Final Project

Kushal Ismael and Matt Kosko

6/23/2021

Contents

1 Introduction

In this report, we will predict violent crime rates using a variety of features plausibly associated with crime. The data comes from the "Communities and Crime" dataset, hosted on the University of California Irvine machine learning repository (Redmond 2009).

1.1 Description of the Data

The dataset, created in July 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) (Redmond 2009). The observations in the dataset are at the community level and contain information about socio-economic indicators (e.g., median family income) and crime/law enforcement (e.g., per capita number of police officers). In total, there are 1994 observations and 128 features in the data.

1.2 Machine Learning Approach

The main model that we will use is a *multi-layer perceptron* (MLP). The multilayer perceptron is a type of artificial neural network, with multiple hidden layers and neurons in each layer. MLPs are extremely useful and versatile in that they can approximate almost any function to an arbitrary degree of accuracy (Hagan, Demuth, and Beale 2014, 29). An example of a 3-layer network is taken from (Hagan, Demuth, and Beale 2014, 364) and shown in Figure 1.

We use the *backpropagation* algorithm to find the appropriate weights and biases for our neural network. The backpropagation is based on optimizing some performance index; inn our case, the performance index is

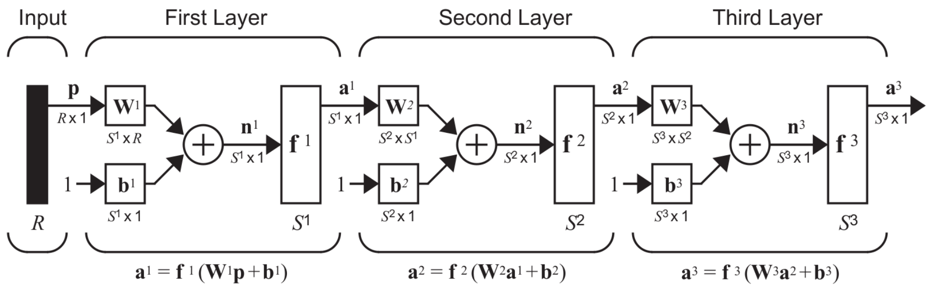

Figure 1: 3-Layer Perceptron

the mean squared error (MSE). Let X be a set of weights and biases in a multi-layer network. The MSE is defined as:

$$MSE(\mathbf{X}) = E(\mathbf{t} - \mathbf{a})^T(\mathbf{t} - \mathbf{a})$$

where E is the expectation operator over the set of input vectors, $\mathbf{t}$ is the target output, and $\mathbf{a}$ is the output of the netwowrk with the weights and biases $\mathbf{x}$ (Hagan, Demuth, and Beale 2014, 364). Because we do not know the probability distribution of the input vectors, we will approximate the MSE by replacing the expecation with the actual squared error.

$$M\hat{S}E(\mathbf{X})_k = (\mathbf{t}(k) - \mathbf{a}(k))^T(\mathbf{t}(k) - \mathbf{a}(k))$$

where k indicates the kth iteration in an optimization algorithm.

2 Experimental Setup

First, we need to preprocess the data to prepare the model for training. First, there are BLANK features that have missing values. Rather than drop incomplete observations, we will use a … impute function to impute missing values from observed data.

We first exclude variables that are highly correlated with one another; in this case, we drop variables that are greater than 0.9 in correlation. This exclues … features, leaving us with .. features. We also standaridzied the data, meaning that, for each observation i and feature X, we construct a new variable Z_i defined as follows:

$$Z_i = \frac{X_i - \bar{X}}{\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

where $\bar{X}$ is the sample mean. This standardizes the data to have mean 0 and standard deviation 1. By standardizing the data in this way, we … This is easily accomplished in Python using the `StandardScaler` function.

In order to accurately estimate the true error of the model, we will split the data into a training set and test set. The training set is use to actually fit the model, estimating the weights and biases. The test set is a hold out set, data that the model has not seen. The performance of the model on the test set will provide a good indication of how the model would perform on future data.

In addition to the MLP, we also train a support vector machine (SVM), linear regression model, and decision tree to compare the performance of classical machine learning methods with the MLP. The accuracy of the different approaches will be compared using the MSE on the test data; the model with the smallest MSE will be considered the "best" model.

3 Results

The result of our k-fold cross validation chose the MLP with .. layers. Table BLANK shows the MSE error for our MLP as well as the classical machine learning methods.

4 Summary and Conclusions

One potential future improvement is to inclue past information in a time series context. A model that predicts crime rates that have already occurred it of limited utility; ideally, we would want to know how the conditions today will affect violent crime rates next year or next month. One obvious application of this kind of model would be in crime prevention. If you could accurately predict violent crime rates for a particular community, policymaker would be able to

References

Hagan, Martin T, Howard B Demuth, and Mark Beale. 2014. *Neural Network Design.* https://hagan.okstate.edu/nnd.html.

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