**Dataset Description and Summary Statistics**

The dataset for implementing the gradient descent algorithm for linear regression is downloaded from UCI Machine Learning repository. The link for downloading the dataset is [appliances+energy+prediction](https://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction).

Dataset consists of 19735 observations on a total of 29 variables. The variables are Appliances and lights energy consumption in the house, 9 temperatures and 9 relative humidity in 9 different locations in the house, 6 weather variables such as outside temperature, outside humidity, Pressure, Windspeed, Visibility and Dewpoint. We also have date column, and two random variables that are being ignored in this model. Additional information about the dataset can be found out in the above mentioned link. The dependent variable for linear regression model is Appliances energy consumption based on the remaining predictors. There are no missing values in any column, so no missing value imputation is required on this dataset.

**Train/Test Split, Classification, Standardization**

The dataset is split into train and test sets using 70-30 split percentage. The number of observations in train and test sets are 13814 and 5921 respectively.

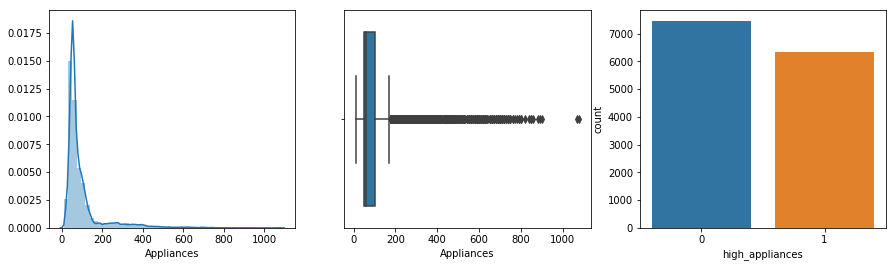
A new target variable, *high appliances* is created using Appliances median energy consumption as cutoff. Houses with appliances energy consumption more than median are denoted as 1 and the remaining as 0.

The independent variables in train set are standardized using mean and standard deviation of respective features from the train set. The features in test set are also standardized using the same train set mean and standard deviation to avoid data leakage from test set into train set which might give overfit model.

**Data Exploration – Target Variable**

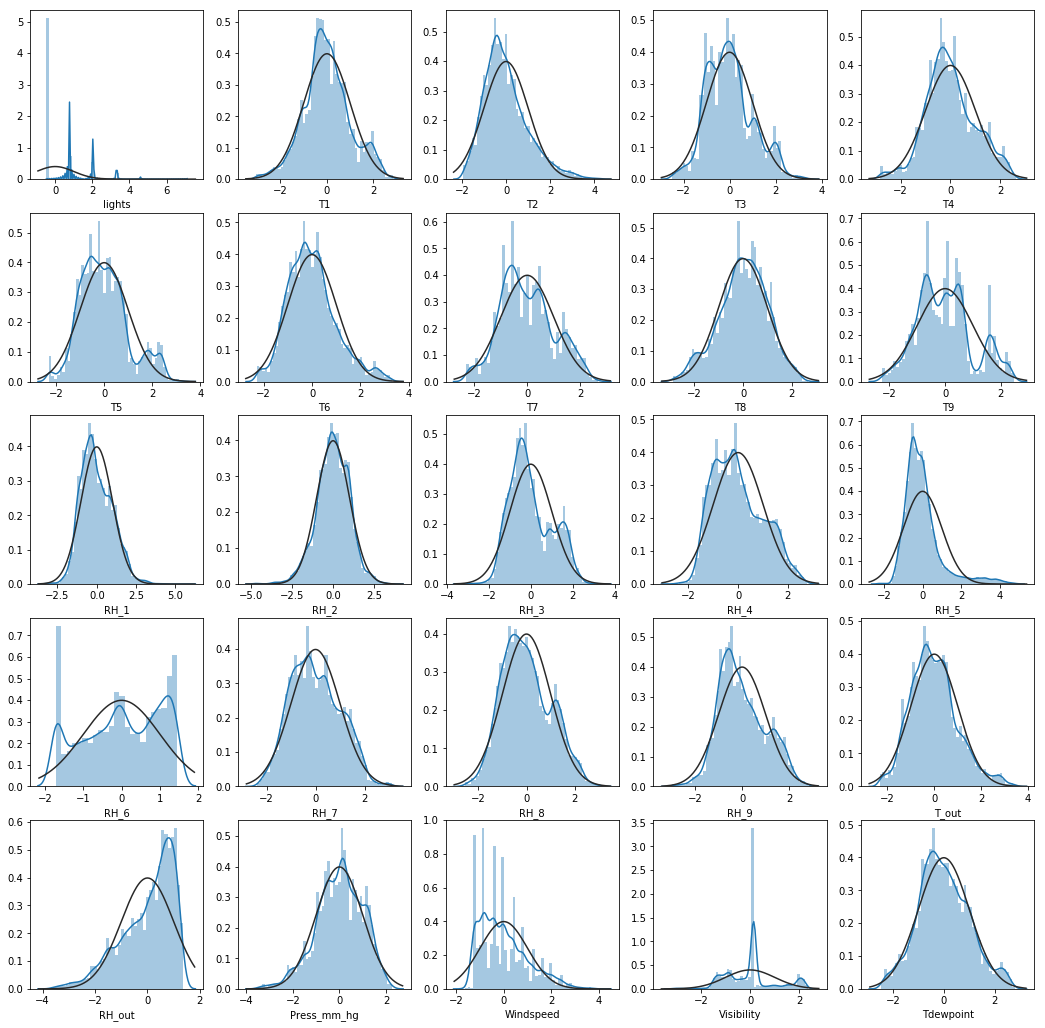
The target variable is positively skewed with few outliers. But they are ignored since there is no strong evidence on the outliers being clerical errors. The data is collected using sensors and faulty data could result only from sensor malfunction. Since, we don’t have information on this, the outliers are taken as legitimate data and are included in the model. Moreover, they contribute only 2.7% of the observations.

