variables as well, in particular the season and a Boolean variable indicating whether or not the day was on a weekend. We also tested adding further feature columns such as multiple past hour's power values and the past temperatures.

We approached the forecasting problem in several different ways. In one of the approaches, we separated the data set into 24 different sub-divisions corresponding to the hour of the day. We ran regression models on each of these data sets (subdivided into training and test sets). We hoped that this would help us account for systematic hourly variations in the inhabitants' behavior. In the alternative approach we averaged the data set further over each day, and then did a simple regression to predict the average power used for the test days. Finally, we subdivided the day into intervals (e.g. morning, afternoon, evening, middle of the night) in

an attempt to both account for systematic patterns within a day but to not be oversensitive to the inhabitants' precise hourly choices. We also briefly tested dividing the data into winter and summer sets in order to probe the effect of season more effectively.

Methods

The data was split randomly using hold-out cross validation with 75 % training data and a 25 % test. We then applied three different learning models to the data: linear regression (LR), locally weighted linear regression (LWLR) and support vector regression (SVR).

Linear regression was applied both using the normal equations according to Equation 1 and using the software package Sci-Kit Learn⁸. The obtained parameters were used to make a prediction on the training set and the test set. The relationship between some of the features and the power is not linear so it was expected that this model would perform poorly. The main nonlinearity is in the power variation with time of day. This issue was mitigated by having a different model for each hour of the day.

$$\theta = (X^T X)^{-1} X^T y$$

Equation 1. The normal equations

A LWLR algorithm was implemented according to Equation 2 to address the issue of nonlinearity in the power variation with time of the day. The results obtained using this algorithm were a bit better than that of the linear regression.

$$\theta = (X^T W X)^{-1} X^T W y,$$

$$\text{where } w^i = \exp\left(-\frac{(x^{(i)} - x)^2}{2\tau^2}\right)$$

Equation 2. The weighted normal equations

The bandwidth parameter was chosen so as to get a good fit for our data by avoiding overfitting or underfitting the data. When $\tau = 0.1$ was used, we obtained a very low training error (about 0.6%) but a large test error and when $\tau = 10$ was used, we obtained a high training and test error. When $\tau = 5$ was used, we got a reasonable value hence this was the bandwidth parameter value used.

We also used Support Vector Regression to address any nonlinearities in our data. This problem tries to solve the following optimization problem:

$$\min_{\omega, b, \xi, \xi^*} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \xi_i + C \sum_{i=1}^l \xi_i^*$$

$$\text{s. t. } \omega^T \phi(x_{\square}) + b - z_i \leq \epsilon + \xi_i,$$

$$z_i + \omega^T \phi(x_i) - b \leq \epsilon + \xi_i^*,$$

$$\xi_i, \xi_i^* \geq 0, i = 1, \dots, l.$$

Equation 3. The primal problem of ϵ -SVR

We applied SVR to our problem using the algorithm implemented in Sci-Kit Learn. We tested multiple kernels: linear, Gaussian (radial basis function) and polynomial. We saw a dramatic increase in computation time with the polynomial kernel and significant overfitting with the Gaussian kernel. Removing features from the data matrix removed the overfitting problem with the Gaussian kernel, but brought performance down to approximately match the linear kernel. In order to get results from the polynomial kernel within a reasonable time frame, we reduced the data matrix to only three features: the temperature and the past two hours' powers. We also limited the polynomial to be a 2nd-degree polynomial. Additionally, after taking these steps the polynomial-kernel SVR model was much more sensitive to the random splitting of training and test data than some of the other models.

Results

The results of the different regression models are shown in Figure 1 (testing sets) and Figure 2 (training sets). We have primarily focused on the approach that divided the data into hourly data sets, since this allowed us to extract more information from the data set. For hours in which the inhabitants of the house could be assumed to be inactive, we could obtain test errors approaching 5%, but the hours in the daytime (especially 7-8am, around noon, and from 5-10pm - potentially the hours near meal times) were especially difficult. The later hours have high error possibly due to the unpredictability and drastic effect on power usage of the inhabitants dining going to bed at different times. We hypothesize that the mixed results of our project are due to the nature of our data set as single-home power data, which is subject to large amounts of irregular events due to the schedule of the

inhabitants. This is in contrast with the typical power consumption prediction problem, which aggregates data from many houses at a utility level.

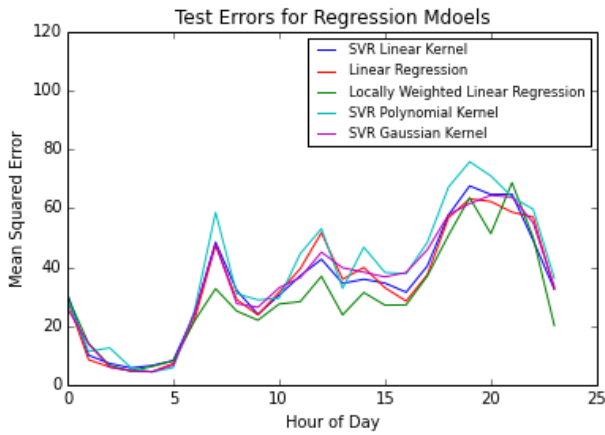


Figure 1. Test Errors for each hour of the day for five different models. Clear peaks are seen in hours of typically greatest human activity.



