Final Report - Predicting Crimes in Philadelphia

Fiona Chin (flc38), Erik Enriquez (eee37), Yumeng Niu (yn227)

Abstract

The rate of crime has been decreasing since 2006 in Philadelphia, but how can policemen use data on different crimes to make the city safer? Our team's objective is to predict the number of crimes on a monthly basis to allow for policemen to allocate their men efficiently. Our models focus on predicting crime numbers on burglary and general crime. We have built multiple models using transformed input spaces, linear regression techniques and Generalized Low Rank Models to predict crime rates. Out of all the models we have trained, validated, and tested, the team concludes the best model to predict crime numbers is applying Proximal Gradient with Quadratic Loss and a Quadratic Regularizer to deseasonalized burglary data. Although this is an effective model, it can be dangerous if used carelessly and might be considered a Weapon of Math Destruction.

Data Description

Our goal for the project is to predict future numbers of different crimes in Philadelphia. That way, the police can prepare for future offenses in Philadelphia. We plan to reach this goal using Philadelphia Crime Data. This dataset comes from OpenDailyPhilly.org, whose primary goal is to make data available to promote business and reach out to its citizens. This dataset dates from 2006 to 2016 and consists of 14 features and 2,154,123 examples of different crime incidents reported in Philadelphia. The data can be split into time-related information, location-related information, and type of crime related information about the crimes in Philadelphia. But, there weren't enough features to build a crime-predicting model in the main dataset. Thus, the team decided to focus on how crime data is affected by weather patterns and yearly general statistic indicators.

Weather information in Philadelphia dates from 2006 to 2015 containing indicators such as temperature, precipitation and snow depth (3652 examples, 10 features). The yearly statistic indicators were extracted from OpenDailyPhilly.org from the Community Health Assessment containing yearly data from 2000 to 2012 (15 examples, 8 features). It consists of unemployment rates, cancer mortality, homicide mortality, and other indicators over the years.

The team decided to reshape the data by aggregating it monthly and summing up the number of crime incidents per month. This allows us to build a model that does not focus on too fine a granularity in detail. After joining the data sets, our team decided to use data from 2006 to 2012 where our training set is from 2006 to 2011 and we use 2012 data for our test set. This selection allows the team to implement time-theory analysis and autoregressive models onto our data. In addition it simulates the environment future policemen would use our model for. Our data set now consists of 44 features and 2,099 examples.

Approaches/Algorithms

Approach 1 Apply Different features to the X Input Space for Burglaries

In our midterm report, we created models using 1-2 features and an offset to predict the number of burglaries. The first model used only average temperature and an offset. The second model was the same as the first model and includes looking at previous 12-month data. The third model is the same as the first model and included unemployment and the number of good air quality days indicators. Thus, the team decided to create another model using all the features in our training set. We split the training and validation sets 80:20.

Approach 2 Apply Different Loss Functions using Proximal Gradient Linear Search for Burglaries

Using the all features model, we want to determine what loss functions would best fit the training set. By applying Proximal Gradient (proxgrad.jl) linear search, we attempted Quadratic loss (QuadLoss), L1 loss, and Huber loss using a stepsize following the Lipshitz derivative and using a max iteration of 1000000. All these models did not include regularizers. Quadratic loss averages out the model, while L1 loss finds the median of the model, and Huber loss is used to make sure the outliers won't have too much impact on the model.

Approach 3 Apply Different Regularizers to the Proximal Gradient Method with QuadLoss for Burglaries

We continued to build on our models by adding regularizers to the QuadLoss model. We implemented QuadLoss with a Quadratic Regularizer and a l1 Regularizer with lambda = 1. Both regularizers produced almost the same model on the training and test set.