For the category variable in train, there are 6351 houses with high appliances energy consumption (greater than median) and 7463 houses with low appliance energy consumption (less or equal to median).

****

**Data Exploration - Predictor Variables**

Below are distribution charts of 25 features. Except for RH\_6, all others roughly fit normal distribution



**Feature Selection**

Features for different experiments are selected using the correlation heatmap presented in next page. 15+ features models - 17 features are shortlisted that have highest correlation with the target variable.

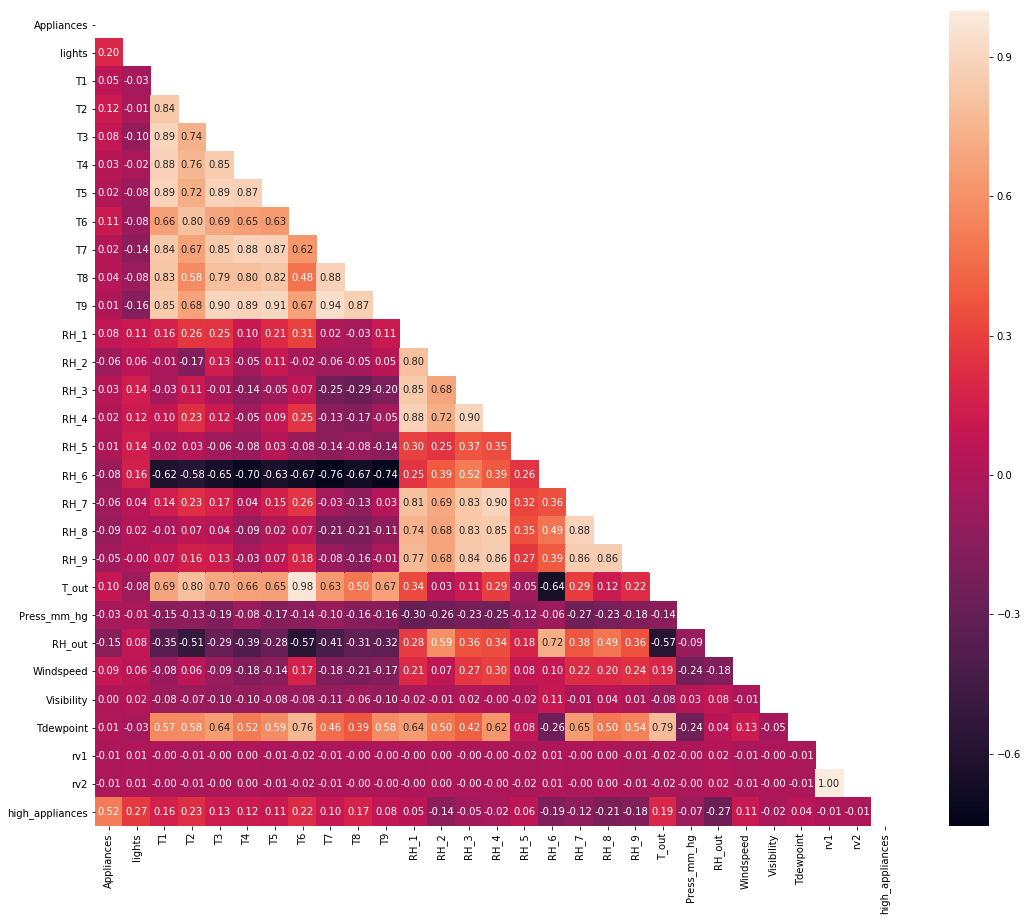
10 random features model - 10 random features are selected form all available 25 predictor variables. 10 best features model - 10 features are shortlisted that have highest correlation with target variable.

