Using Deep Learning to Forecast Political Instability

Christopher Boylan

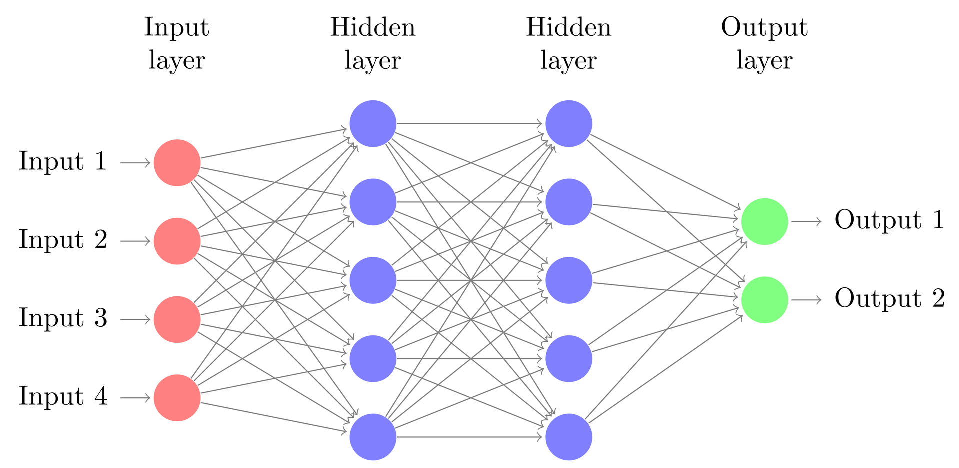
Pennsylvania State University

cib117@psu.edu

Introduction

Prediction is becoming increasingly recognized as an important task in political science, particularly in the domain of political instability. I build on existing political instability research by employing deep feedforward neural networks, which have led to major advances in areas such as image recognition, to forecast irregular leadership changes. I find that multilayer neural networks provide a small improvement in forecasting performance over other methods.

