

Proposal

Kushal Ismael and Matt Kosko

6/23/2021

In this project, we will use the “Communities and Crime” dataset from the UCI Machine Learning Repository to analyze violent crime rates (Redmond 2009). Specifically, we will attempt to predict the violent crime rate using the 128 attributes which are plausibly related to crime rates.

Because the outcome variable is a real number, this is a *regression* problem. Because there are missing data in some of the features, we will first preprocess the data by imputing missing values. Then, because there are only 1994 observations and 128 features, and it is likely that several of the features are correlated, we will remove highly correlated features (i.e., features with correlation coefficients greater than some threshold like 0.95). We will also scale and center the data to put all the features on the same scale.

We will make use of a multilayer perceptron (MLP) to predict the outcome variable. We plan to split the data into train and test sets and use k-fold cross validation (CV) on the training data; the CV accuracy metric will help us tune the number of hidden layers and number of neurons in each layer. In addition, we will also fit regular linear regression models, support vector machine (SVM) models, and decision trees, again using k-fold cross validation to tune hyperparameters. We will then present the accuracy of these methods using mean squared error (MSE) on the test data to compare it with the MLP.

All of the coding will be done in Python using the `sklearn` package. The textbook and course materials will be the background materials we use. A rough schedule is shown below:

Table 1: Rough Schedule

Date	Task
June 13	Select dataset
	Create GitHub repo
	Add proposal to GitHub
June 19	Finish preprocessing data for models
	Apply models to data
June 21	Finish final report
	Create presentation
June 23	Finalize project materials
	Present
	Submit project

References

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