Bayesian Machine Learning: Final Project

1 Introduction

The 'Energy Efficiency' dataset supplied by the University of Oxford consists of 8 input variables $x_1, x_2, \ldots, x_8$ which are predictors for the 'Heating Load'(y) a house can maintain. Predictor variables are the architectural attributes of the house such as the 'Roof Area', 'Relative Compactness and 'Glazing Area' amongst others. Heating Load is a real-value variable. The dataset consists of 768 data points, split evenly into test and training data. There is also a constant bias term (x_0) to assist linear regression modelling.

The aim of this project is to utilise Bayesian methods to accurately predict the Heating Load for a given house and its respective architectural attributes. This information can help inform policies to maximise the energy efficiency of housing in the U.K.

A Bayesian linear regression (BLR) will be used to predict the Heating Load (y) of a house given its attributes (x):

$$y_i = \mathbf{x}_i^\mathsf{T} \boldsymbol{w} + \varepsilon_i, \tag{1}$$

where w is a vector of weight coefficients, and ε_i is the Gaussian error term.

We will assess deterministic approximation in the form of Variational Inference (VI) and stochastic approximation in the form of the Hamiltonian Monte Carlo (HMC) method to approximate the nuisance parameters of the BLR - α and β. Where α is the precision of the weights $1/\sigma_w^2$ and β is the precision of the Gaussian noise $1/\sigma_e^2$.

2 Exploratory Data Analysis

2.1 Predictor Relevance

Initial exploratory analysis was performed to assess the x variables potential of explaining Heating Load. First, the x variables were standardised. The relevance of these variables can be highlighted by the plotting each x variable against Heating Load, as depicted in Figure 1.

It is clear from the visual depiction of each of the x variables against Heating Load that the x variables exhibit no obvious, strong linear relationship. This is a result of the discretisation of many of the x variables. Despite the apparent continuous nature of variables such as 'Roof Area', standardised measurement or data recording limitations means that many variables are represented as discrete variables. As Heating Load is a continuous variable, this will increase the difficulty of the prediction task, as the discrete variables may not capture the underlying complexity or variability of the continuous Heating Load. However, these discrete variables do contain information which may be helpful in predicting Heating Load.

