Optimization: Project 3
Group 13

Names: Aditya Soni, Anant Gupta, Casey Copeland, Jessie Lee

1. Objective

LASSO- Least Absolute Shrinkage and Selection Operator is frequently used in our firm for linear regression. In this study, we will test a direct variable selection method and see if it builds a better, more accurate model, and save computation time using the sample set provided.

The direct variable selection approach is MIQP - Mixed Integer Quadratic Programming which utilizes Gurobi optimizer. We will then compare the result with LASSO using the ScikitLearn package as our baseline approach.

The MIQP method involves 2 main steps. Step 1 is to use 10-fold cross validation of the training set to select the optimal numbers of X variables to predict Y with minimum loss (square error). Step 2 is then to find the weights of the chosen number of X variables in the training set to train the model and then to predict the y of the test set.

LASSO is conventionally used in our firm where we use 10-fold cross validation of the training set to search for the optimal shrinkage factor λ. Then we use the chosen λ to build models based on the training set. Then we predict the y and compare the overall performance of the model with the MIQP model.

2. Methods

2.1 Direct Variable Selection - MIQP

The picture below depicts our objective function, constraints, and variable types.

$$\min_{\beta,z} \sum_{i=1}^{n} (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} - y_i)^2$$

$$s.t. -Mz_j \leq \beta_j \leq Mz_j \quad for\ j = 1, 2, 3, \ldots, m$$

$$\sum_{j=1}^{m} z_j \leq k$$

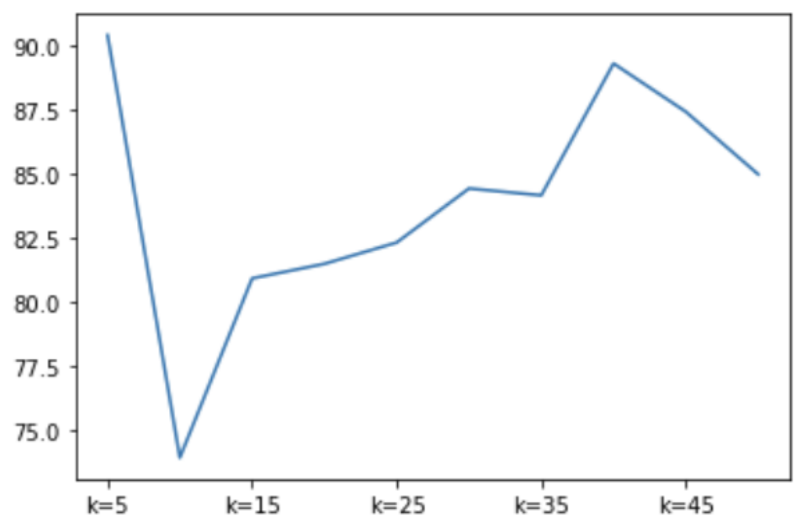
$$z_j \ \ are\ binary.$$

Since there are 50 predictors (X) so we have a total of 50 weights plus a constant in the constraint. In addition, we add 50 binary variables to constrain the total of weights in the model, resulting in 101 constraint variables.

The first 100 constraints were such that when a predictor X is chosen as it's corresponding binary variable is set to be 1, or turned on, their value has to be between [-M,M], where M is chosen to be 1000 here (big M). And then the chosen number of 'turned on' weights sums to be K of choice, resulting in a total of 101 constraints with 101 constraints variables.

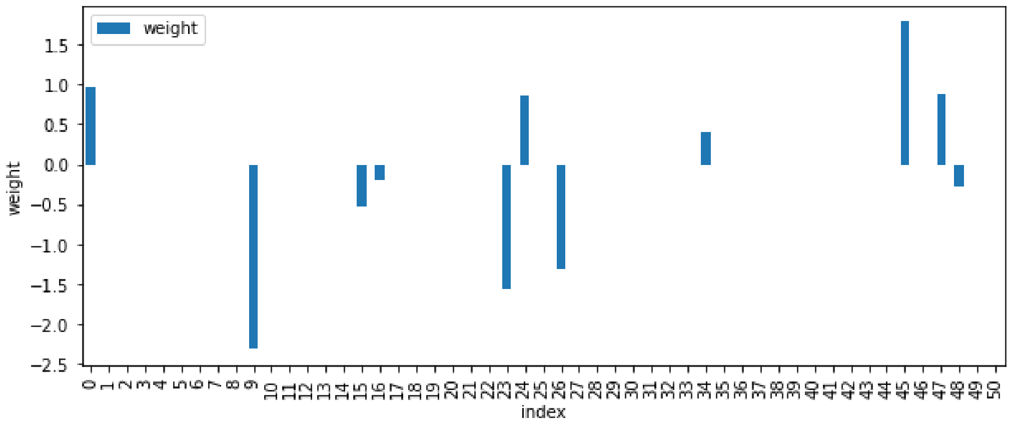
We then conducted cross validation to test different k, increasing the amount of X variables included in the regression and seeing how average error changes. The below plot reports the sum of squared errors at different levels of k. The model has the smallest cross validation error when there are 10 X variables included in the model. This corresponds with the gurobi MIQP optimization output (above). The output includes 11 β s, an intercept term and 10 X variables, that create the optimal regression model.

