Predicting Energy Cons

Stanford University

CS229: Machine Learning, Fall 2016

Introduction and Motivation

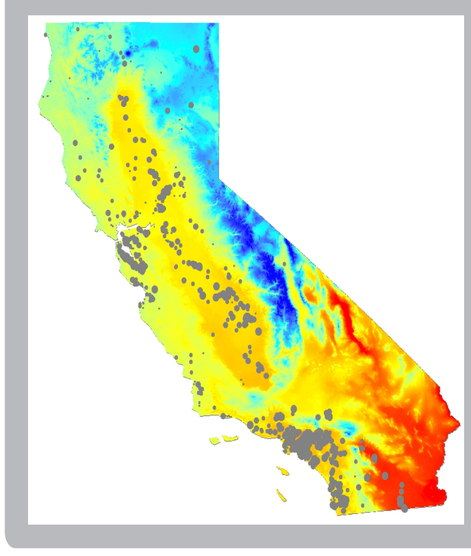
To reduce energy consumption of buildings, understanding patterns in this energy use is paramount. School campuses are especially interesting to solve these energy issues as they act as microgrids. To fit a new building with renewable systems, it is important to predict the future energy use based on design parameters to match generation with the load, so this model uses square footage, energy consumption, natural gas usage for electricity production, peak demand, demand charges etc. to predict the energy consumption by regression.

For existing buildings, the goal is to retrofit with the optimal mix of technologies to gain the best value for money. Here, the model uses inputs such as budget, average temperature, energy consumption, savings obtained, investment for the project etc. to classify the retrofit that provides the best bang for buck. A bill by the CEC provided open source consisting of energy use for over 14,000 schools in California to fund clean energy projects.

Based on these regression and classfication model results, the analysis helps in determining whether a new school building construction or fitting the existing building with retrofits would provide the maximum savings.

Dataset

The data set is composed of several features per school organized by row, including building area, annual energy use etc. organized by column. Additionally, for each school the data provides retrofits that the school implemented, often more than one per school with the resulting energy savings to investment ratio (SIR) per retrofit, the metric by which to gauge the success of a retrofit. To enrich the data set further, a GIS analysis was performed to obtain climate and demographic information for each school to model the climatic impact on energy use.



Feature	Unit
Site Conditioned Area	ft^2
Average Peak Demand	kW
Electricity Density	kWh/ ft^2
Natural Gas Density	therms/ft^2
Energy Use Intensity (EUI)	kBTU/ft^2/yr
Energy Efficiency Measure Category (EEM)	-
EEM Estimated Annual Electric Savings	kWh
Total Combined Energy Costs	$/yr
EEM Estimated Measure Cost	$
EEM Savings to Investment Ratio	-
Total Degree Days	-
Total Housing Units	-
Average Age of Housing Structure	-
Median Gross Rent	$
Median Value of Housing Units	$
Median Household Income from last year	$
Median Age of Population	-

Analysis: Objective 1

To train the model, 80% of the data was chosen at random, leaving 20% for testing. Cross validation was done for selecting the optimum parameters for the models.

1. Polynomial Regression

Firstly, a polynomial regression model was run for predicting the annual energy consumption based on the features provided in the dataset. The cross validation results (in Fig.1) showed that the quadratic model would be efficient in predicting as it has the lowest RMSE. The results for the regression model are plotted in Fig.2 below.

