

Cyber Data Analytics

Assignment 1

Ahmet Güdek Tabe Etta

May 13, 2018

Introduction

To be able to safely and trustingly perform online payments, it is important to be able to identify fraudulent transactions. But identifying fraudulent transactions in a sea of valid ones is a difficult task, mainly due to two reasons. First, all transactions look alike with little to distinguish fraudulent transactions from valid ones. And second, the number of fraudulent data points is extremely small in comparison to the entire set of transaction data. This asymmetry in class probabilities makes it difficult to create robust and reliable classifiers, leading to high false negative rates. In this report we will explore data sampling and manipulation methods to improve the data imbalance and create classifiers building upon this.

Finding patterns

Balancing data

Building classifiers