# PRACTICAL DATA ANALYSIS ASSIGNMENT

MIS41120 - Statistical Learning

UCD DUBLIN
UCD Michael Smurfit Graduate Business School


**Group 9**

**Siva Thirumavalavan** **20200235**
**Unnikrishnan Thekke Puthumana** **20200006**

# Assessment Submission Form

| | |
|---|---|
| **Student Name** | Siva Thirumavalavan, Unnikrishnan Puthumana |
| **Student Number** | 20200235, 20200006 |
| **Assessment Title** | Practical Data Analysis Assignment |
| **Module Code** | MIS41120 |
| **Module Title** | Statistical Learning |
| **Module Co-ordinator** | Prof. Sean McGarraghy |
| **Tutor (if applicable)** | |
| **Date Submitted** | 16-07-2021 |
| **Date Received** | |
| **Grade/Mark** | |

**A SIGNED COPY OF THIS FORM MUST ACCOMPANY ALL SUBMISSIONS FOR ASSESSMENT.**

**STUDENTS SHOULD KEEP A COPY OF ALL WORK SUBMITTED.**

**Procedures for Submission and Late Submission**

Ensure that you have checked the School's procedures for the submission of assessments.
**Note:** There are penalties for the late submission of assessments. For further information please see the University's ***Policy on Late Submission of Coursework***, (http://www.ucd.ie/registrar/)

**Plagiarism:** the unacknowledged inclusion of another person's writings or ideas or works, in any formally presented work (including essays, examinations, projects, laboratory reports or presentations). The penalties associated with plagiarism designed to impose sanctions that reflect the seriousness of University's commitment to academic integrity. Ensure that you have read the University's ***Briefing for Students on Academic Integrity and Plagiarism and the UCD Plagiarism Statement, Plagiarism Policy and Procedures,*** (http://www.ucd.ie/registrar/)

**Declaration of Authorship**
I declare that all material in this assessment is my own work except where there is clear acknowledgement and appropriate reference to the work of others.

Signed:
Date: 16-07-2021

# Contents

# List of Figures

# List of Tables

# 1. Statement of work

Though we are a team of two, we wanted to explore all the models and test our learning on them. Hence, we have analysed all the 3 models (MLR, SVM and NeuralNet). For evaluation, we would recommend our analysis on Linear and SVM models.

| Team Member | Tasks Completed |
|---|---|
| Siva Thirumavalavan | • Prepared a skeleton document including all the tasks to be completed<br>• Created functions in R<br>• Built LR and SVM models in R<br>• Created the report for the public dataset results |
| Unnikrishnan Thekke Puthumana | • Referred to the practical problems in ISLR textbook<br>• Exploration of data sources<br>• Data Cleaning and exploratory data analysis<br>• Built NeuralNet models in python<br>• Created the report for the Boston dataset results |

# 2. Data Analysis – Boston Dataset

The first set of tests we conducted on the Boston dataset, which comprises demographic and other important information about several Boston neighbourhoods. The column labelled 'crime rate' is utilized as the target in predictive modelling.

## 2.1 Exploratory Data Analysis

We created scatter plots between each predictor factors and the target variable as part of the exploratory data analysis. A trend line is added to these plots based on the least square fit between an individual predictor and the target. Based on this trend line, we can obtain a general idea of which variables might be useful in predicting the crime rate. We can see from the scatter plots that property tax and access to radial roads have notable slopes for their trendlines, which might be significant in estimating the per capita crime rate.