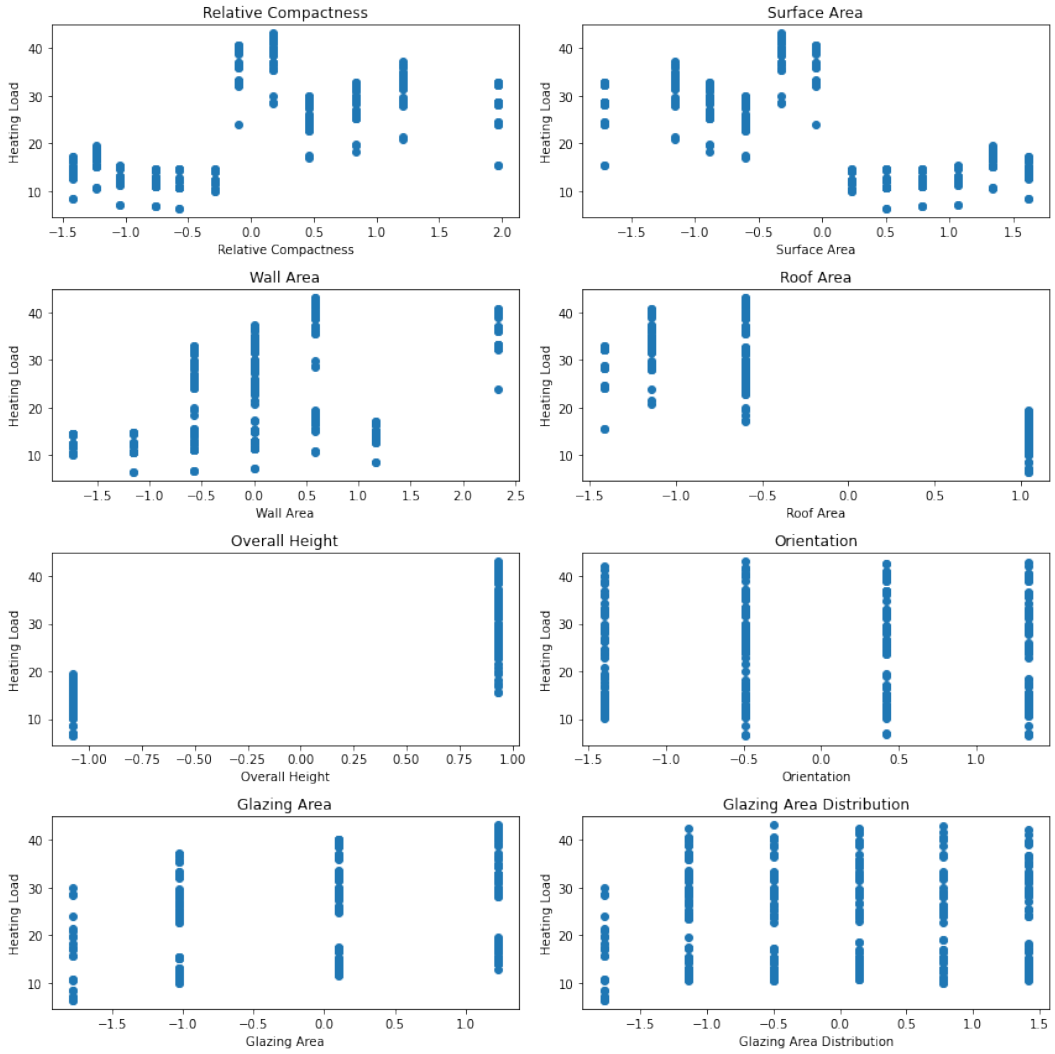


Figure 1: Exploratory Data Analysis: x variables against Heating Load plots

The difficulty of this task is further compounded by the correlation matrix shown in Figure 2. By inspection of the correlation coefficients, 'Overall Height' and 'Roof Area' show signs of strong predictors with absolute correlation coefficients >0.8, 'Relative Compactness' and 'Surface Area' appear relatively strong predictors with absolute values >0.6, but the other variables show little promise in predicting y.

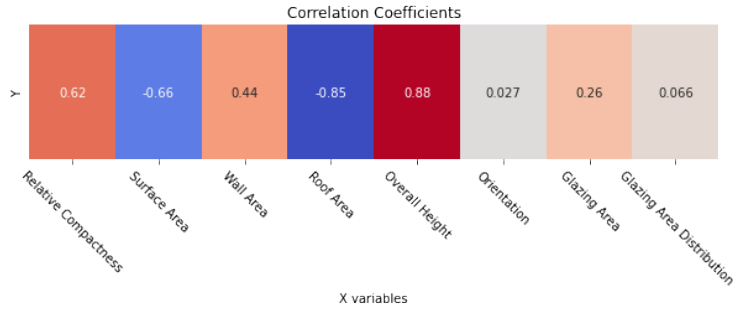
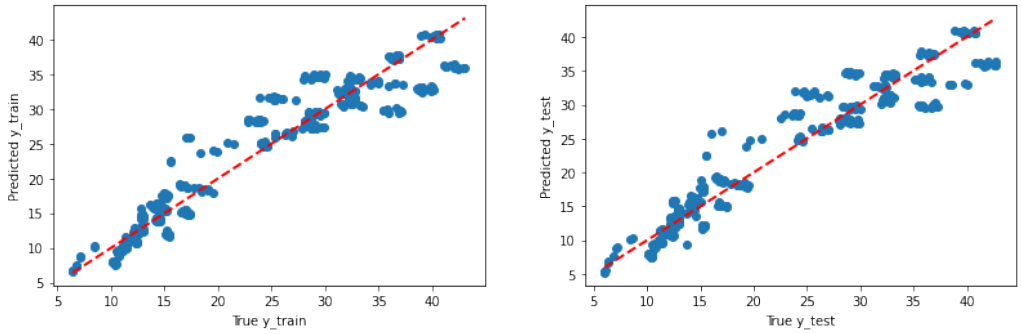


Figure 2: Correlation Coefficients

2.2 Least-Squares Linear Regression

As a baseline for comparison, a least-squares linear regression was fit using the X variables against Heating load. As is evident in Figure 3, the model offers a solid prediction of Heating Load but as shown in the MAEs and RMSEs there is more that can be done.



