Final Project

Kushal Ismael and Matt Kosko

6/23/2021

Predicting Violent Crime Rates

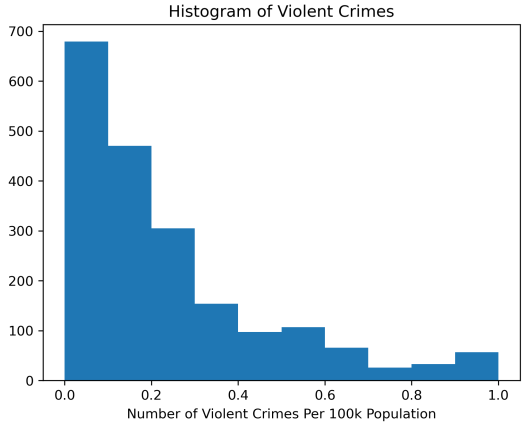
Dataset

- Data hosted on the University of California Irvine machine learning repository (Redmond 2009)
- Combines data from three datasets, the 1990 Census, 1995 FBI Uniform Crime Report (UCR), and the 1990 US Law Enforcement Management and Administrative Statistics Survey (LEMAS)
- Contain information about socio-economic indicators (e.g., median family income) and crime/law enforcement (e.g., per capita number of police officers)
- 1994 observations and 122 predictive features.

Problem Statement

- In this project, we want to predict violent crime rates from a variety of features using the "Communities and Crime" dataset
- Main challenges are processing the dataset, choosing predictors, and models

Target


Histogram of Violent Crimes
700
600
500
400
300
200
100
0
0.0
0.2
0.4
0.6
0.8
1.0
Number of Violent Crimes Per 100k Population

Plan

- Choose a method for dealing with missing data
- Filter features to choose the most relevant
- Pick the best model
  - Compare MLP with classical machine learning methods

Contribution

- This data set does not have any published papers relevant to it (Redmond 2009)
  - So any analysis with this particular data set is new
- There is a large literature on the use of machine learning in so-called "predictive policing" (Hardyns and Rummens 2018)
- Many predictive policing systems use a single method ML method, like gradient boosting
- We have tried many different methods and compared them

Missing Data

- Data are “missing not at random” (Rubin 1976)
- Impute data using *MICE* algorithm (Van Buuren 2018)
  - Implemented as part of `MultipeImputer`
- Check analysis as part of robustness check

Preprocessing

- Filter features based on correlation, drop variables with correlation greater than a cutoff value
- Algorithm for removing highly correlated features is taken from (Kuhn, Johnson, and others 2013, 26:47)

1. Calculate correlation matrix (`.corr()`)
2. Find two predictors with largest absolute pairwise correlation (A and B)
3. Find average correlation for all features with A and all features with B
4. If A greater than B, remove A. Else, B
5. Repeat until no correlations above threshold

Solver

- Rather than stochastic gradient descent, use `L-BFGS` solver (limited memory Broyden–Fletcher–Goldfarb–Shanno) (Aggarwal and others 2018, 148)
- Approximates Newton method:

$$\mathbf{W}(t+1) = \mathbf{W}(t) - \mathbf{H}^{-1}\nabla F(t)$$

- Replace $\mathbf{H}^{-1}$ with an approximation $\mathbf{G}(t)$

$$\mathbf{W}(t+1) = \mathbf{W}(t) - \alpha(t)\mathbf{G}(t)\nabla F(t)$$

Hyperparameter Tuning

- For every model type, there are many parameters to choose
  - For MLP, can choose number of hidden layers, activation function, learning rate, etc.
- We use 5-fold cross validation to choose hyperparameters for each model type
- Implemented with `GridSearchCV`, scored by `neg_mean_squared_error`