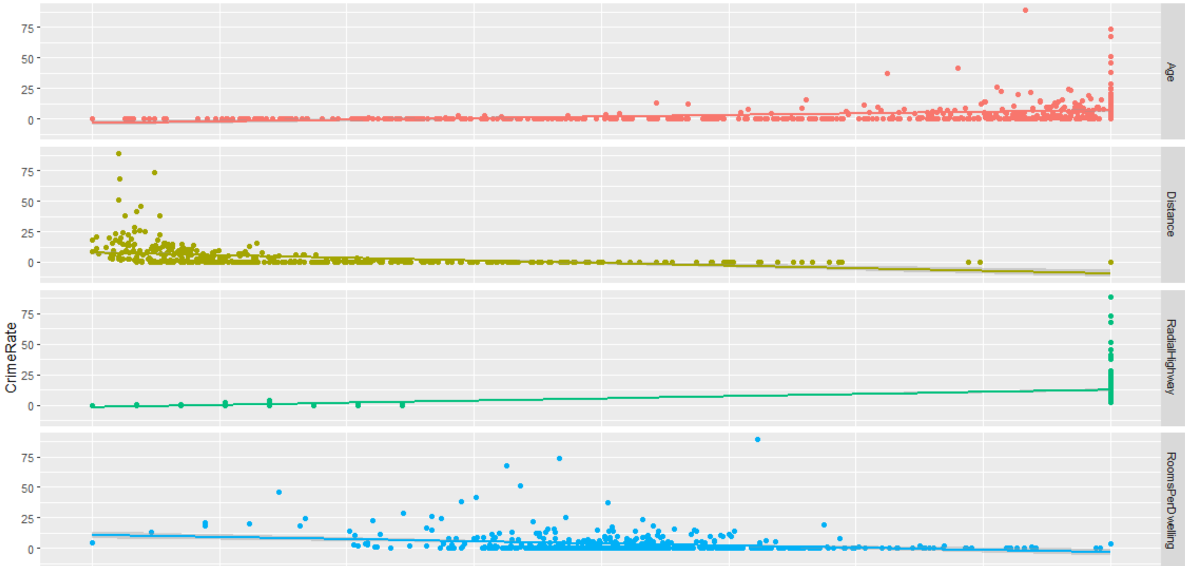


*Figure 1 : Snippet of Target vs Predictors - Boston*

The correlation plot as seen in *figure 2* provides us with the correlation coefficient between the target variable and individual predictor variables. We observe no significant correlation for any of the 13 predictor variables. The direct distance to highway has the most noticeable effect on crime rate out of all the predictors although the correlation coefficient is only 0.625. The variable 'property tax' also shows an equally strong effect on crime rate. However, we notice that accessibility to redial highways and property tax cannot be used simultaneously in the final as there is a strong correlation between these two variables evidently seen in the above figure. Other variables like percentage of lower living status, percentage of black population, and Nitric Oxide concentration show no visibly strong relationship and are likely to be not useful when predicting crime rate. So, in overall we expect that the absence of a good predictor variable is likely to reflect in modelling stage causing larger errors.

As part of our data preparing the data, we scaled the predictor variables using a minimum-maximum scaling technique to bring them to the same range of values between 0 and 1. Doing so enables us to compare the coefficients of various features in the Multiple Linear Regression model and assess their feature importance. After scaling, a multiple linear regression model was built in R to predict crime rate using all the variables available in the Boston dataset. To consistently measure the true performance, we used k-fold cross validation technique which tests the model by evaluating its predictions on validation sets containing unseen data points. With this technique, the data is split into multiple sets of equal size and the model is tested on each of these folds in successive iterations after being trained on the rest of the data. For the Multiple Linear Regression model, we used average RMSE across all the folds to be the metric for evaluating the model and comparing it to other models later. In addition to RMSE, we also report $R^2$ score which measures how good the model predictions are compared to the baseline estimate of mean of the target variable. It represents the variance in the target variable data as explained by the regression line. Due to the varying complexity of various models, we use in this project, we also take note of total time taken for the models to train.

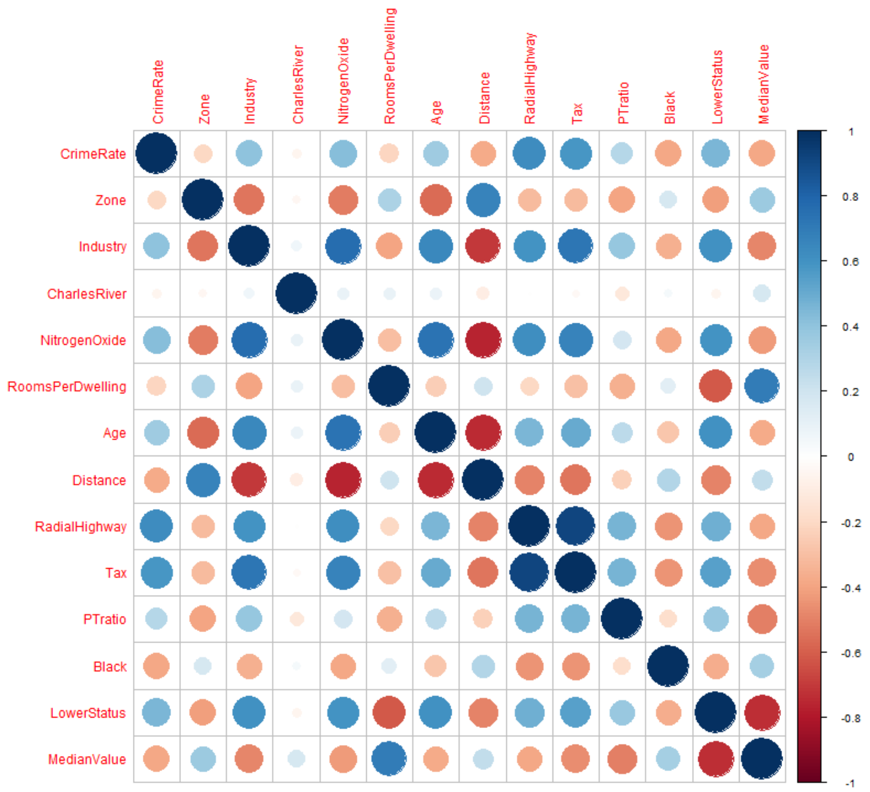


*Figure 2 : Boston - Correlation plot*

## 2.2 Multiple Linear Regression

### 2.2.1 Base model

The multiple linear regression model we built using all the scaled predictor variables resulted in an RMSE value of 8.995. This means that we have made an average error of 8.995 in predicting per capita crime rate in Boston. It corresponds to an R2 score of 0.369 which we find to be insufficiently low and indicates that the model has not learned much from the training data. However, from the exploratory data analysis, it has been clearly seen that the dataset is lacking a good predictor variable. In order obtain the quality of various predictor variables in this model, we performed statistical hypothesis testing for the variable coefficient estimates. Following are the hypotheses that form the basis for the test:

$$Null\ Hypothesis, \quad H_0 : \beta_j = 0\ (The\ estimate\ of\ a\ variable\ is\ zero)$$
$$Alternate\ Hypothesis\ H_a : \beta_j \neq 0\ (The\ estimate\ is\ not\ zero)$$

Using the built-in functionality of R, we get the results for above test for each variable. Using a confidence interval of 95% which corresponds to 2 standard deviations in the standard normal distribution, we select 7 variables as shown in TABLE. The selected variables have p-values less than 0.005 which means that the null hypothesis was rejected for those variables with a confidence of 95%.