k = 10 has the lowest CV error.



Then using the max allowed predictor of 10, we get the following weights with weight index 0 being the intercept term and all weights between ~-2.5 to ~2.

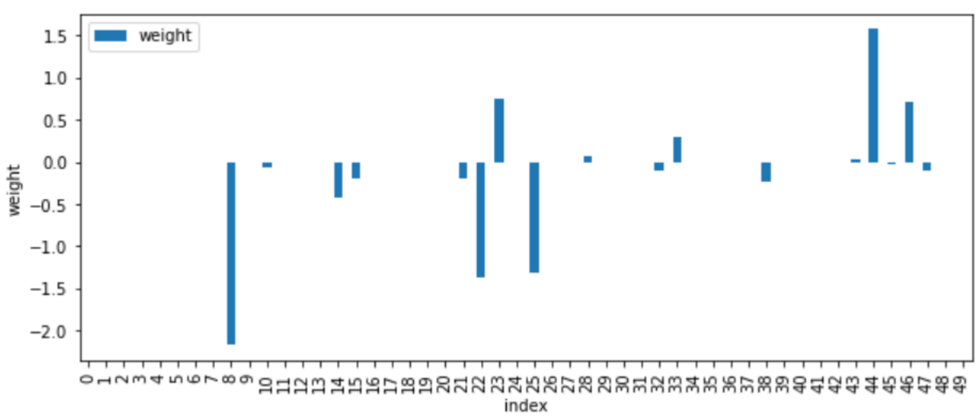
```
[ 0.97252408,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        , -2.30820726,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
 -0.51832612, -0.20416201,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        , -1.55914318,  0.86697336,
  0.        , -1.31191942,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.4081653 ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  1.78147489,  0.        ,  0.88738292, -0.28229213,  0.        ,
  0.        ])
```



2.2 Indirect Variable Selection - LASSO

For our LASSO variable selection method we used Scikit-learn to pick the optimal tuning parameter λ of 0.070548, based on a grid search with a 10-fold cross validation on training data. LASSO cross validation resulted in a model with 18 predictors.

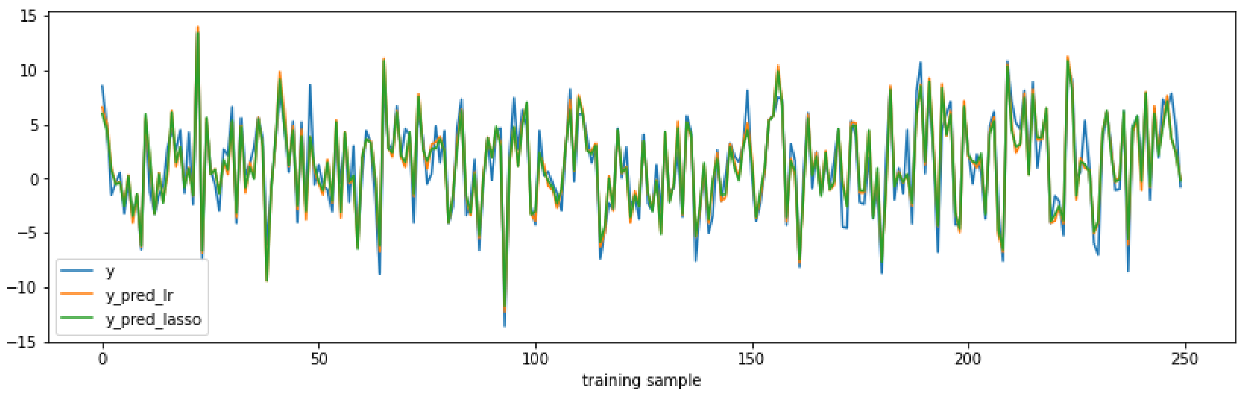
```
array([-0.        , -0.        ,  0.        ,  0.        , -0.        ,
        0.        , -0.        , -0.        , -2.16759923,  0.        ,
       -0.06543049, -0.        , -0.        , -0.        , -0.41968302,
       -0.1952272 ,  0.        ,  0.        , -0.        ,  0.        ,
        0.        , -0.1939836 , -1.36867927,  0.74620669, -0.00605903,
       -1.30816494, -0.        ,  0.        ,  0.05717854,  0.        ,
       -0.        ,  0.        , -0.09717768,  0.29386359,  0.        ,
        0.        ,  0.        ,  0.        , -0.23972899,  0.        ,
       -0.        ,  0.        ,  0.        ,  0.03733622,  1.56934772,
       -0.02949255,  0.71091105, -0.09588212,  0.        ,  0.        ])
```