(a) Training data

(b) Test data

Figure 3: Least-squares Linear Regression

Least-squares Linear Regression MAE and RMSE:

- Train MAE: 2.1307
- Test MAE: 2.0690
- Train RMSE: 3.0116
- Test RMSE: 2.8436

3 Bayesian Linear Regression

3.1 Type-II Maximum Likelihood

Bayesian Type-II maximum likelihood methodology was used to estimate the BLR hyper-parameters α and β, both of which are assumed to have uniform priors. The aim is to estimate Heating Load given X, the corresponding weights w and the hyper-parameters: α and β.

Figure 4 shows the log-posterior of the hyper-parameters α & β using Type-II Bayesian estimation. The most probable $log\alpha$ and $log\beta$ are marked in red.

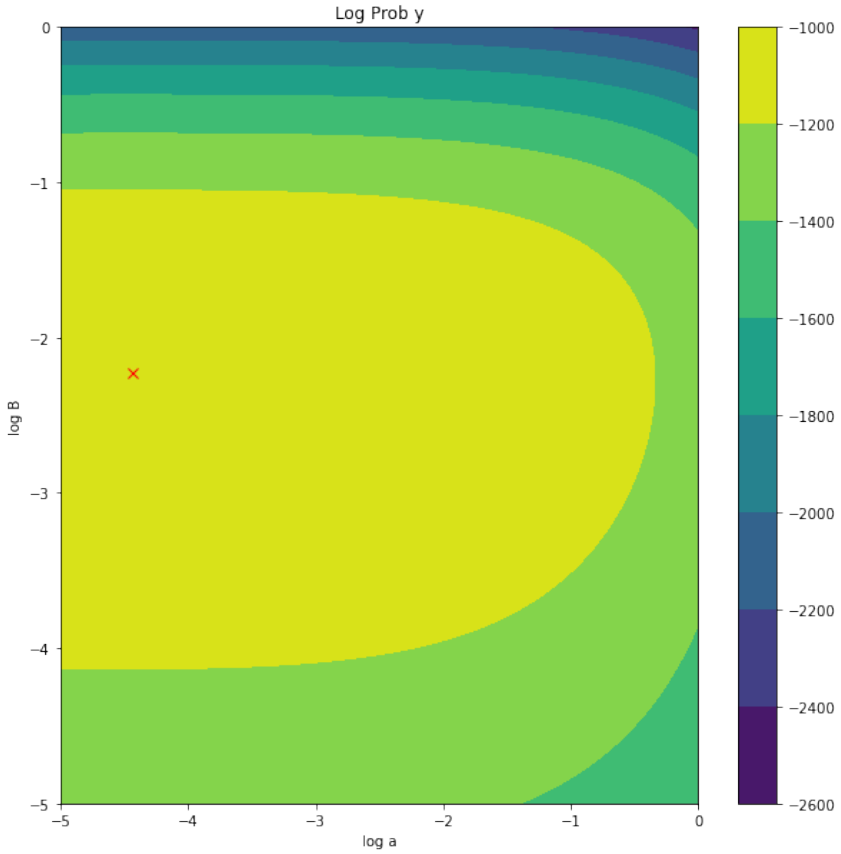


Figure 4: Log-posterior with most probable α and β using Type-II maximum likelihood

- Most probable α: 0.0117
- Most probable β: 0.1084
- Log prob y given α & β: -1,001.4576

The MAE of the test and training sets takes a slight improvement versus least-squares linear regression.

Type II maximum likelihood MAE and RMSEs:

- Train MAE: 2.1302
- Test MAE: 2.0668
- Train RMSE: 3.0117
- Test RMSE: 2.8434

3.2 Variational Inference

Next, we employ a deterministic approximation technique in the form of Variational Inference (VI) with Mean-Field Theory factorisation, to approximate the values of the hyper-parameters α & β of the BLR. Figure 5 shows the log-posterior of the hyper-parameters α & β using Variational Inference (VI). The VI most probable $log\alpha$ and $log\beta$ are marked in black.