RH\_4 has 0.90 correlation with RH\_3,RH\_7 which have better correlation with target than RH\_4.

T9 has more than 0.90 correlation with T3,T5,T7 which have better correlation with target than T9

T\_out has 0.98 correlation with T6 which has better correlation with the target variable than T\_out.

**Correlation Heatmap**



**Models Specification**

**Model 1 (Experiment 1)** - linear (target = *Appliances*) and logistic (target = *high\_appliances*) regression

target = + \**lights* + \**T1* + \**T2* + \**T3* + \**T4* + \**T6* + \**T8* + \**RH1* + \**RH2* + \**RH3* + \**RH6* + \**RH7* + \**RH8* + \**RH9* + \**Press\_mm\_hg* + \**RH\_out* + \**Windspeed*

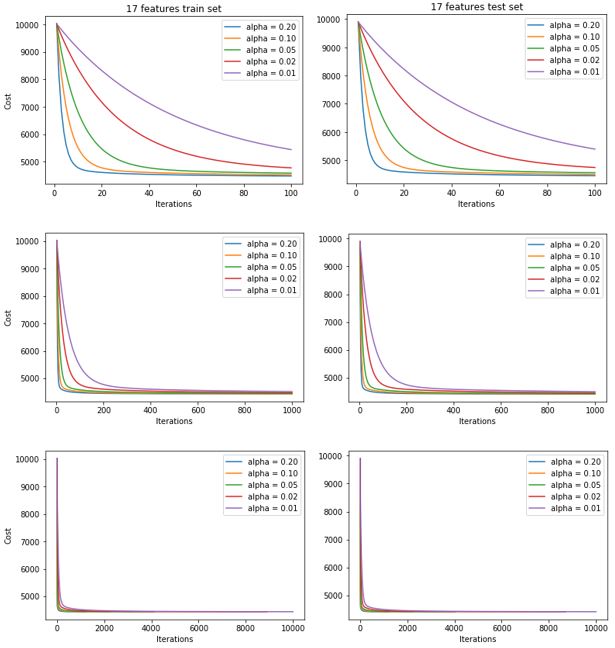
**Model 2 (Experiment 3)**

target = + \**T1* + \**T3* + \**T5* + \**T8* + \**RH3* + \**RH5* + \**Press\_mm\_hg* + \**RH\_out* + \**Visibility* + \**Tdewpoint*

**Model 3 (Experiment 4)**

target = + \**lights* + \**T2* + \**T3* + \**T6* + \**RH1* + \**RH2* + \**RH6* + \**RH8* + \**RH\_out* + \**Windspeed*

**Experiment 1a** - Tuning alpha parameter with gradient descent for linear regression with convergence threshold of 0.0001. Following graphs shows the decrease in cost function against number of iterations.



We can observe that as value of alpha decreases, the algorithm needs more iterations for convergence. Following table shows the actual number of iterations needed for above alphas with threshold of .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Learning Rate | Convergence Threshold | Convergence Time | Iterations | Cost after convergence |
| **0.20** |  | **7 seconds** | **3999** | **4440.6617709813245** |
| 0.10 |  | 14 seconds | 7773 | 4440.6617709815050 |
| 0.05 |  | 26 seconds | 14978 | 4440.6617709820380 |
| 0.02 |  | 67 seconds | 35874 | 4440.6617709836500 |
| 0.01 |  | 133 seconds | 69513 | 4440.6617709863895 |

For alpha = 0.2 and convergence threshold = , below beta parameters came from gradient descent