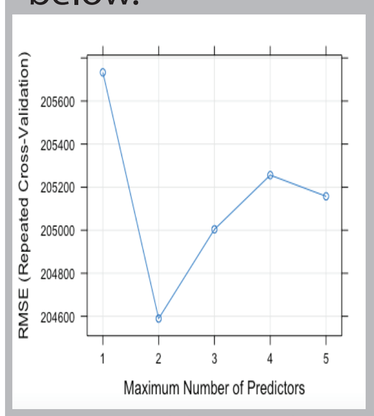


Fig.1: Cross validation results

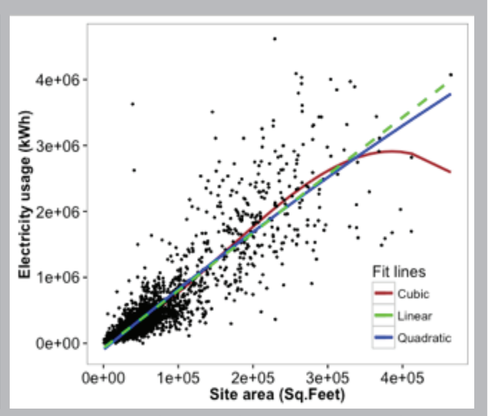
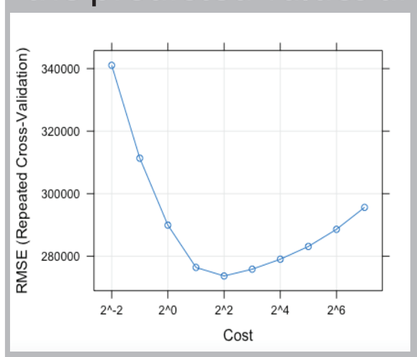


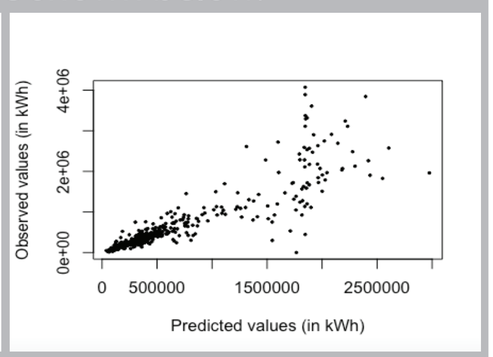
Fig.2: Polynomial regression results

Analysis: Objective 1

2. Support Vector Machine (SVM) Regression

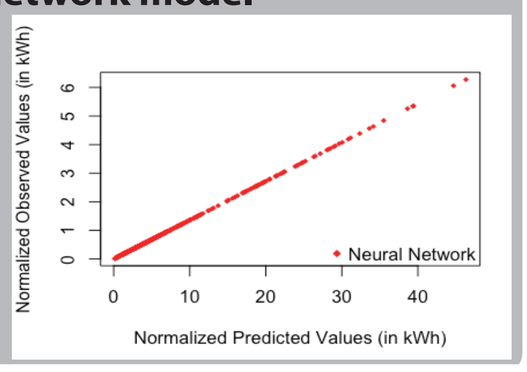
Based on SVM regression, we found the best parameters for the model based on 10 fold cross validation. The results for cross validation and the predicted values are shown below.





3. Artificial Neural Network model

ANN gives the best results for the regression model. The model had a RMSE of 1.4%. The results are shown.



umption of School Buildings

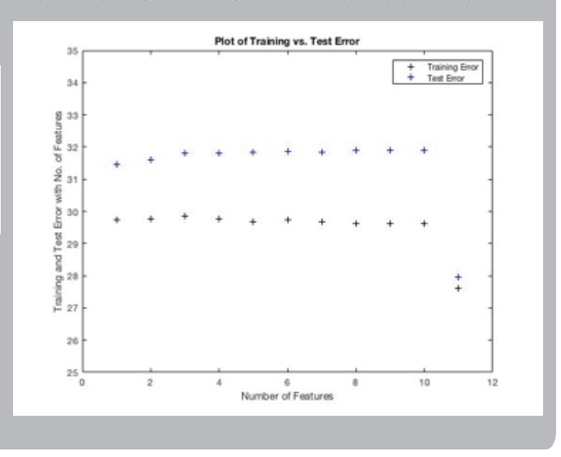
Vikhyat Chaudhry, Dan Sambor, Rohith Desikan

Analysis: Objective 2

Softmax Regression

Given that there are numerous classes of data, classification using a softmax regression model based on a multinomial distribution is logical method for analysis. There are nine classes of retrofit options in the dataset. The Softmax model was run with the features outlined in the dataset table. Ultimately, the feature of cost of retrofit proved to be most important as expected. Principal Component Analysis (PCA) was also performed which showed that retrofit cost along with household income of the neighborhood was also important. The results of the softmax regression model are shown in the plot below which outlines the training and test error of the classification.