*Table 1: Predictors variables selected for Linear Regression Models - Boston*

| Variable | Multiple Linear Regression | Lasso Regression | Ridge Regression | Elastic-Net Regression |
|---|---|---|---|---|
| Zone | | | | |
| Industry | | | | |
| Charles River | | | | |
| Nitrogen Oxide | | | | |
| Rooms per Dwelling | | | | |
| Age | | | | |
| Distance | | | | |
| Radial Highway | | | | |
| Tax | | | | |
| PT Ratio | | | | |
| Black | | | | |
| Lower Status | | | | |
| Median Value | | | | |

| Null Hypothesis Rejected | Null Hypothesis Accepted |
|---|---|

### 2.2.2 Lasso (L1 Regularisation)

In the following stage, we applied regularisation techniques to fit the same data using Lasso, Ridge and Elastic-Net models in R. The main purpose of regularisation was to simplify the model without compromising its performance on the test sets in cross validation. Although regularisation does not improve the RMSE, it helps in eliminating some of the unnecessary variables that have little to no effect in predicting the crime rate. Regularised linear models do this by penalising the addition of more variables on top of the total sum of squared errors.

In our experiment, the Lasso regression model resulted in an RMSE of 9.05 which is marginally worse than the unregularized multiple regression model. However, it ended up being a simpler model by eliminating 3 of the 13 predictor variables: Rooms per dwelling, Age, and Property Tax. It is worth noting that from our exploratory analysis on the data, these variables showed no noticeable correlation with the target variable crime rate. The accessibility to radial highways once again comes out to be the best predictor with a high positive correlation to crime rate. The average distance to employment centres also seems to have a noticeable effect on crime rate but it shows an inverse relationship. This could mean that the farther a locality is from the employment centres, less the chance for crimes to happen. An interesting aspect of this relationship is that employment centres are usually located close to those localities with higher unemployment rates which in turn would have led to increased crime rate per capita. We also note that two more variables are slightly correlated with crime rate: percentage lower status and median value of owner-occupied homes. According to the model predictions, higher percent of population with lower status might result in increased crime rate. On the other hand, large number of owner-occupied houses could potentially mean less crime rate as there is a slightly inverse correlation between them.

We do not perform hypothesis testing for the variable estimates in any of the regularised regression models since they are not useful. When ordinary linear regression yields unbiased estimates of the coefficients, penalised regression results in biased estimates. So, performing hypothesis testing and comparing its p-values against the confidence interval gives us an unrealistic indication of the true significance of a variable.

### 2.2.3 Ridge (L2 Regularisation)

Our ridge regression model also resulted in the exact same RMSE of 9.05 value although the weights of the coefficients are slightly different. The main difference we see between lasso and ridge is in the complexity of these models. As shown in *table 1*, the Ridge regressor fails to eliminate a single variable thus resulting in a relatively more complex model. Given the fact that they both have the same performance based on RMSE metric, we can confidently say that Lasso regularisation is better compared to Ridge as the latter tends to penalise more aggressively.

### 2.2.4 Elastic Net (L1 + L2 Regularisation)

We also used Elastic-Net regularisation with the multiple linear regression model which combines both lasso and ridge penalties. The two parameters in this model were tuned using a simple grid search method keeping minimum RMSE to be the criteria for selecting the best parameter values. FIGURE shows the results from all the trials in our experiment to determine the best values of both these parameters. As seen in the figure, the best version of the elastic net model has its penalty term split in a 30% and 70% combination of lasso and ridge penalties. Although this split makes it closer to the Ridge regression model, we see that the results are more similar to Lasso regression with the same 3 variables being eliminated and the remaining variables being close to that of Lasso regression model. Accessibility to radial highways and average distance to employment centres remain the best two predictor variables as before.