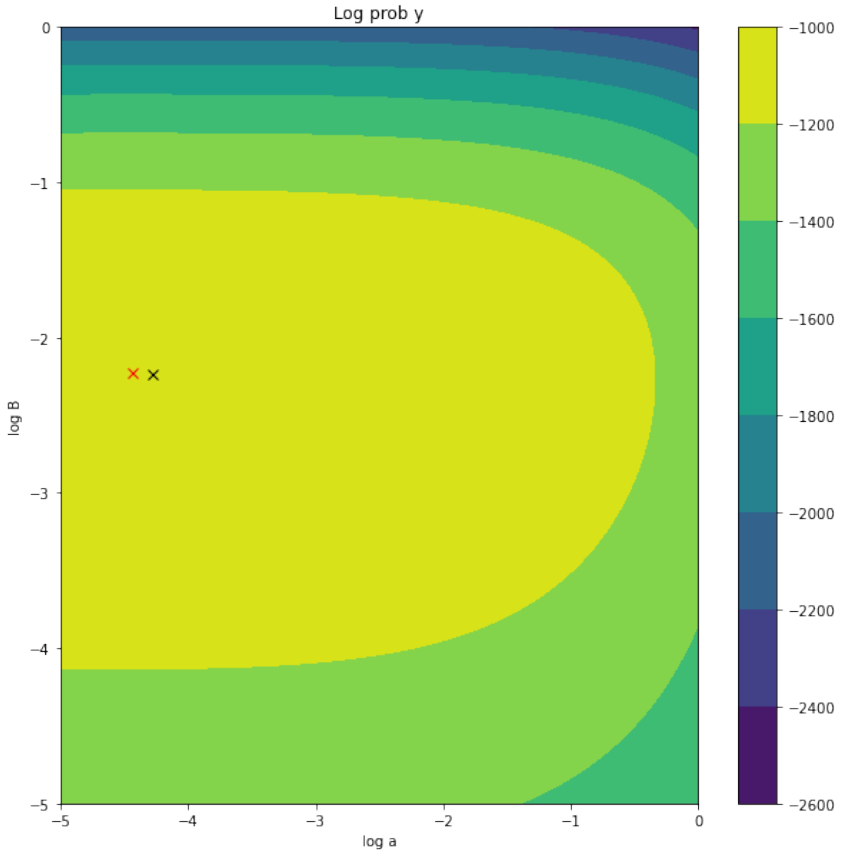


Figure 5: Log-posterior with most probable α and β using VI (black) and Type-II (red)

- Most probable α: 0.0139
- Most probable β: 0.1076
- Log prob y given α & β: -1,001.5037

The MAEs and RMSEs are negligibly different to those produced by Type-II - a sign of the approximation power the algorithm possesses.

Varational Inference MAE and RMSEs:

- Train MAE: 2.1339
- Test MAE: 2.0616
- Train RMSE: 3.0177
- Test RMSE: 2.8485

4 HMC - 2-D Gaussian Illustration

The Hamiltonian Monte Carlo (HMC) algorithm is a stochastic approximation method that uses Hamiltonian dynamics to randomly sample from a pre-defined distribution to approximate parameters

of interest. As a demonstration of the process, we apply HMC to a simple 2-dimensional Gaussian, we generate samples to estimate a target probability distribution.

The mathematical formula of the 2-dimensional Gaussian follows:

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{1/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{2}$$

Where x is a 2-D vector, μ is the mean and Σ is the covariance matrix.

The following functions were used for the HMC:

```
def energy_func(x, covar):
    neglgp = -multivariate_normal.logpdf(x, mean=None, cov=covar)
    return neglgp

def energy_grad(x, covar):
    g = np.empty(2)
    cov_inv = np.linalg.inv(covar)
    delta = x - [0,0]
    g = np.matmul(cov_inv, delta)
    return g
```

The HMC hyper-parameter values used to generate the results:

- R: 10,000
- L: 25
- eps: 0.375

The effectiveness of the HMC is best demonstrated by Figure 6. The blue samples converge to the target distribution shown by the red contours.