$$\begin{aligned}\ell(\theta) &= \sum_{i=1}^{m} \log p(y^{(i)}|x^{(i)};\theta) \\ &= \sum_{i=1}^{m} \log \prod_{l=1}^{k} \left(\frac{e^{\theta_l^T x^{(i)}}}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \right)^{1\{y^{(i)}=l\}}\end{aligned}$$



Discussions and Conclusions

The results of the analysis for both the objectives are listed below:

- Objective 1:

Model	RMSE
Neural Network	1.40%
Polynomial Regression (Quadratic model)	39.0%
SVM	55.8%

-Objective 2:

Model	Test Error
Neural Network	3.36%
Softmax	27.9%

While lighting dominates the data set, the model recommends an even larger percentage of lighting retrofits. This makes sense as LED lighting technology has become so efficient that it makes significant economic sense especially for those regions without significant HVAC loads.

As per the analysis, the neural network was much more efficient in prediction and classification for both the objectives.

So based on the given design features in the dataset, the analysis will allow a school official to predict the energy consumption if a new building is to be built or the best retrofits that lead to largest savings per investment if they were to remodel the existing campus.

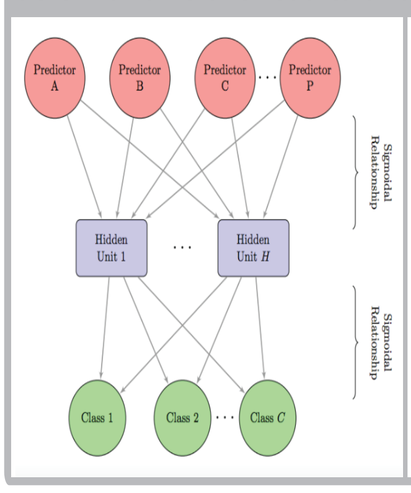
Analysis: Objective 2

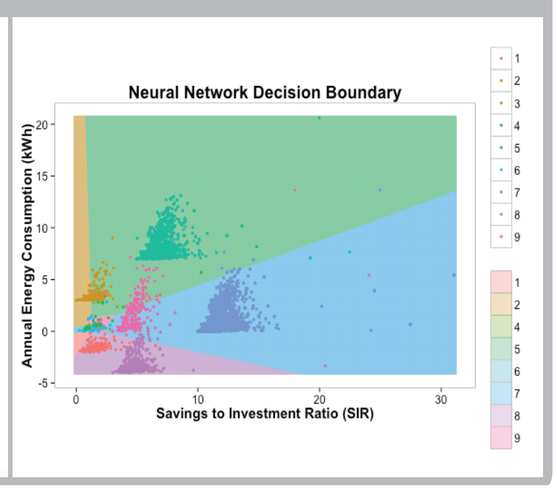
Artificial Neural Network (ANN)

ANN was used for classification of retrofits based on SIR and annual energy use values. This model is a 3-layer ANN with the hidden units represented by neurons and their corresponding functions and interconnection weights. ANN is defined by three types of parameters:

- Interconnection pattern between the different layers of neurons.
- Learning process for updating the weights of interconnections.
- Activation function that converts a neuron's input to its output activation.

The flowchart shows the working of ANN algorithm. The plot below shows the results of the classification of the nine retrofits.





Future Work and References

These models can also be extrapolated beyond California to predict for other areas of the country. Given California's relatively mild climate, there may be a more significant link between weather and energy use for the rest of the US.

Other algorithms can also be tested including falling rule lists and decision trees, which have been demonstrated in the literature for step-wise installation of retrofits (Marasco et al, 2016).

References

[1] E. Mocanu, P. Nguyen, M. Gibescu, and W. Kling, "Comparison of Machine Learning Methods for Estimating Energy Consumption in Buildings," Probabilistic Methods Applied to Power Systems, Conference, 2014.

[2] A. Foucquier, S. Robert, F. Suard, L. Stphan, and A. Jay, "State of the art in building modelling and energy performances prediction: A review," Renewable and Sustainable Energy Reviews, vol. 23, no. 0, pp. 272 – 288, 2013

[3] D. E. Marasco and C. E. Kontokosta, "Applications of machine learning methods to identifying and predicting building retrofit opportunities," vol. 128, pp. 431–441, 2016.