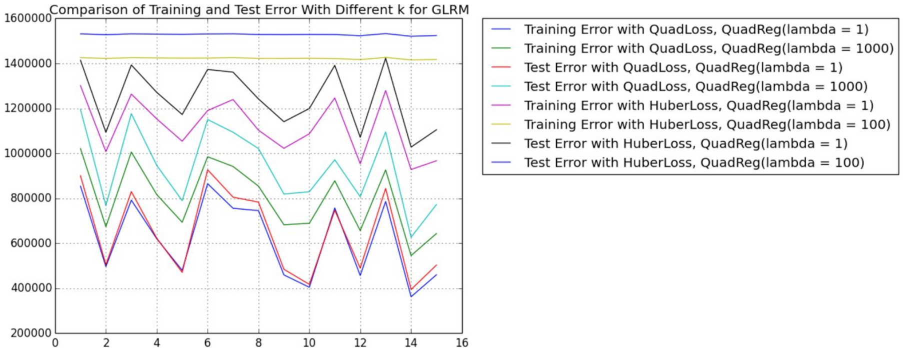
Approach 4 Apply One Hot Encoding for Types of Crimes

One concern we have with our previous approaches is that we are only using data for one type of crime when we are training our models. Therefore, we also want to explore ways to utilize all types of crimes. The first idea we tested is to feature transform the type of crimes with one hot encoding. With this encoding, we could test whether there could be a general model that produces good prediction based on which type of crime it is. More importantly, it would be able to use all the data.

Approach 5 Apply GLRM to Find Vector Representation of Types of Crimes

As the one hot encoding is rather primitive, especially given the fact those types of crime is a categorical feature with high dimension (25 types of crimes). Hence, in addition to one hot encoding, which add 25 boolean features to our dataset, we also experimented vectorizing this feature into numerical matrices of lower rank. In other words, we fitted a generalized low rank model to our dataset, using OvALoss(25) for the feature types of crimes.

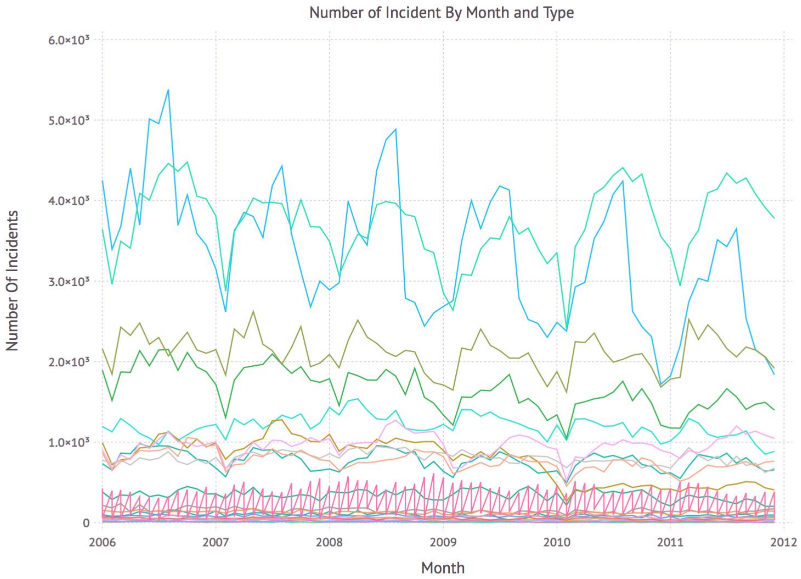
For this approach, a key decision is the rank k of GLRM. In order to find the optimal k for our dataset, we experimented k = 1:15, and plotted the training and testing error for the supervised learning problem with the type of crimes feature replaced by k numerical features.



(x-axis is the rank k, y-axis is the mean square error)
As shown in the graph above, the optimal k for the supervised learning model should be Quadratic Loss with Quadratic Regularizer of lambda = 1, and k = 14.

Approach 6 Applying Quad Loss with Quad Regularizer to Deseasonalized Burglary Data

When analyzing our data we discovered that crimes followed an annual seasonal trend (see picture . Crime would increase in the summer and decrease in the winter. So we decided to try deseasonalizing the number of crimes data. We deseasonalized crime data by calculating and then subtracting the mean number of each type crime for each month from the actual number of crime. The means were calculated using only training set data. For example, the deseasonalized 2006-01 data for Burglary is computed by subtracting the mean number of Burglary incidents for January of 2006-2011. We then applied a quadratic loss with a quadratic regularizer to our burglary data and measured the mean squared error.