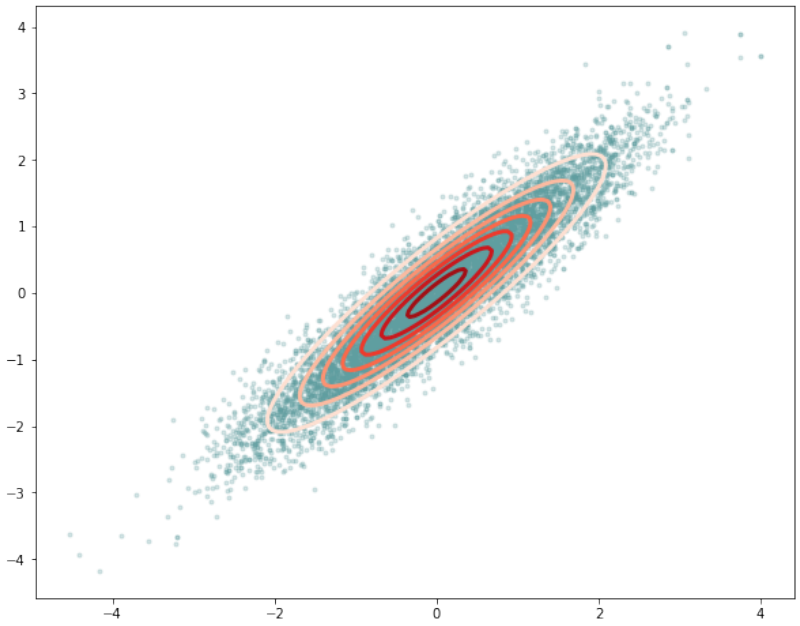


Figure 6: 2-D Gaussian HMC Illustration

5 HMC - Bayesian Linear Regression

The HMC algorithm was utilised to approximate the hyper-parameters of the BLR; α and β as well as the weight w coefficients. Log transformation of the hyper-parameters α and β was performed to avoid negative values of α and β. The trace plots of the parameters in Figure 7 demonstrates the effectiveness of the HMC algorithm. Most of the parameters show convergence, with exception to 'Surface Area', 'Wall Area' and 'Roof Area' which exhbit irregularities (Gelman et al. 2014). Thus, there is less uncertainty regarding their posteriors.

The HMC hyper-parameter values used to generate the results:

- R: 5,000
- L: 100
- eps: 0.00962

Appendix A shows the sample space for α and β has been appropriately explored by the HMC algorithm.

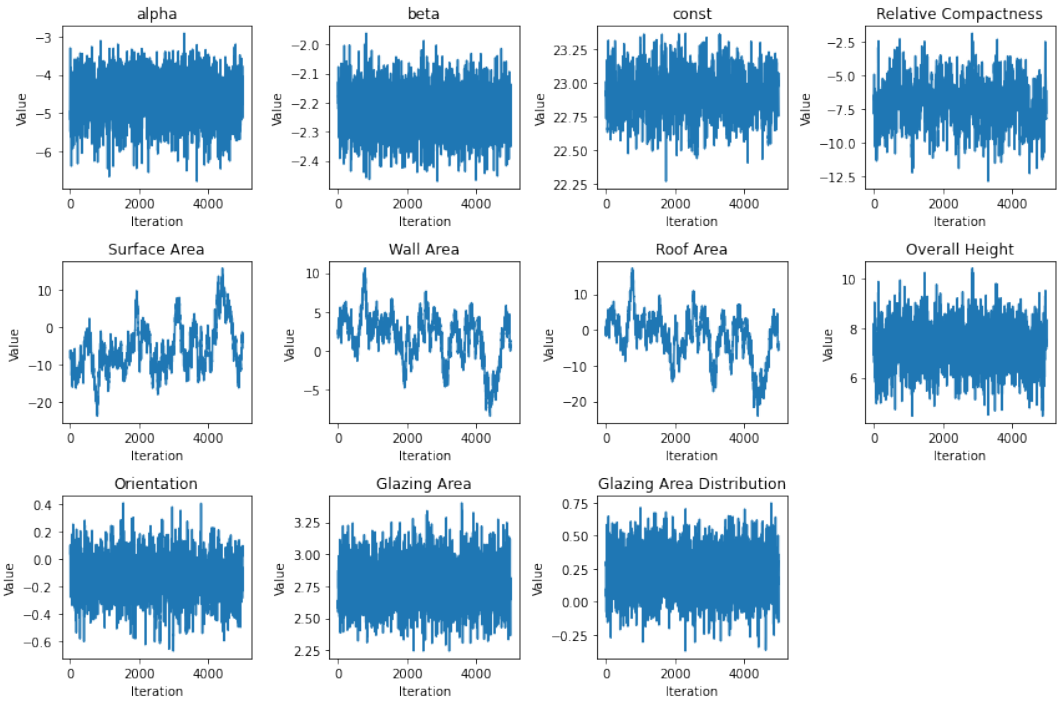


Figure 7: HMC linear regression: trace plots of parameters

The most probable α and β using HMC is depicted in Figure 8 in blue.