Multilayer Neural Networks

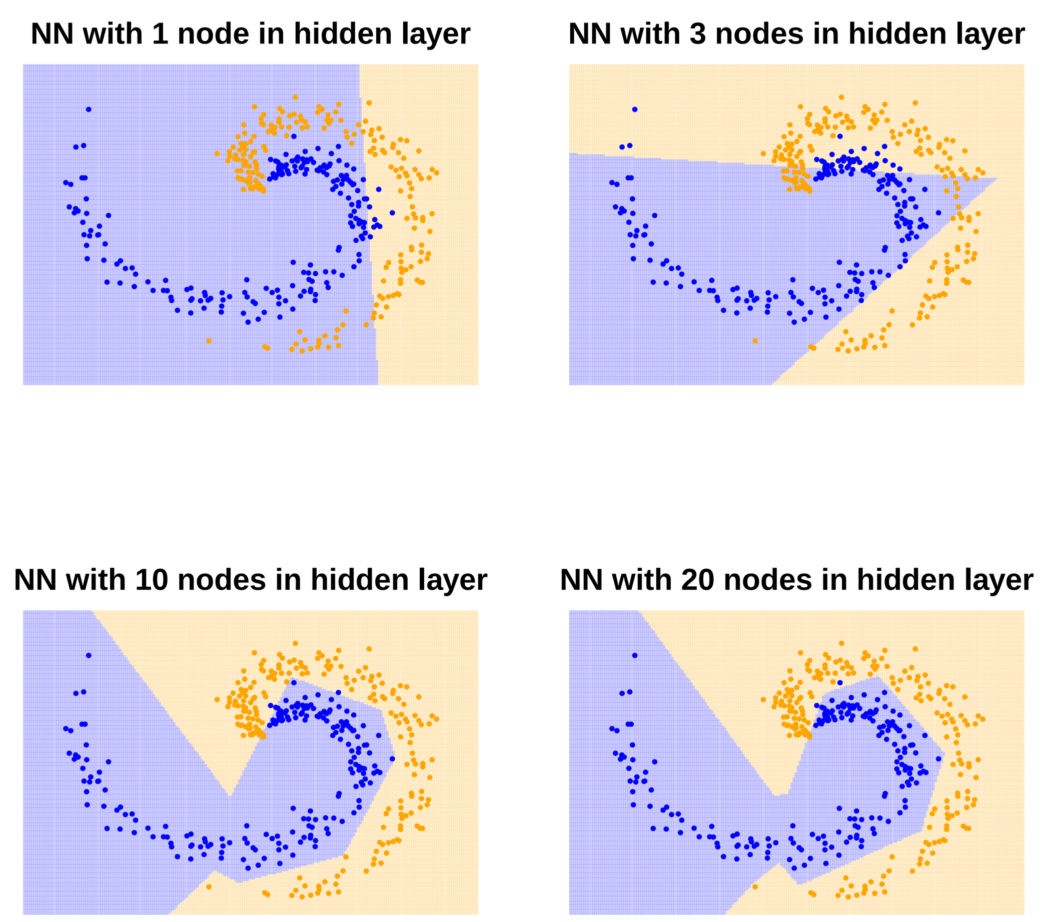


Architecture

A multilayer feedforward neural network produces predictions using forward propogation. For binary classification, the structure of a feedforward network with two hidden layers is as follows:

$$\begin{aligned} a_1 &= g(xW_1 + b_1) \\ a_2 &= g(a_1W_2 + b_2) \\ \hat{y} &= \sigma(a_2W_3 + b_3) \end{aligned}$$

where x is the input, g is the activation function, σ is the logistic function, and $(W_1, b_1, W_2, b_2, W_3, b_3)$ are the parameters learned from the training data. Commonly used activation functions are logistic, tanh, and the recitified linear unit (ReLU).

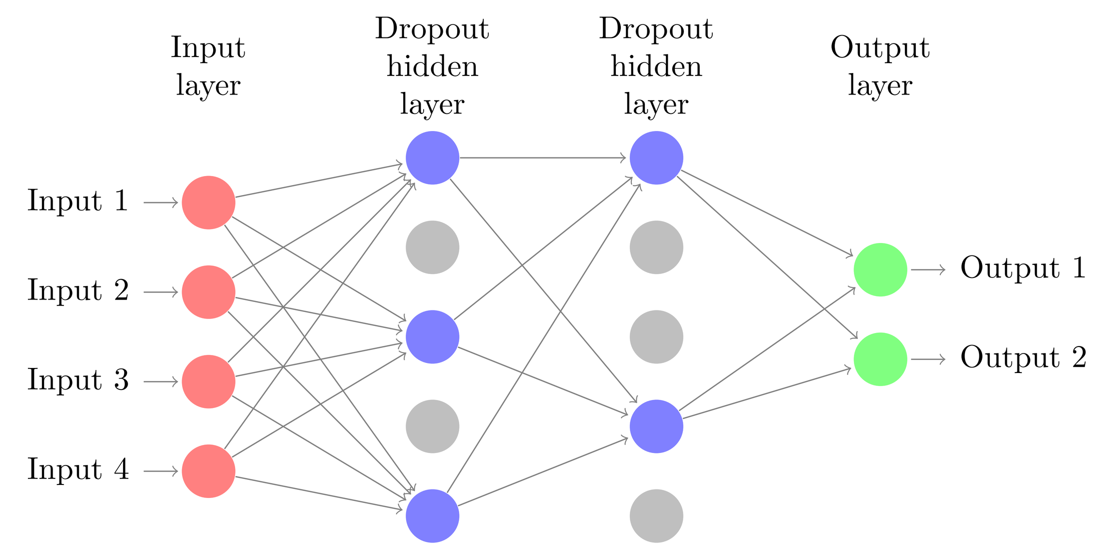


Training

Stochastic gradient descent and backproprogration are used to learn the parameters that minimize the error in the training data, as measured by a chosen loss function. For binary classification, log loss is commonly employed:

