**RM 294 – Optimization I**

**Project 3 – Mixed-integer quadratic programming**

**Nittala Venkata Sai Aditya (vn5227), Muskaan Singhania, Varun Kausika (vsk394), Yingjia Shang**

**Table of Contents**

[**I. Introduction**](https://docs.google.com/document/d/1wlwXOdm0PY3sLgUUHnL5MMqJazZccwFNW8tLZsGZEOg/edit#heading=h.6dll2qejuld)[**2**](https://docs.google.com/document/d/1wlwXOdm0PY3sLgUUHnL5MMqJazZccwFNW8tLZsGZEOg/edit#heading=h.6dll2qejuld)

[**II. Methodology**](https://docs.google.com/document/d/1wlwXOdm0PY3sLgUUHnL5MMqJazZccwFNW8tLZsGZEOg/edit#heading=h.et8r3vaspc3i)[**2**](https://docs.google.com/document/d/1wlwXOdm0PY3sLgUUHnL5MMqJazZccwFNW8tLZsGZEOg/edit#heading=h.et8r3vaspc3i)

1. Direct Variable Selection – MIQP **3**
2. Indirect Variable Selection – LASSO **6**

[**III.**](https://docs.google.com/document/d/1wlwXOdm0PY3sLgUUHnL5MMqJazZccwFNW8tLZsGZEOg/edit#heading=h.6fjm43b4voec) Comparing Lasso and MIQP Models **7**

[**IV.**](https://docs.google.com/document/d/1wlwXOdm0PY3sLgUUHnL5MMqJazZccwFNW8tLZsGZEOg/edit#heading=h.c61bhg72e9pi) Recommendations **9**

1. **INTRODUCTION**

One of the most common problems in predictive analytics is variable selection for regression. Direct variable selection using optimization has long been dismissed by the statistics/analytics community because of computational difficulties. It is important we select only the best ones which give all the information while ensuring that the model complexity is kept to a minimum. But it is not possible always so we have to have a trade-off between the model complexity and the variables.

The total number of models that can be picked are:

Diagram

Description automatically generated with low confidence

This computational issue was part of the motivation for the development of LASSO and ridge regression. With the advancement of new technologies, we can better decide and compute beta coefficients for computationally expensive operations. This is the motivation behind this project where we will aim to compare the results we get from Gurobi and Lasso regression. Either of those two can be used in future for such problems.

1. **METHODOLOGY**

We try to solve the problem with 2 approaches. The first one is via Gurobi where we will use the Mixed Integer Quadratic Problem and the second approach is via LASSO regression.

The first thing we need to check is if there any null values in both train and test datasets. Null values can give false results with our models so it is important we take care of them.

A picture containing chart

Description automatically generated

A picture containing table

Description automatically generated

We see that both train and test dataset don’t have any missing values so we can proceed with the model building.

1. **Direct Variable Selection – MIQP**

There are a few steps we need to consider while formulating the MIQP problem:

**Decision variables:**

The 𝑚 coefficients of the variables:

****

The binary 𝑧𝑗 variables which decide which 𝛽𝑗 variables are 0



The objective is to minimize the loss function.

A screenshot of a computer

Description automatically generated with low confidence

where m is the number of variables to be selected

**Constraints:**

1. First 𝑚 set of Big-M constraints:



1. Second 𝑚 set of Big-M constraints



We choose the value of 𝑀 to arbitrarily be 100 (sufficiently large value).

1. Total number of variables picked should be less than or equal to k:



The main objective while training this model is to select the optimal value of k which gives the lowest error. We will select the value of k by devising our own cross validation function

**10 – Fold Cross Validation**

The main advantage of cross validation is that we use the same dataset for getting the best fit model by plotting 9 parts into training set and keeping the last one as validation dataset. This ensures that the model is not overfitted and is able to cater to all data points and insights. Once all parts are modelled and tested, the average of the errors between the prediction obtained from the one validation set and the actual value is calculated. The aim is to select the value of K (number of variables selected) which gives the lowest error. We have 50 predictor variables in our dataset. We will start our k value from 5 to 50 with an increment of 5 i.e., k = 5,10,15,20,...,50.

We used np.split function on a sample created from training dataset and split it into 10 folds. The loss metric used was Sum of Squared Errors. The aim was to reduce this by the formula

A picture containing text, watch

Description automatically generated

We plotted the SSE against various values of k as seen below and found that the best number of variables for the MIQP model came out to be 10,  the other variables are noise and contribute to overfitting.

Table

Description automatically generated with medium confidence

Chart, line chart

Description automatically generated

Once we get the optimal number of variables, we use that to fit the MIQP model and get the coefficients of beta

Table

Description automatically generated

1. **Indirect Variable Selection – LASSO**

The second model we will try is by doing Lasso Regularisation. The main purpose of the Lasso Regularisation is to penalise high overfitting. Lasso shrinks the coefficient estimates towards zero and it has the effect of setting variables exactly equal to zero. When lambda is small, the result is essentially the least squares estimates. As lambda increases, shrinkage occurs so that variables that are at zero can be thrown away.

In Lasso, the variables are selected as

A close-up of a document

Description automatically generated with low confidence

Here, variables are set to 0 when lambda is large, penalizing the cost function. Note that β0 is not included in the penalty term. When lambda is very large, the prediction of the regression is just the mean of the target variables.

Similar to MIQP problem, we need to tune our hyperparameter (in this case lambda) to give the best possible fit which will reduce the SSE.

We used the LassoCV function in python which has a built in Cross validation function that performs the same task as in MIQP problem and helps select the best lambda.

On training the model, we found the best parameters as below:

Graphical user interface, text

Description automatically generated

Optimal number of variables in Lasso is 17 and the lambda that came out was 0.076.

1. **Comparing Lasso and MIQP Models**

On comparing the two methods, we found that there is not much difference in the predicted values used from both models but the optimal number of variables selected is different in case of lasso (17) and MIQP (10)

Table

Description automatically generated

We further plotted a scatter plot between the predicted values of both methods as shown below:

Chart, line chart

Description automatically generated

As can be seen from the graph, there is not much difference between the two in terms of predictions. The points plotted from same methods are very close to the diagonal.

**Comparing coefficients**

**Chart, box and whisker chart

Description automatically generated**

On comparing the beta values that were obtained form both methods, we find that there is not much difference between the two. The graph looks identical for both cases which further validates our original inference that there is not much difference in the two models. There are many coefficients that are almost zero in lasso. This implies that as far as subset selection, it is not as good as the MIQP, but very close.

1. **Recommendations**

In terms of subset selection, we find that solving the MIQP is more accurate and exactly zeros out the required variables, whereas if only prediction is necessary, there is not much use to solving the harder problem (MIQP) and would be more advantageous to select the computationally cheaper LASSO.