Executive Summary: Neural Network-Based Fraud Detection

Di Kang, Fang Li

Overview

Credit card fraud is a persistent problem in the world of finance. Detecting fraud is essential to protect cardholders and financial institutions. A common issue in fraud detection datasets is that fraudulent transactions are significantly outnumbered by legitimate ones. We address the challenge of data imbalance typical in such datasets and configure a neural network model to detect credit card transaction fraud. Finally we evaluate the metrics and computational efficiency of two approaches.

Methods to Address Imbalance

Two techniques were employed:

1. **Oversampling with SMOTE**: SMOTE, or Synthetic Minority Over-sampling Technique, is a statistical technique designed to address class imbalance in machine learning datasets. It works by generating synthetic samples from the minority class to balance the dataset without losing valuable information. The method involves selecting instances of the minority class and then introducing synthetic examples along the line segments joining any/all of the k-nearest neighbors of these instances.
2. **Random Downsampling**: Random downsampling is a technique used to address class imbalance in machine learning datasets, particularly when one class significantly outnumbers another. This method involves reducing the size of the majority class to match the minority class's size, thereby creating a more balanced dataset.

Classifier

The classifier is configured using PyTorch with one hidden layer of 16 neurons and ReLU activation function.

Model Evaluation

We evaluate performance based on accuracy to ensure the highest likelihood of correctly identifying both fraudulent and legitimate transactions. Additionally, computational efficiency is crucial in business contexts, making time an important metric. Below, we list the accuracy metrics and computational times for both data balancing approaches:

1. **Oversampling with SMOTE**: Accuracy is 99.8%, time spent is 9min 36s.
2. **Random Downsampling**: Accuracy is 96.7% , time spent is 3s.

Conclusion

Both SMOTE and downsampling effectively balance class distribution, enhancing the neural network's performance in fraud detection. Oversampling shows a slight advantage in handling imbalanced datasets at a cost of significantly longer computational time. The choice of method should consider the trade-off between accuracy and computational efficiency and adjust to specific business demands.