$$L(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

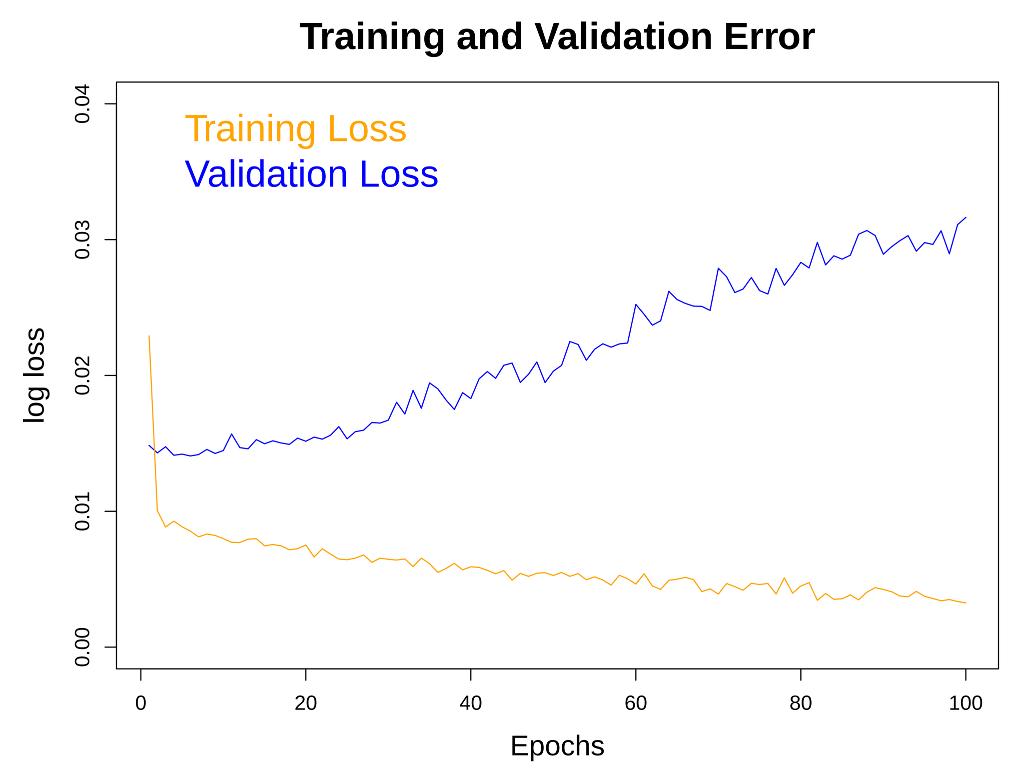
Networks with more hidden layers and nodes can represent more complicated functions. Techniques such as dropout, L_1/L_2 regularization, and input noise can be used to prevent networks from overfitting (Srivastava et al. 2014).