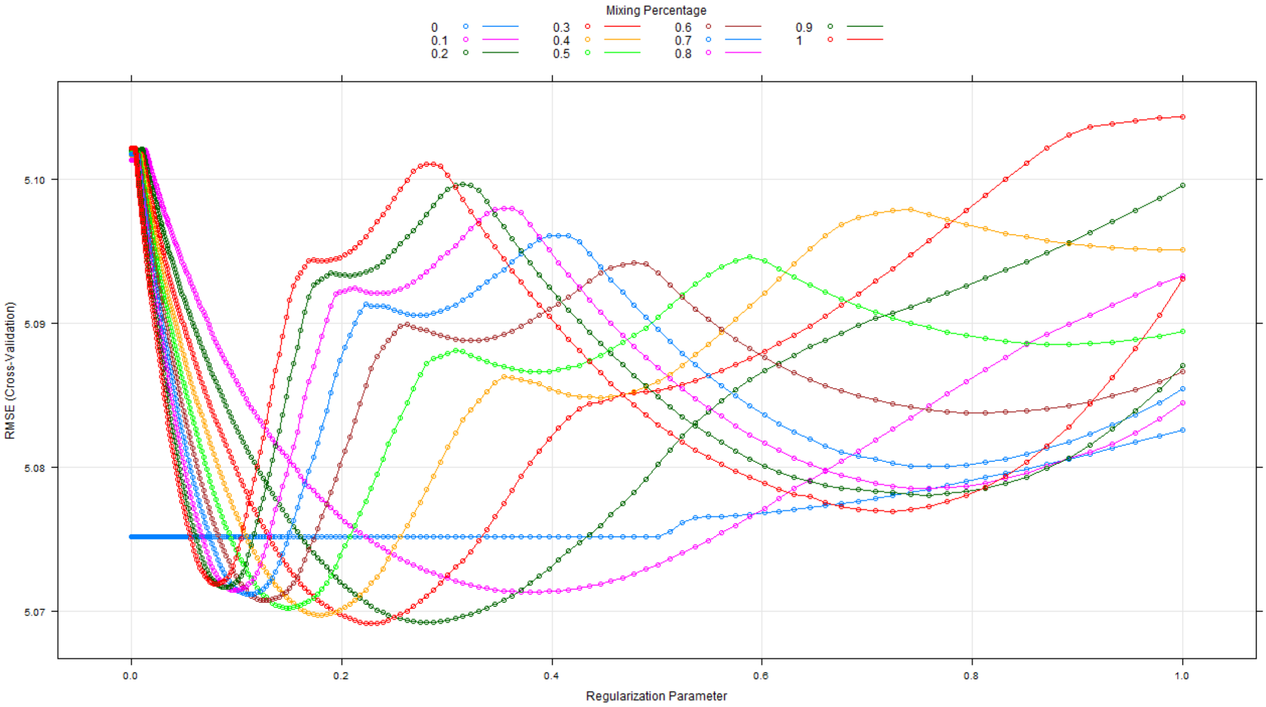


*Figure 3: Grid search for regularised linear model parameters*

## 2.3 SVM

### 2.3.1 Base model

As part of our regression analysis, we also tested a Support Vector Regression (SVR) model on the Boston Dataset. This model came out to be the worst performing model with an RMSE of 9.6 and $R^2$ score of 0.35. We believe this could be because the SVR model would be struggling to learn from the small size of the Boston dataset which has only 516 records. The significance of variables in the model is evaluated using the variable importance functionality in R.

*Table 2: Variable Importance - Boston SVR model*

| **Variable** | **Importance** |
|---|---|
| Zone | 0.05 |
| Industry | 0 |
| Charles River | 0 |
| Nitrogen Oxide | 0.03 |
| Rooms Per Dwelling | 0.07 |
| Age | 0.02 |
| Distance | 0.09 |
| Radial Highway | 0.39 |
| Tax | 0 |
| PT Ratio | 0.03 |
| Black | 0.08 |
| Lower Status | 0.15 |
| Median Value | 0.05 |

Upon a close examination of the variable importance results, we see that the only significant variable in the SVR model is the accessibility to radial highways. All the other variables show negligible effect on crime rate.

### 2.3.2 Regularised SVC Models

As the SVR model does not support regularisation, we converted the regression problem into a classification equivalent. To do so, the target column of crime rate was converted into a binary column using a threshold of 50$^{th}$ percentile. This translates to the lower half of the target column as low crime rate and the rest to be high crime rate. As it is a classification model and the dataset is balanced, we use accuracy as the metric to evaluate the model performance. To make predictions on the new binary target, we use the Support Vector Classifier model with all the 13 predictor variables we used before. In the first model, we applied Lasso regularisation and obtained an accuracy of 80% using a penalty of 0.46. This also resulted in the elimination of several variables which makes it the simplest of all the models we have tested so far. The ridge regularised version of the same model reported a slightly improved accuracy of 82%. However, we do not recommend using it since it used all the variables to predict crime rate. The elastic net SVC model is equally complex as it too kept all the variables although the accuracy has been improved slightly once again to 85%.

## 2.4 Neural Network

In the next part of modelling the Boston dataset, we used a simple perceptron with a single node hidden layer to predict crime rate. Although this is a neural network, it is structurally and operationally very close to a multiple linear regression model. However, from our experiments, we find that this model is able to predict crime rate with a lower RMSE value of 8.01 and higher R2 score of 0.533. An interesting observation from this trial is that the relative significance of the variables has changed with median value of owner-occupied houses being the best predictor variable. Like all the linear models, the neural network model also finds accessibility to highways and distance to employment centres to be important. A noteworthy contradiction we see between all the regression models and this model is that neural network model finds age to be an important predictor whereas it was eliminated by the Lasso regressor. We also see another new predictor being deemed significant: proportion of land for large residential zones. A higher fraction of large residential areas is likely to cause less per capita crime rates, according to this model.