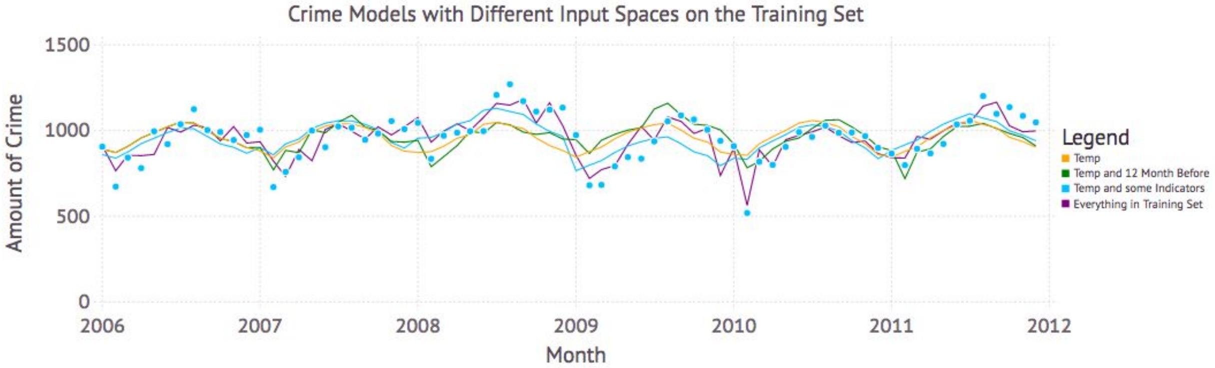


Results

Results From Fitting Burglary Data (Approach 1-3)

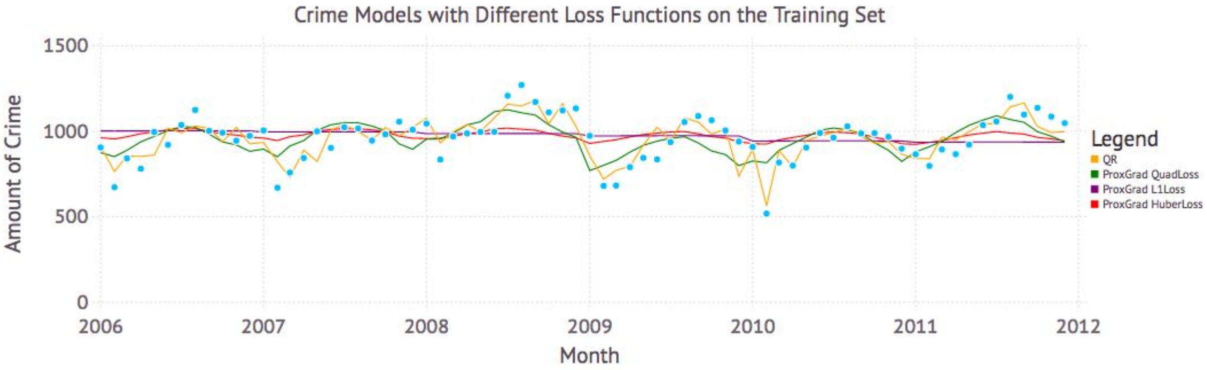
1. **Approach 1 Apply Different features to the X Input Space for Burglaries**

    Comparing the three simple models using QR factorization, we observed that the model using all the features trained the best, as shown by the graph below.