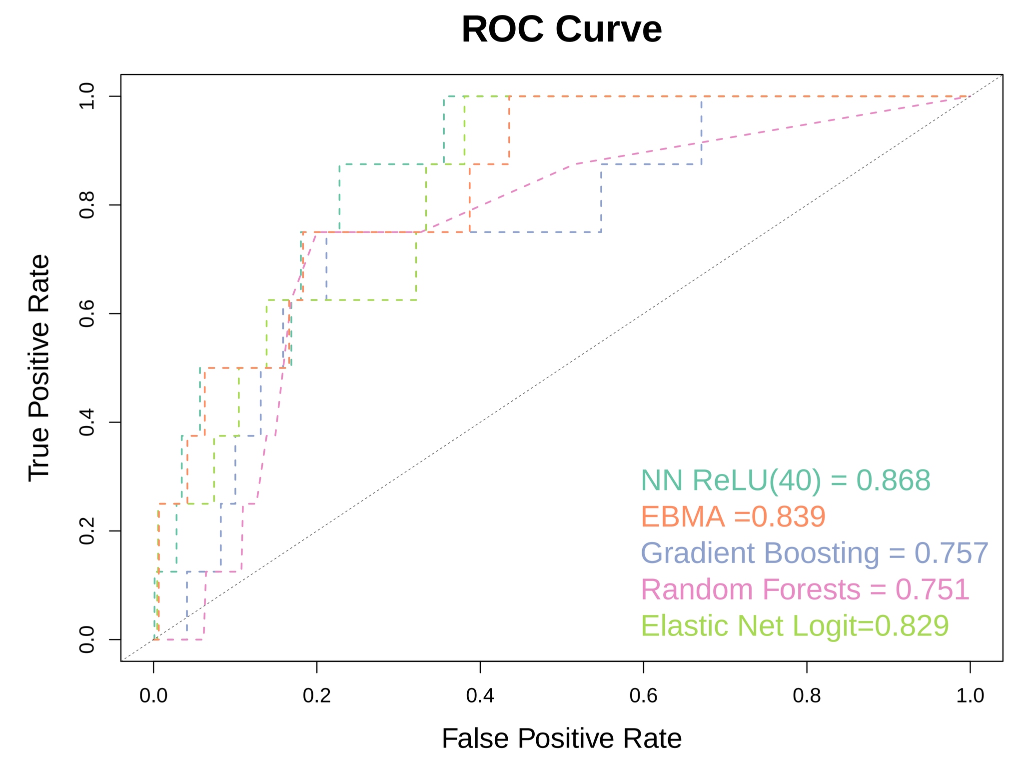
Data and Analysis

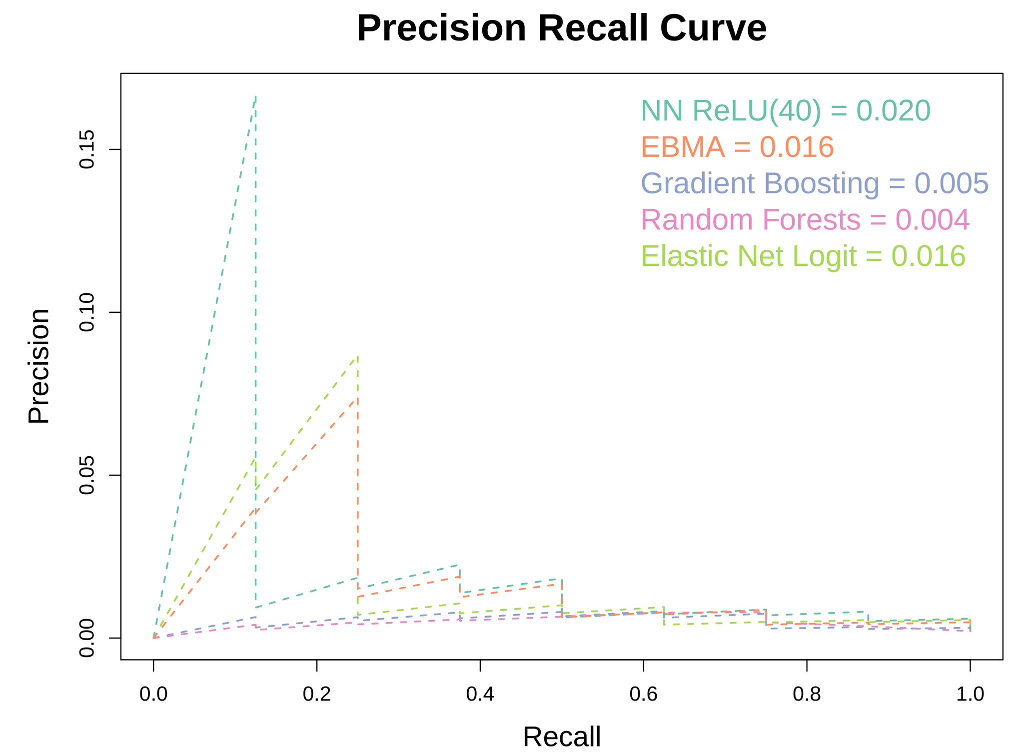
I predict irregular leader changes using data compiled by Beger et al. (2014). An irregular leadership change is the removal of a leader from office which contravenes a country's legal framework for leadership changes.

I search over different activation functions, the number of hidden layers, the numbers of nodes in each hidden layers, and different regularization techniques to find the optimal hyperparameters. I find that a ReLU with one hidden layer and 40 nodes using Dropout works best.



Performance Comparison





The neural network model performs best based on both the ROC curve and the Precision Recall curve.

Future Work

Explore the utility of recurrent neural networks, which work well with sequential data and can capture temporal dependencies.

References

- Beger, A., Dorff, C. L., & Ward, M. D. 2014. "Ensemble forecasting of irregular leadership change." *Research & Politics* 1(3).
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." *Journal of Machine Learning Research* 15(Jun):1929-1958.