Results

Model Architecture

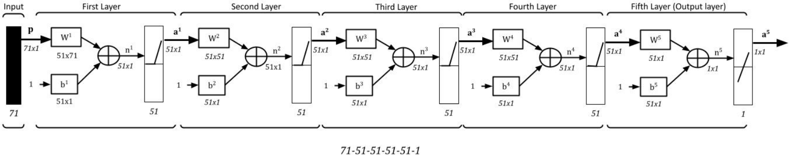


Figure 1: Chosen MLP

Results

Table 1: Model Performance for Non-Imputed Data

Model Type	Best 5-Fold MSE	Test MSE
SVR	0.0186	0.0203
DTR	0.0356	0.0453
RFR	0.0181	0.0187
MLP	0.0180	0.0130

Table 2: Model Performance for Imputed Data

Model Type	Best 5-Fold MSE	Test MSE
SVR	0.0185	0.0199
DTR	0.0352	0.0389
RFR	0.0178	0.0187
MLP	0.0181	0.0167

Predicted Values

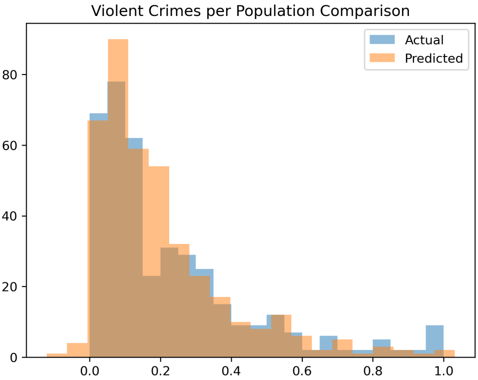
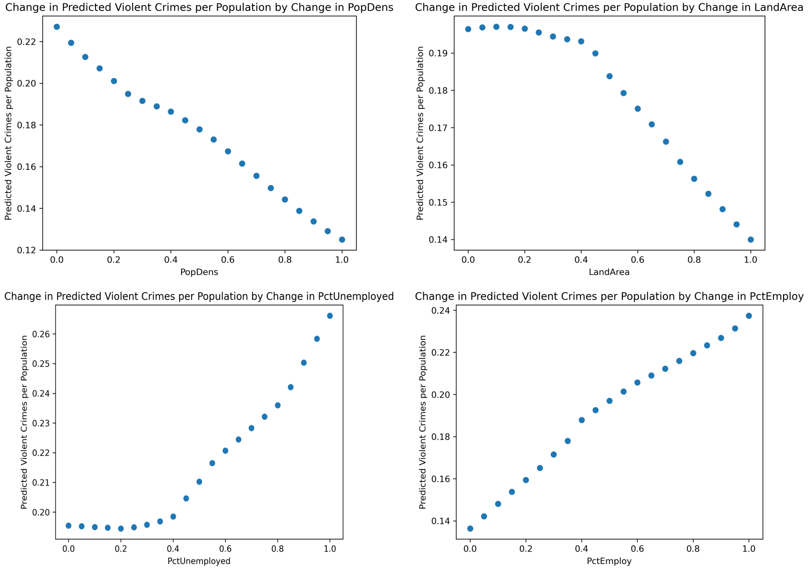


Figure 2: Violent Crime Per Population Comparison (Test Data)

Predicted Crime Rate Changes



Further Work

- Better data imputation algorithm
  - `IterativeImputer` has multiple estimations methods or `KNNImputer`
- More up to date data
  - Potentially incorporate time series data
- More comprehensive hyperparameter grid to search

References I

Aggarwal, Charu C, and others. 2018. “Neural Networks and Deep Learning.” *Springer* 10. Springer: 978–3.

Hardyns, Wim, and Anneleen Rummens. 2018. “Predictive Policing as a New Tool for Law Enforcement? Recent Developments and Challenges.” *European Journal on Criminal Policy and Research* 24 (3). Springer: 201–18.

Kuhn, Max, Kjell Johnson, and others. 2013. *Applied Predictive Modeling*. Vol. 26. Springer.

Redmond, Michael. 2009. “Communities and Crime.” https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime.

Rubin, Donald B. 1976. “Inference and Missing Data.” *Biometrika* 63 (3). Oxford University Press: 581–92.

Van Buuren, Stef. 2018. *Flexible Imputation of Missing Data*. CRC press.