The most probable α & β:

- Most probable α: 0.0105
- Most probable β: 0.1076
- Log prob y given α & β: -1,001.4877

Again, the MAEs and RMSEs show negligible difference to those produced by the VI and Type-II maximum likelihood as shown below.

HMC MAE and RMSEs:

- Train MAE: 2.1293

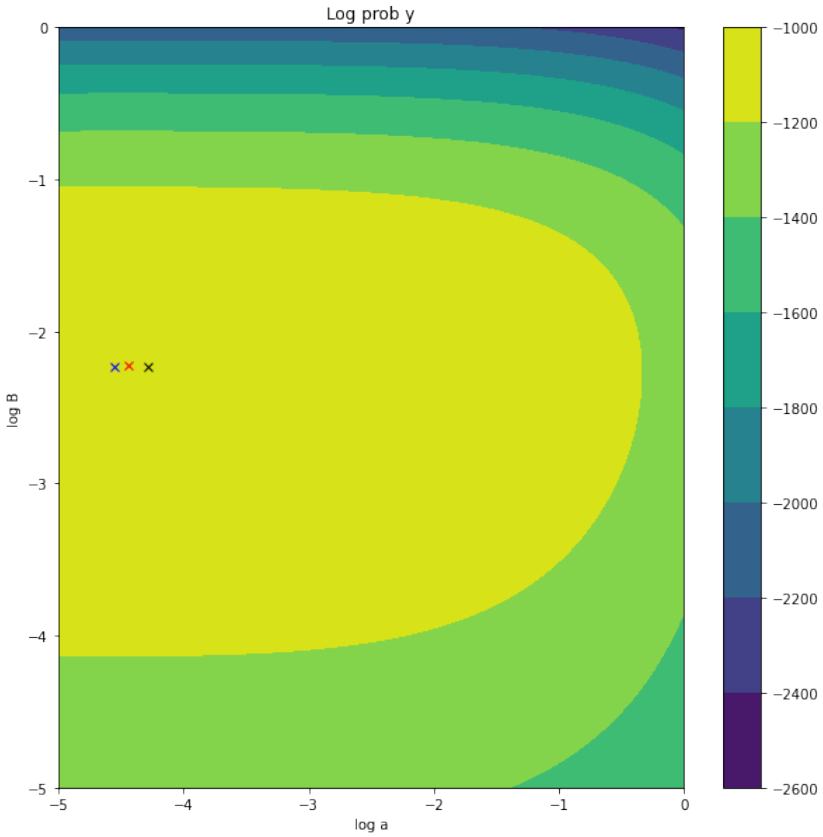


Figure 8: Log-posterior with most probable α and β using HMC (blue), VI (black), Type-II (red)

- Test MAE: 2.0691
- Train RMSE: 3.0116
- Test RMSE: 2.8431

6 HMC - Bayesian Classifier

The target variable Heating Load was transformed into a binary variable with a value of 1 for Heating Load greater than 23.0 (high) and 0 if less than 23.0 (low). A Bernoulli likelihood with a sigmoid activation function was used to predict high and low Heating Load. This essentially transforms the model into a logistic regression model.

The misclassification results came down to 1.30% for the training and 1.04% for the testing data sets using the HMC classifier, which is an impressive result given the data-set was split evenly between high and low samples. The convergence of the parameters is demonstrated in Figure 9. Appen

The HMC hyper-parameter values used to generate the results:

- R: 5,000
- L: 100
- eps: 0.01

Appendix B shows the sample space for α and β has been appropriately explored by the HMC algorithm.