Due to the ease of implementing, we used python to build a multi-layer neural network model using Tensorflow library. Due to the simple nature of the Boston dataset under consideration, we decided to use only two hidden layers with 16, 32 units in the first and second layers respectively. The model was trained using mean squared error as the metric and we ensured convergence of the model by closely monitoring the loss function and metric at every epoch. This model performed extremely well resulting in an exceptional improvement in performance with an RMSE of 5.64. Although we can say that this is the best model and far better than other models we have built, it is innately difficult to explain the predictions made by a multi-layer neural network due to its inherent complexity and non-linear nature. However, we can confidently say that the model is reliable as it was tested on unseen data using K-Fold cross validation in the same way as other models. We also applied regularisation to the multi-layer neural network using the in-built functionalities of Tensorflow. The regularisation penalty is applied for each hidden layer simultaneously which prevents both the layers from overfitting to the data. For both Lasso and Ridge regularisations, we found the optimum value of regularisation penalty using a simple grid search similar to what we had done for linear models. The results indicate that higher penalties result in a low RMSE score on the test set with the optimum value at 9.97.

*Table 3: Predictors selected for NeuralNet and SVM models - Boston*

| Variable | Neural Network (R) | SVR | SVC Lasso | SVC Ridge | SVC Elastic-Net |
|---|---|---|---|---|---|
| Zone | | | | | |
| Industry | | | | | |
| Charles River | | | | | |
| Nitrogen Oxide | | | | | |
| Rooms per Dwelling | | | | | |
| Age | | | | | |
| Distance | | | | | |
| Radial Highway | | | | | |
| Tax | | | | | |
| PT Ratio | | | | | |
| Black | | | | | |
| Lower Status | | | | | |
| Median Value | | | | | |

| | |
|---|---|
| Null Hypothesis Rejected | Null Hypothesis Accepted |

# 3. U.S. Census Dataset

## 3.1 Data description

We utilised a dataset comprising demographic and economic data from US counties, which is available at [Kaggle].

There were 74001 rows and 37 columns in the raw data. Each row represents a small geographic area (Tracts) in the respective county. Based on demographic and economic variables, we built a model to estimate the per capita income in each tract. We removed columns that we believed were irrelevant to our model (Tract ID, County and State). We eliminated the dependent columns such as average income and economic return rates since per capita income was our target variable. We also eliminated Male and Female Population columns since they were highly correlated with the Total Population Column.

We first created our models using the full dataset, but due to its magnitude, it required a lot of computational effort. To address this, we sampled a subset of the original dataset. The final dataset has 10,000 rows and 31 columns.

*Table 4: List of variables in the U.S Census data*

| | | | | | |
|---|---|---|---|---|---|
| TotalPop | Service | PublicWork | Native | Transit | PrivateWork |
| Men | Office | SelfEmployed | Asian | Walk | |
| Women | Construction | FamilyWork | Pacific | OtherTransp | |
| Hispanic | Production | Unemployment | VotingAgeCitizen | WorkAtHome | |
| White | Drive | IncomePerCap | Poverty | MeanCommute | |
| Black | Carpool | Professional | ChildPoverty | Employed | |

## 3.2 Exploratory Data Analysis

We conducted a similar study on the U.S. census data as we did on the Boston data. We plotted each of our predictor variables against our target variable and fitted a trendline. This provided us with an idea of the predictors that will assist us in predicting per capita income.

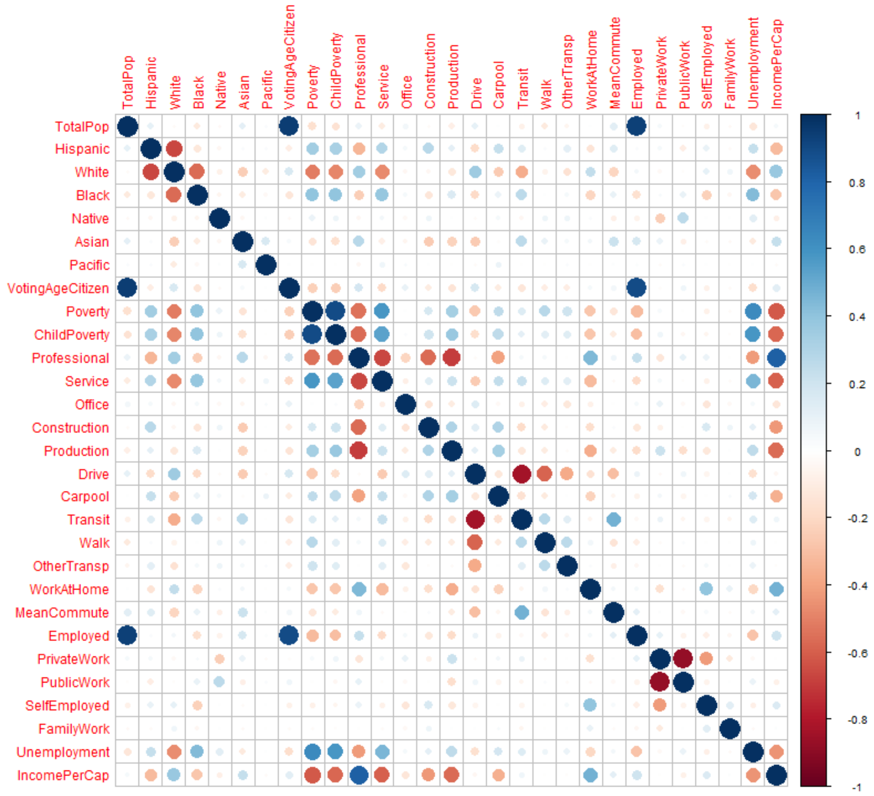


*Figure 4: Correlation Plot - U.S. Census*

*Figure 4* shows that our target variable has a very strong correlation with predictors such as Poverty, ChildPoverty, Professional, Service, and Production. We expect the r-square values to be higher than the Boston models since this data includes some good predictors.