Figure 2. Training errors for the different regression models used.

By further aggregating the data into daily averages, we were able to predict daily power consumption with errors down to 13%. The test and training errors for this approach are shown in Table 1. Finally, by dividing the day into 4 time intervals corresponding from the hours of 1am-5am, 6am-9am, 10am-4pm, and 5pm-midnight, we were able to obtain errors that remained flat at around 30% (both test and training) for the latter three intervals, and from 7-10% for the first (early morning) interval. This represented an improvement over the average errors from some of the hourly predictions, especially the

evening/night ones; however, this approach also lost us the ability to predict the finer peaks in the data. In contrast, the hourly predictions allowed us to capture much of the peak variation in the data (see Figure 3). However, we believe this is primarily due to the use of the previous hour's power as a feature in the data set.

Model	Test Error	Training Error
Linear Regression	16.1	15.0
LWLR	15.0	10.5
SVR (linear)	13.0	13.0
SVR (Gaussian)	16.7	2.5

Table 1. Errors for Daily Power Consumption Prediction.

In general, we concluded that the features we used were able to explain some of the variability of power consumption for a single home, but cannot explain many of the hour-to-hour individual variations during times of peak power usage. In order to more reliably make predictions for a single home, it would be necessary to include features that describe the behavior of the inhabitants such as if the inhabitants are at home and perhaps what room they are in. For example, knowing that an inhabitant is showering might allow predicting the high power peak when they eat breakfast shortly thereafter. We discuss further improvements at the end of the report.

Interpretation

To better understand the data, we examined how the features were related to the data in multiple ways. The covariances between the features and the power were calculated in order to give an idea which features are more relevant than others. The covariance between the power and the temperature was calculated to be -0.55 while the covariance between the power from one hour before and the current power was 0.57. Features like humidity and day of the week had low covariances: -0.087 and 0.086 respectively. It is somewhat surprising that the day of the week feature has such a low covariance

with the power. Especially during the weekends one might expect greater power consumption than during the week.

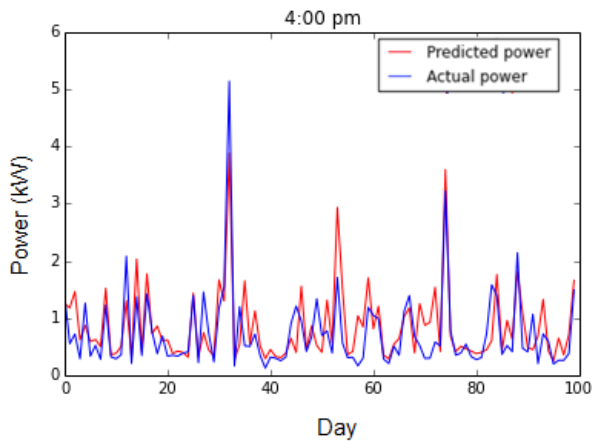


Figure 3. A typical comparison of output from a regression model and the actual power measured. This shows the two for a test set at 4:00pm with predictions done by the SVR-linear model.

A plot of the current power against the power from the previous hour is shown in Figure 4. The relationship seems to follow a linear trend, albeit with a large amount of noise.

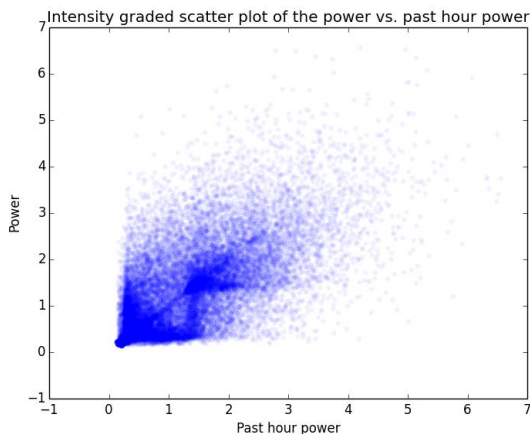


Figure 4. Scatter plot of the relationship of the power consumed and the power consumed in the previous hour of the same day.

The relationship between the various features and the power data can also be probed using the built-in feature selection module in Sci-Kit learn. This allowed us to select the k best features by using a univariate linear regression test to look at a single feature’s relationship with the power data, and then

take a certain number of features that scored the best. This showed us that the most significant features for all hours were the past hour’s power and 2nd-closest hour’s power. When we included three previous hours, all three of them were at the top of the list. The third- or fourth-best feature (depending on how many previous hours were included in the feature set) was the temperature, and the order of the others depended on the hour of the day, alternating between humidity and the “is it a weekend” feature. We found that the Gaussian kernel required us to remove the less important features from the data matrix in order to get the best results, due to its propensity to overfit the data.