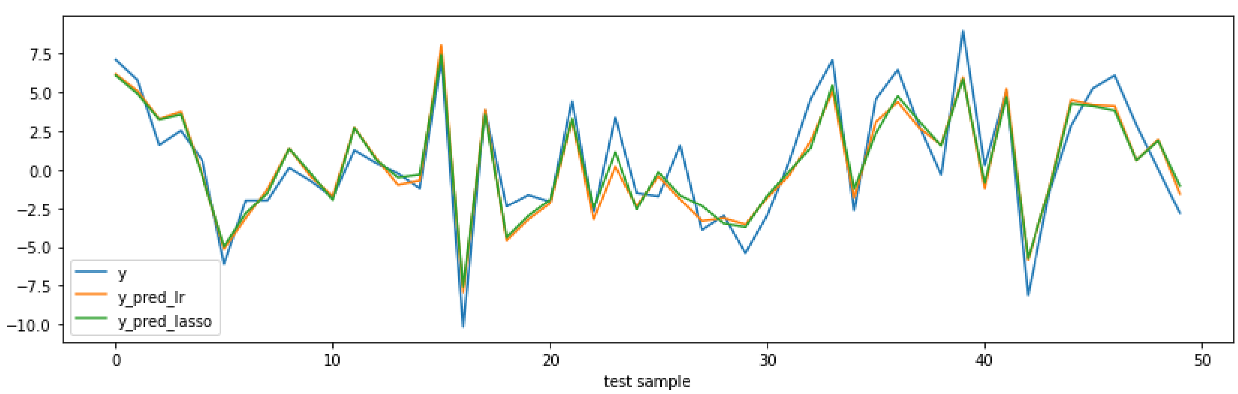
3. Comparison of MIQP and LASSO model selection

The two model selection methods result in a similar MSE for both training and test sets, with LASSO performing slightly better for the training set while linear regression performing slightly better in the test set. The predicted Y are similar to each other (see figures below).

	lasso err	linear reg err
train	2.356777	2.391985
test	2.346139	2.336544



.



The main difference is the computing expense. The MIQP model selection with cross validation involves a loop of 100 model fitting processes using Gurobi optimizer, which involves integer programming with quadratic objective function. This cross validation process is very computational intensive and took a few hours to find the optimal matrix. The prediction is pretty fast when the optimal K is decided from cross validation. To be comparable to LASSO, we may need a denser grid of k with increments of 1 or 2, which would future increase computational demands.

In contrast, our conventional LASSO cross validation uses out of the box tool kit and the process is very computationally efficient. Compared with a few hours, it only took a few minutes to compute.

This study suggests that for this particular dataset, LASSO and MIQP perform similarly in loss metrics, with LASSO being much more efficient in finding the solutions, thus we recommend continuing to use LASSO for this type of dataset. Though, it’s always good to perform a similar exercise to cross check for the best prediction model using MIQP when time and resources allow.

While they produce similar metrics and it is sufficient to use either method, we must consider how the betas chosen will impact our application of the model and the recommendations we give clients. We should keep in mind what predictive relationship we are modeling, and the importance of our chosen independent variables. Say we are predicting demand y, using different business functions and economic factors, the x values. The x variables chosen could shift company operations and focuses. This is where expertise beyond data comes into play. Since both methods are available, it may be optimal to cross reference betas chosen by each MIQP and LASSO, and further conclude on business decisions. If the team is most familiar with LASSO, it is valuable to continue using it. Moving forward, we recommend to cross reference both LASSO and MIQP, taking into consideration the business relationship you are modeling, as well as staying up to date on MIQP advancements as programming continues to develop and evolve.