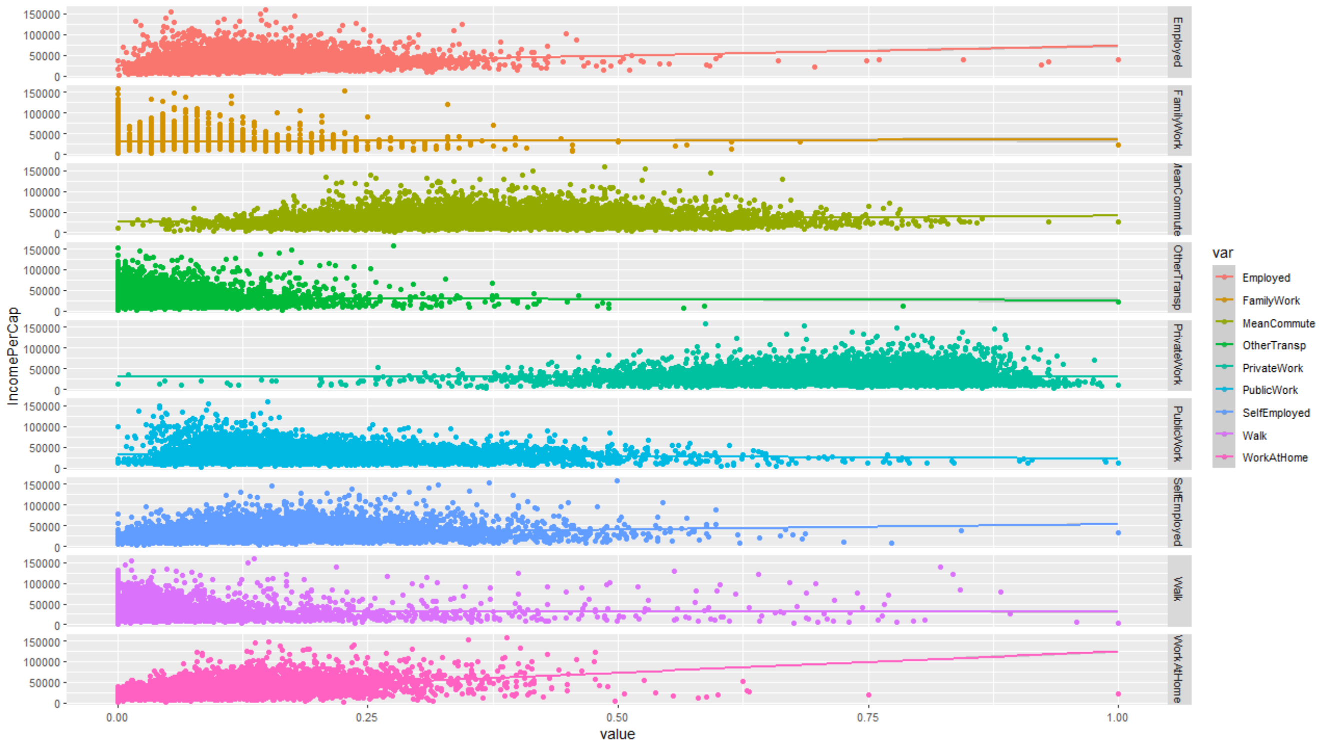


*Figure 5: Snippet of Target vs Predictors - U.S. Census*

## 3.3 Multiple Linear Regression

We scaled the predictors of US census data using the min-max scaling approach, as we did with our Boston dataset, to get all predictors into the same range. This standardisation will allow us to compare predictor estimations. We also use 10-fold cross validation method to assess our model's performance. We recorded the computation time of each model in addition to the RMSE and R-square values.

### 3.3.1 Base model

When we included all the predictor variables in the multiple linear models, the RMSE was 7,438 and the adjusted r-square was 0.756. This indicates that our independent variables explain 76 percent of the variation in Per Capita Income, indicating this is a strong model. By looking at the summary of the model, we can see variables such as TotalPop, Pacific, Poverty, ChildPoverty, Drive, MeanCommute, Employed, PrivateWork, PublicWork, SelfEmployed, FamilyWork and Unemployment have p-value less than 0.05. i.e., we can reject the null hypothesis ($H_0: \beta_j = 0$) with a 95% confidence interval for these variables.

### 3.3.2 Lasso (L1 Regularisation)

We obtained an RMSE of 7,436 and an r-square of 0.756 by using Lasso regularisation on our multiple linear models. While the improvement in RMSE is small, the lasso model achieved the same R-square value as our base model, while eliminating 5 predictors from our model. This is a better model than our base model since it reduces model complexity without sacrificing efficiency.