Another issue with feature selection was whether to use the “classification-like” variables as features or as dividers for our data set (to use with different regression models). If we had more time, we could have probed this more thoroughly. Since the “season” and “weekend” variables were not particularly significant as features, we briefly tested dividing the data sets along those markers (resulting in $24 \times 2 \times 2 = 96$ sets, using just cold and warm seasons). We saw an improvement in errors for some of these sets, but a slight worsening for other sets, indicating that perhaps further investigation is required.

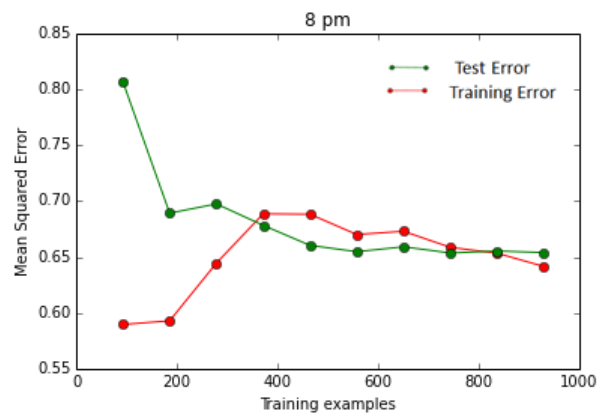


Figure 5. A typical learning curve for one of the hourly data sets. This figure shows the test error and training error as a function of the number of training examples for an SVR model with linear kernel at 8:00 pm.

We used the learning curve calculation package in Sci-Kit Learn to examine the dependence of MSE on the number of training samples and test samples. The precise number of the training and test errors tends to vary slightly with the run due to the

randomness of the cross-validation data splitting, so one advantage of this was the built-in averaging over many runs. The results are shown in Figure 5. It is clear that for the hourly predictions the errors converge quickly at around 400-500 training samples. The learning curve of daily data (not shown) has a slower convergence (800-900 data points) but we do have enough samples to allow this one to converge as well.

Finally, we also note that the errors are not significantly different for the three different regression methods, and the locally weighted linear regression τ -value that gave the best fit approaches linear regression. Changing models in this case improved our errors only slightly. Importantly, this suggests that our data may be fundamentally somewhat unstructured. This is likely due to the inherent variability in the actions of a single family.

Conclusions & Future Work

We have demonstrated that we can predict some aspects of power variability of single-home power consumption using basic regression models. We tested linear regression, locally weighted regression, and ϵ -SVR with three different kernels. We also tested various methods of temporally dividing the data to account for different classes, including running separate regression models on each hour of the day, weekdays vs. weekends, and different seasons. We obtained good results in early morning hours, and larger errors at times when the inhabitants may be more active. Furthermore, we found that all the models had similar results, indicating that the errors may be due to a lack of clean structure in our data despite our efforts to separate temporally. We attribute this to the inherent unpredictability of a single family's actions and consequent power usage.

In the future, to improve our single-home model we ideally would have an increased number of relevant features. For example, we initially planned to analyze the SMART data set from the University of Massachusetts at Amherst⁵, which is equipped with disaggregated power consumption including every appliance in the home and room-specific sensor data. Unfortunately, this data set was only two months and thus too short for prediction purposes. We hypothesize that a longer data set with these detailed features would allow us to account for the individual variability of the inhabitants more precisely. We would also be able to test the

regression algorithms more thoroughly on different feature subsets to understand which are most significant for power prediction. Also, for both models we are interested in trying to discretize the feature space. We currently have some features that are continuous and some that are Boolean or take on discrete values.

In addition, a related problem is one that utility companies are interested in - the ability to predict power for a sector or neighborhood. This relates to our problem in that it would resolve some of the issues of studying a single home by averaging out the variability due to individual decisions.

References

1. Al-Alawi, S. and Islam, S. "Principles of electricity demand forecasting." *Power Engineering Journal* p. 139 (1996)
2. Hobbs, Benjamin F. "Optimization methods for electric utility resource planning." *European Journal of Operational Research* 83.1 (1995): 1-20.
3. Mishra, A., Irwin, D., Shenoy, P., Kurose, J., & Zhu, T. (2012, May). SmartCharge: cutting the electricity bill in smart homes with energy storage. In *Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet* (p. 29). ACM.
4. Veit, A. et al. "Household Electricity Demand Forecasting – Benchmarking State-of-the-Art Methods." Eprint arXiv:1040.0200 (2014).
5. J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In *proceedings of the SustKDD workshop on Data Mining Applications in Sustainability*, 2011.
6. Bache, K. & Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
7. Weather Underground API: see required logo.
8. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011