= 98.0129 = 12.5884 = -9.1124

= 18.5415 = 7.1482 = -23.3118

= 1.0971 = 63.3958 = -6.7171

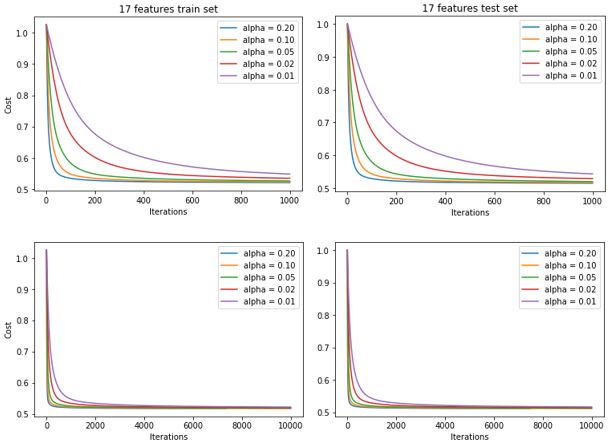
= -40.1456 = -62.2118 = 1.6765

= 43.7432 = 20.4086 = 0.8448

= -19.2265 = 8.6189 = 4.2329

Coefficients from gradient descent are accurate to the closed form linear regression model coefficients. With linear model package, -statisitc is found as 15.6% and T2, Pressure, RH are insignificant features.

**Experiment 1b** - Tuning alpha parameter with gradient descent for logistic regression with convergence threshold of . Following graphs shows the decrease in cost function against number of iterations.



We can observe that as value of alpha decreases, the algorithm needs more iterations for convergence. Following table shows the actual number of iterations needed for above alphas with threshold of .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Learning Rate | Convergence Threshold | Convergence Time | Iterations | Cost after convergence |
| **0.20** |  | **48 seconds** | **29004** | **0.5175267603268712** |
| 0.10 |  | 83 seconds | 55549 | 0.5175267603272693 |
| 0.05 |  | 179 seconds | 109001 | 0.5175267603276069 |
| 0.02 |  | 563 seconds | 258778 | 0.5175267603296202 |
| 0.01 |  | 1325 seconds | 502194 | 0.5175267603322398 |

For alpha = 0.2 and convergence threshold = , below beta parameters came from gradient descent

= -0.15078 = 0.56904 = -0.03968

= 0.68980 = 0.76855 = -0.78382

= 0.47667 = 2.18618 = -0.72758

= -0.70678 = -1.43367 = 0.01965

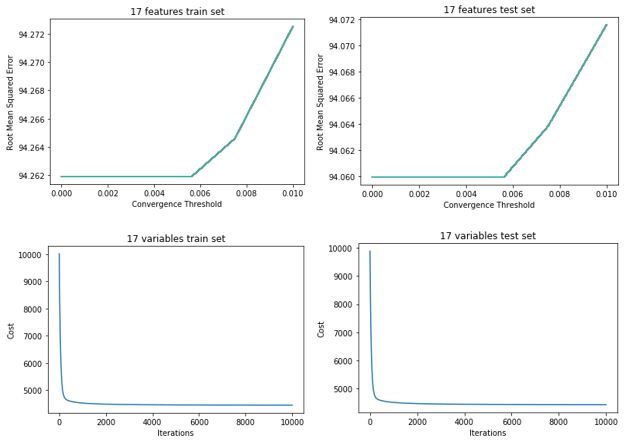
= 0.35108 = 0.08040 = -0.11551

= -0.85176 = 0.54689 = 0.20180

Coefficients from gradient descent are accurate to the closed form logistic regression model coefficients.

**Experiment 2**

Following graphs shows the Root Mean Squared Error as a function of different convergence thresholds



Above graphs show cost function for threshold at 0.005 and its relationship with number of iterations.

**Experiment 3**

Following table shows the errors and RMSE obtained from 17 selected features and 10 random features.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Iterations | Cost | RMSE |
| **17 features train** | **3999** | **4440.66** | **94.24** |
| 10 random train | 781 | 4972.68 | 99.73 |
| **17 features test** | **3845** | **4421.81** | **94.04** |
| 10 random test | 807 | 4934.00 | 99.34 |

Model with 17 features has lower cost and error statistics, however it takes more iterations to converge.

**Experiment 4**

Following table shows the errors and RMSE obtained from 17 selected features and 10 best features.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Iterations | Cost | RMSE |
| **17 features train** | **3999** | **4440.66** | **94.24** |
| 10 best train | 1938 | 4504.05 | 94.91 |
| **17 features test** | **3845** | **4421.81** | **94.04** |
| 10 best test | 1900 | 4470.48 | 94.56 |

Model with 17 features has lower cost and error statistics, however it takes more iterations to converge.