### 3.3.3 Ridge (L2 Regularisation)

We obtained an RMSE of 7,493 and an R-square of 0.752 for the ridge model. None of our model estimations are made 0 by the ridge model. However, it decreases the estimates of a few of the predictors to a relatively low value and hence lowering their influence on our target variable. We may conclude that the lasso model is superior to the ridge model since it fails to strictly reject variables and does not improve the model’s efficiency.

### 3.3.4 Elastic Net (L1 + L2 Regularisation)

We implemented Elastic net by combing the ridge and lasso regularisation and employed cross validation to fine-tune our hyper-parameters. The resulting model has an R-square of 0.756 and an RMSE of 7434. The Elastic Net behaved quite similarly to our lasso model. This is supported by the alpha (0.6), which makes the elastic net more lasso-like.

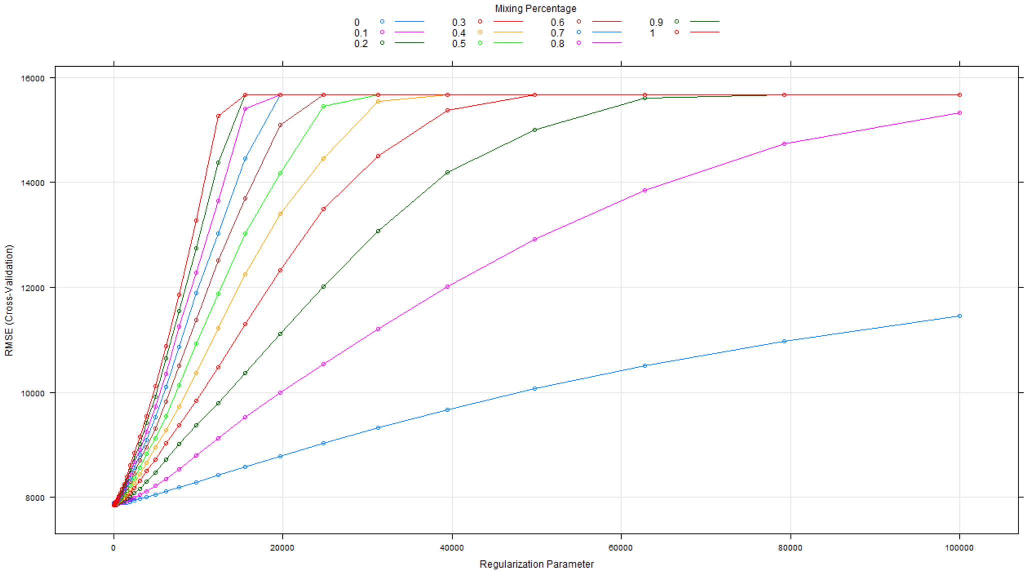


*Figure 6: Grid search for regularised linear models - U.S. Census*

| Variable | Multiple Linear Regression | Lasso Regression | Ridge Regression | Elastic Net Regression |
|---|---|---|---|---|
| TotalPop | | | | |
| Hispanic | | | | |
| White | | | | |
| Black | | | | |
| Native | | | | |
| Asian | | | | |
| Pacific | | | | |
| VotingAgeCitizen | | | | |
| Poverty | | | | |
| ChildPoverty | | | | |
| Professional | | | | |
| Service | | | | |
| Office | | | | |
| Construction | | | | |
| Production | | | | |
| Drive | | | | |
| Carpool | | | | |
| Transit | | | | |
| Walk | | | | |
| OtherTransp | | | | |
| WorkAtHome | | | | |
| MeanCommute | | | | |
| Employed | | | | |
| PrivateWork | | | | |
| PublicWork | | | | |
| SelfEmployed | | | | |
| FamilyWork | | | | |
| Unemployment | | | | |

*Figure 7: Predictors variables selected for Linear Regression Models - U.S. Census*

| | |
|---|---|
| Null Hypothesis Rejected | Null Hypothesis Accepted |

## 3.4 SVM

### 3.4.1 Base model

With Census data, the base SVM model performs the worst of all models. It has the highest RMSE (7,752) and the lowest R-square value (0.752). The standard error and p-value of the variables cannot be extracted from our model summary. Instead, we may compute the variable importance (using the rminer package ) and label the variables with the lowest importance as insignificant by failing to reject the null hypothesis for those variables.

### 3.4.2 Regularised SVC models

We convert the target variable of the Census data to a binary column in the same way that we did with the target variable of the Boston data. To preserve the class distribution, we set the threshold at the 50th percentile (> 50th percentile – high, otherwise – low). We performed cross validation to fine-tune the hyper-parameters, resulting in a model with an accuracy of 89 percent.

On the same data, we used the SVC model with Ridge and Elastic Net regularisation methods. As sparseSVM does not support alpha tuning, we iterated the model for several alpha values and chose the combination that provided the highest accuracy.