2. **Approach 2 Apply Different Loss Functions using Proximal Gradient Linear Search for Burglaries**

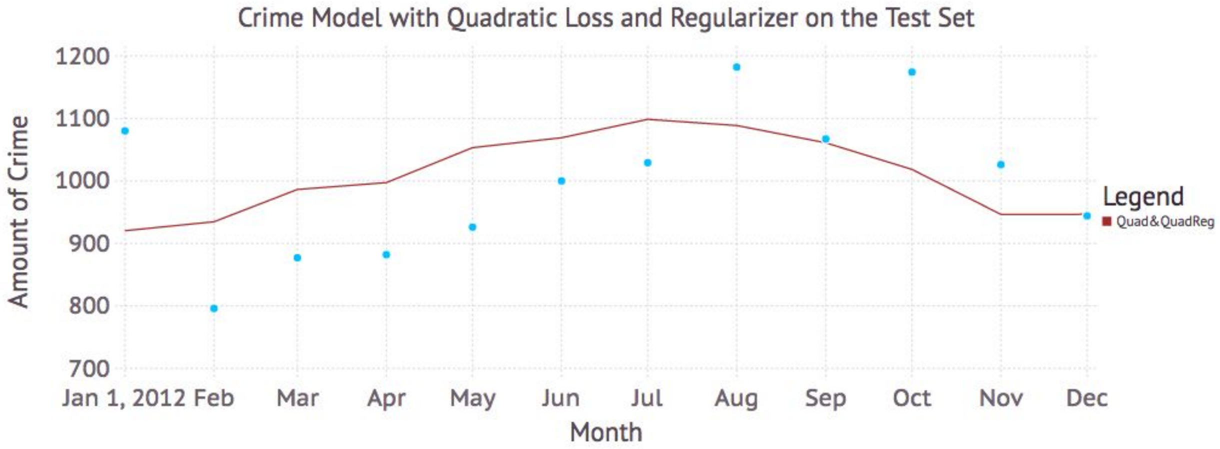
    When compared to QR factorization, we found that QR factorization best modeled the training set, followed by the Proximal Gradient (ProxGrad) with QuadLoss. The models had Mean Square Error (MSE) scores on the training set of 3664.89 and 10175.82. However, these models MSE are different when fitting on the validation set. The QR factorization model and ProxGrad model with QuadLoss had MSE scores of 18649.51 and 17973.82 respectively. Thus, we looked further into generalizing the QuadLoss function.



3. **Approach 3 Apply Different Regularizers to the Proximal Gradient Method with QuadLoss for Burglaries**

    As we have mentioned previously, both L1 and quadratic regularizers produced almost the same model on the training and test set. The model with the Quadratic Regularizer scored an MSE of 10175.92 on the training set and a score of 17973.25 on the validation set. The model with the L1 Regularizer scored an MSE of 10175.89 on the training set and a score of 17973.66 on the test set. The regularizers did not impact the model too much. Thus, the best model is the Proximal Gradient Method with Quadratic Loss and a Quadratic Regularizer. The model scored an MSE of 11252.77. Although the MSE score changed between the train and test set, the error is still pretty high for both sets. Thus we look at other models for general crime numbers.

Method	Training Error[1]	Validation Error	Test Error
QR	3,664.89	18,649.51	
QuadLoss ZeroReg	10,175.82	17,973.82	
HuberLoss ZeroReg	17,845.94	24,492.71	
L1Loss ZeroReg	17,898.02	24,663.03	
QuadLoss QuadReg	**10,175.92**	**17,973.25**	**11,252.77**
QuadLoss L1Reg	10,175.89	17,973.66	



Note about Generalization of Burglaries on Other Types of Crimes

We have found that Quadratic Loss with Quadratic Regularizer fits the Burglary crime incidents well. However, we also want to examine how well this model generalizes to 24 other types of crimes in our dataset. After fitting a different w using Quadratic Loss with Quadratic Regularizer model, we noticed that some types of crimes with more monthly criminal incidences do not work well with this model. For example, the model for Vandalism and Other Assaults under fits the data. In addition, Thefts, All Other Assaults, and All Other Offenses have test error dramatically larger than that of Burglary fitting.