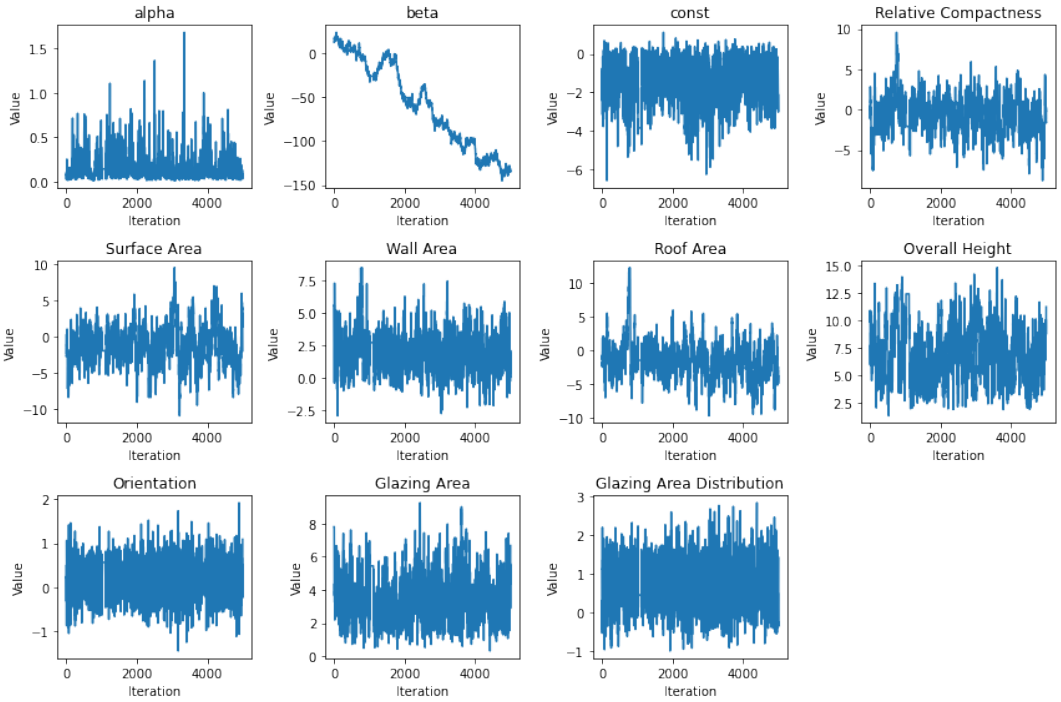


Figure 9: HMC classifier: trace plots of parameters

7 Summary

This report offers an evaluation of four linear regression models for predicting the Housing Load of a house. The models exhibit strong predictive performance, as MAE of around 2.13 and 2.06 on the training and testing data-sets, respectively. Notably, these values represent a fifth of the standard deviation of the Housing Load across the aforementioned data-sets. Despite these favorable results, there is potential for further improvement in prediction accuracy. This may involve the adoption of non-linear prediction techniques, such as Gaussian Processes, which can effectively capture the intricate patterns inherent in non-linear data.

A comparative analysis of the effectiveness of different linear regression methods for predicting the Housing Load is shown in Table 1. All four methods exhibit similar MAE and RMSE results on both the training and test data-sets. HMC produces marginally superior results in RMSE Train, RMSE Test and MAE Train, but is surpassed by VI in MAE Test. HMC has been proved to be an efficient approximation method in high-dimensional spaces such as the 'Energy Efficiency' data-set (Gelman et al. 2014). Nevertheless, the observed differences between the methods are statistically negligible.

	Least-squares	Type-II	VI	HMC
MAE Train	2.1307	2.1302	2.1339	2.1293
MAE Test	2.0690	2.0668	2.0616	2.0691
RMSE Train	3.0116	3.0117	3.0117	3.0116
RMSE Test	2.8436	2.8434	2.8485	2.8431

Table 1: Results Summary: MAE and RMSE

Similar conclusions are made with respect to the approximation of the values of α and β, with values for α around 0.01 and 0.11 for β for the 3 approximation methods. The HMC algorithms trace plots also enlighten us as to which parameters are of particular importance in the prediction of Housing Load. As previously highlighted, those variables which show greater signs of convergence Figure 7 offer more reliable posterior distributions, and thus more reliable predictions.

	Type-II	VI	HMC
α	0.0117	0.0139	0.0105
β	0.1084	0.1076	0.1076

Table 2: Summary: estimated α & β

8 References

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B., 2014. *Bayesian Data Analysis*. CRC press.