*Table 5: Variable Importance SVR - U.S. Census*

| Variable | Importance |
|---|---|
| PrivateWork | 14% |
| Employed | 13% |
| MeanCommute | 10% |
| VotingAgeCitizen | 6% |
| PublicWork | 6% |
| ChildPoverty | 5% |
| Production | 5% |
| Professional | 3% |
| Construction | 3% |
| Carpool | 3% |
| Transit | 3% |
| Service | 2% |
| Office | 2% |
| Drive | 2% |
| SelfEmployed | 2% |
| Asian | 1% |
| Pacific | 1% |
| Poverty | 1% |
| Walk | 1% |
| OtherTransp | 1% |
| WorkAtHome | 1% |
| FamilyWork | 1% |
| TotalPop | 0% |
| Hispanic | 0% |
| White | 0% |
| Black | 0% |
| Native | 0% |
| Unemployment | 0% |

| | |
|---|---|
| Null Hypothesis Rejected | Null Hypothesis Accepted |

*Table 6 : Predictors variables selected for SVM models*

| Variable | SVM | SVM.lasso | SVM.ridge | SVM.elastic |
|---|---|---|---|---|
| TotalPop | | | | |
| Hispanic | | | | |
| White | | | | |
| Black | | | | |
| Native | | | | |
| Asian | | | | |
| Pacific | | | | |
| VotingAgeCitizen | | | | |
| Poverty | | | | |
| ChildPoverty | | | | |
| Professional | | | | |
| Service | | | | |
| Office | | | | |
| Construction | | | | |
| Production | | | | |
| Drive | | | | |
| Carpool | | | | |
| Transit | | | | |
| Walk | | | | |
| OtherTransp | | | | |
| WorkAtHome | | | | |
| MeanCommute | | | | |
| Employed | | | | |
| PrivateWork | | | | |
| PublicWork | | | | |
| SelfEmployed | | | | |
| FamilyWork | | | | |
| Unemployment | | | | |

| | |
|---|---|
| Null Hypothesis Rejected | Null Hypothesis Accepted |

## 3.5 Neural Net

We ran regularised neural net models in R, but as the data amount grew, the computational time grew exponentially. We attempted to create a neural net model with one hidden layer there by emulating a simple linear regression model. Despite achieving good RMSE and R-square values, we were unable to evaluate the model estimates since they did not reflect the observed correlation.

# 4. Conclusion

*Table 7 : Model Comparison - Boston*

| Model | RMSE | R-square | Runtime | Accuracy | alpha | lambda | Comments |
|---|---|---|---|---|---|---|---|
| **MLR** | 8.995 | 0.369 | 0.01 | | | | Provides a good baseline Model is simplified using Hypothesis testing |
| **MLR + L1** | 9.047 | 0.365 | 0.125 | | 1 | 0.089 | Simplified the MLR model by eliminiting 3 variables Without compromising model performance |
| **MLR + L2** | 9.058 | 0.365 | 0.238 | | 0 | 0.484 | Would not recommend it as it has same performance as Lasso Regression yet it is more complex |
| **MLR + L1 + L2** | 9.048 | 0.366 | 1.44 | | 0.3 | 0.223 | Same in performance and complexity as compared to Lasso Regression |
| **NN** | 8.098 | 0.533 | 0.004 | | | | Provides a noticeable improvement over MLR models |
| **SVR** | 9.603 | 0.349 | 0.008 | | | | Worst performing model Struggles to learn from a small dataset |
| **SVC + L1** | | | 0.002 | 0.8 | 1 | 0.46 | Strong baseline for the classification variant Extremely simple model which eliminates many unnecessary predictors |
| **SVC + L2** | | | 0.002 | 0.82 | 0 | 0.46 | Complex model yet no noticeable improvement in accuracy over Lasso SVC |
| **SVC + L1 +L2** | | | 0.003 | 0.85 | 0.1 | 0.48 | Relatively more complex compared to Lasso SVM yet no significantly better accuracy |

*Table 8 : Model Comparison - U.S. Census*

| Model | RMSE | R-square | Runtime | Accuracy | alpha | lambda | Comments |
|---|---|---|---|---|---|---|---|
| **MLR** | 7439 | 0.756 | 0.022 | | | | R-square dropped to 0.46 when only the significant variables were considered |
| **MLR + L1** | 7435 | 0.756 | 0.064 | | 1 | 11.498 | Best model. Less variables but as effective as other linear models |
| **MLR + L2** | 7493 | 0.752 | 0.064 | | 0 | 1204.504 | High lambda values. Could not reject any variables based on estimates |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **MLR + L1 + L2** | 7435 | 0.756 | 0.625 | | 0.6 | 18.307 | Performs similar to lambda but high computational time |
| **NN** | 7539 | 0.758 | 0.055 | | | | Good r-square values. But we cannot accept or reject null hypothesis of any variables |
| **NN + L1** | 7469 | 0.754 | 0.058 | | 1 | | |
| **NN + L2** | 7437 | 0.777 | 0.059 | | 0 | | |
| **NN + L1 + L2** | 7500 | 0.773 | 0.063 | | | | |
| **SVR** | 7752 | 0.752 | 5.731 | | | | Worst model. Has high run time and the least r-square value |
| **SVC + L1** | | | 0.042 | 0.8895 | 1 | 0.007 | All SVC models are similar in performance |
| **SVC + L2** | | | 0.041 | 0.892 | 0 | 0.011 | Least run-time and the best Accuracy |
| **SVC + L1 +L2** | | | 0.052 | 0.892 | 0 | 0.008 | Very similar to the ridge model but higher runtime |