Energy efficiency

This study looks into using linear regression to predict energy efficiency in the form of heating load.

**1 Introduction**

The heating load of the building is defined as the amount of heat energy that must be added to the space in order to maintain the temperature in an acceptable range. This is of particular interest as not only by designing more efficient buildings can heating costs be reduced but additionally, the carbon footprint from heating such buildings is reduced.

The dataset was taken from the UCI machine learning repository, accessible at <https://archive.ics.uci.edu/ml/datasets/Energy+efficiency> and this report builds on work done by Tsanas and Xifara[1].

**2 Data Exploration**

The dataset has eight features denoted by X1...X8 and two target variables denoted by Y1 and Y2 and shown in Table 1.

|  |  |
| --- | --- |
| Code | Variable Name |
| X1 | Relative Compactness |
| X2 | Surface Area |
| X3 | Wall Area |
| X4 | Roof Area |
| X5 | Overall Height |
| X6 | Orientation |
| X7 | Glazing Area |
| X8 | Glazing Area Distribution |
| Y1 | Heating Load |
| Y2 | Cooling Load |
| Table 1: Variables | |

Main objective of the analysis that specifies whether your model will be focused on prediction or interpretation.

Brief description of the data set you chose and a summary of its attributes.

Brief summary of data exploration and actions taken for data cleaning and feature engineering.

Summary of training at least three linear regression models which should be variations that cover using a simple linear regression as a baseline, adding polynomial effects, and using a regularization regression. Preferably, all use the same training and test splits, or the same cross-validation method.

A paragraph explaining which of your regressions you recommend as a final model that best fits your needs in terms of accuracy and explainability.

Summary Key Findings and Insights, which walks your reader through the main drivers of your model and insights from your data derived from your linear regression model.

Suggestions for next steps in analyzing this data, which may include suggesting revisiting this model adding specific data features to achieve a better explanation or a better prediction.

**References**

[1] A. Tsanas, A. Xifara: 'Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools', Energy and Buildings, Vol. 49, pp. 560-567, 2012 (the paper can be accessed from [[Web Link]](http://people.maths.ox.ac.uk/tsanas/publications.html))   
  
For further details on the data analysis methodology:   
A. Tsanas, 'Accurate telemonitoring of Parkinsonâ€™s disease symptom severity using nonlinear speech signal processing and statistical machine learning', D.Phil. thesis, University of Oxford, 2012 (which can be accessed from [[Web Link]](http://people.maths.ox.ac.uk/tsanas/publications.html))