9 Appendix

9.1 Appendix A

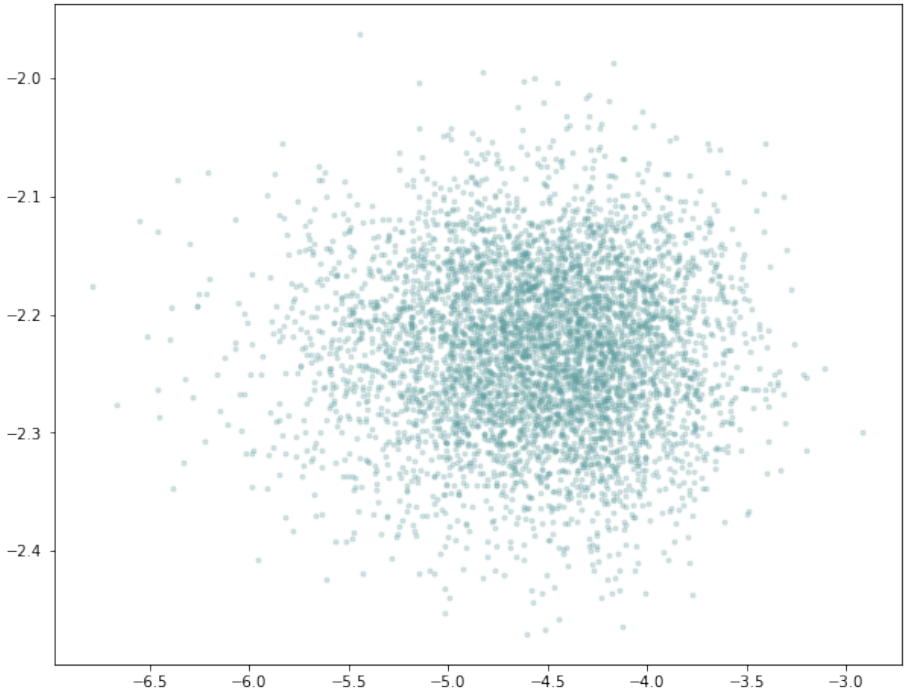


Figure 10: HMC Linear Regression: α and β exploration samples

9.2 Appendix B

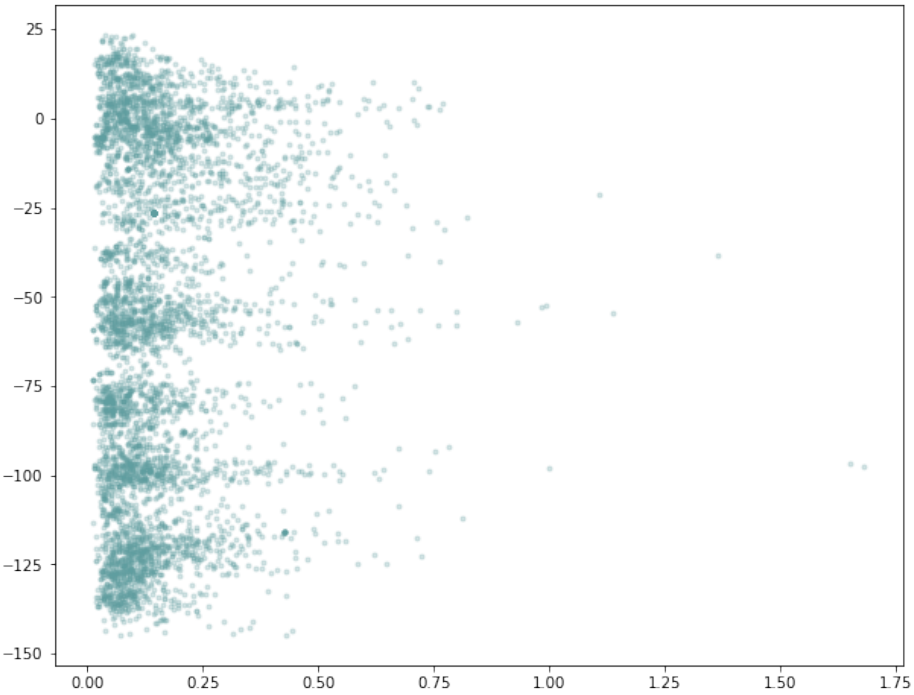


Figure 11: HMC Classifier: α and β exploration samples