Type of Crime	Training Error	Test Error
Vandalism/Criminal Mischief	11,281.48	7,088.00
Thefts	58,452.48	650,165.01
Other Assaults	17,240.66	35,009.47
All Other Offenses	160,909.71	1,492,132.69

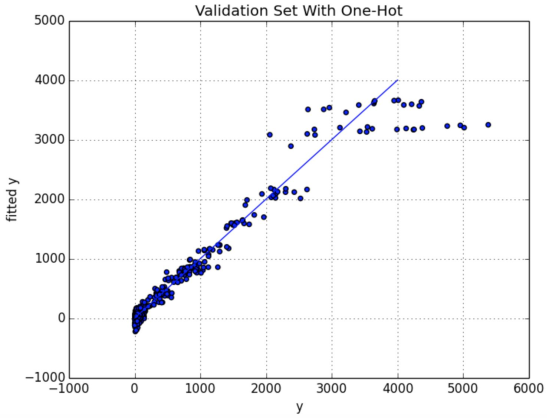
[1] The errors shown here are mean square error

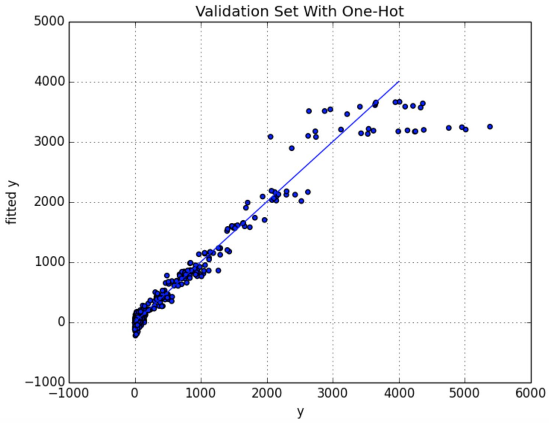
Results From Fitting All Types of Crimes Together (Approach 4,5)

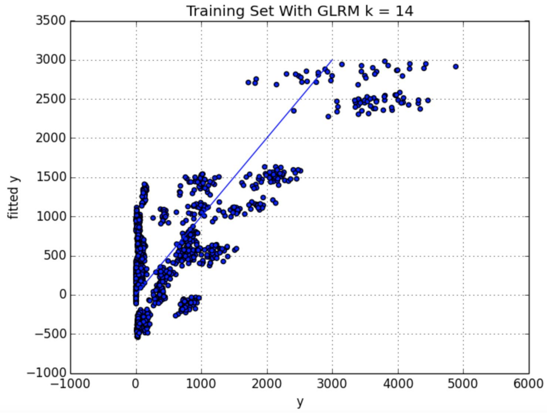
1. Analyze the results from two approaches

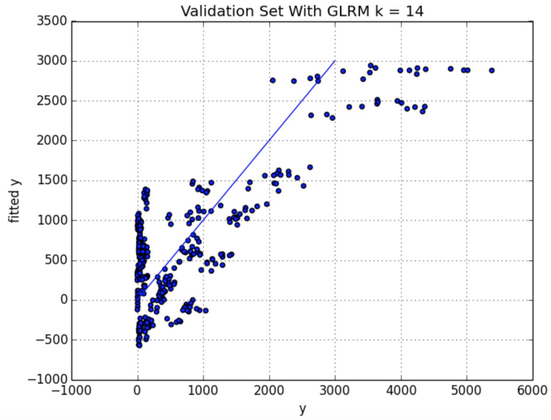
Method	Training Error	Validation Error	Test Error
<u>One-Hot</u>	**34,000.85**	**77,455.92**	**34,589.03**
GLRM	378,383.80	462,365.70	NA

From the error shown above, it seems like Approach 4: one-hot encoding of the types of crimes feature is performing better than Approach 5: GLRM with k = 14. Similar observation could be obtained by the graphs below comparing numbers of criminal incidents and the predicted numbers.









Meanwhile, it seems like types of crimes with monthly number of incidents greater than 2000 is not predicted well with Approach 4, whereas crimes with fewer incidents was fitted much better.

2. Removing Outliers

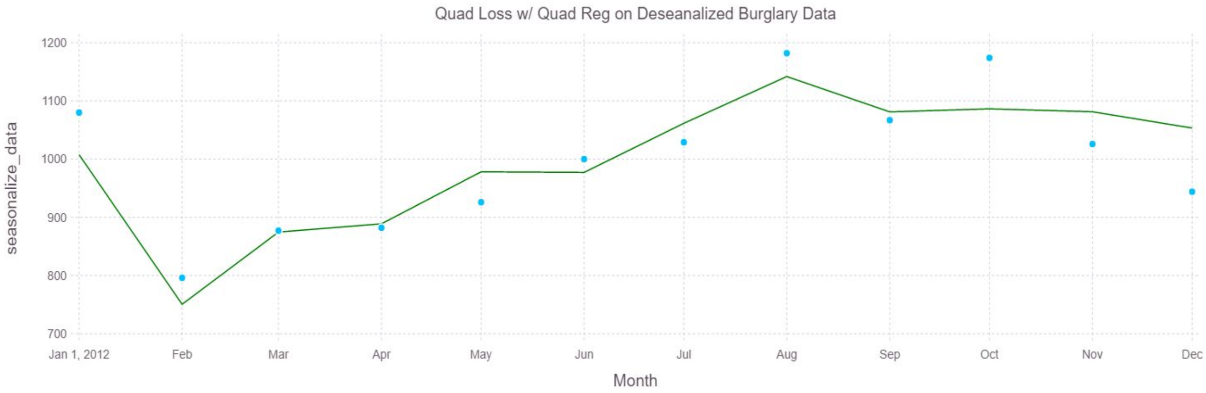
As discussed previously, some types of crimes are outliers to our model. Therefore, we removed the three types of crimes that we found for Approach 1-3 that had higher test errors, Thefts, All Other Assaults, and All Other Offenses, and repeated our analysis. As shown below, errors for both models drop by one magnitude from the original result.

Method	Training Error	Validation Error	Test Error
One-Hot	**8,417.66**	**9,401.39**	**7,874.77**
GLRM	68,720.03	77,257.31	

Results From Fitting on Deseasonalized Burglary Data (Approach 6)

After fitting a model by applying quad loss with quad regularizer to deseasonalized burglary data we re-seasonalized our fit and our data to understand and visualize our prediction. We obtained a mean squared error of 3014.64 on our test datasets using this model. Below is a plot of the model on the test set.

Method	Training Error	Test Error
Deseasonalization	2,759.75	3,014.64



We have also tested how well this method generalizes to other types of crimes. Similar to previous results, the method generalizes well to most types of crimes, except for outliers Thefts, All Other Offenses, and All Other Assaults.

Conclusion

Overall, we have been trying to predict the monthly number of incidents for each types of crime in the city of Philadelphia. We have tested three main categories of models to this problem.

Categories of Approaches	Approaches	Best Training Error	Best Test Error
Fitting One Type of Crimes - Burglary	1, 2, 3	10,175.92	11,252.77
Fitting All Types of Crimes Together	4, 5	8,417.66	7,874.77
Fitting Deseasonalized Burglary Data	**6**	**2,759.75**	**3,014.64**

The table on the top shows that **our best model was Quad Loss with Quad Regularizer to Deseasonalized Burglary Data, which has a test mean square error of 3,014.64**. Intuitively, this is a reasonable result, as we have noticed the existence of annual seasonality during the data exploration phase. Hence, it is likely that each month of the year has a distinct baseline for the number of criminal incidents. A linear model of other indicators like temperature and unemployment might be capable of modeling the number of criminal incidents beyond this baseline, but not modeling directly the seasonality.

In addition, we have observed across the three categories of approaches that 3 types of crimes are consistently different from the 22 other types of crimes. They are Thefts, All Other Offenses, and All Other Assaults. All the approaches generalize well with most types of crimes (with different fitted w), excluding the outliers.

Moreover, we are confident of the out-of-sample error, because we used the train - validation - test method to select our models when multiple models were compared with the same input set. Therefore, the test error should be a fair estimation of the out-of-sample error.

However, we do not recommend that our results be used in production because our predictions can have negative consequences in society. It is possible that our predictions could create self-fulfilling (or defeating) feedback loops. By utilizing more officers on months where we predict a high level of crime, it is possible that more crime will be reported on those months because of the increase of police surveillance. In months where a low level of crime is predicted, less crime could be reported on those months because of the decrease in police surveillance.