Stable Fraud Detection via Batch Balancing

Ruohan Zhan       Advisor: Susan Athey
Collaborators: Nitin Sharma, Vidyut Naware

Introduction

- Where: a digital two-sided marketplace of buyers and sellers;
- What: decide whether or not to decline a transaction;
- How: fraud detection algorithm with seller-buyer information as input and decision as output.

Stable Learning and Covariate Balancing

Research Problem --- Generalizability
With data from one environment, learn a model that can be generalized to other environments.

Stable Learning --- Covariate balancing
Reweight samples such that other features are balanced with respect to each current feature.

Continuous features (see [2] for details)
Adjust sample weights to decorrelate features.

\[
W = \arg \min_W \sum_{j=1}^{\infty} \sum_{l=1}^{P} \sum_{i=1}^{n} [E[(X_{ij}^l)^T \Sigma_W X_{ij}^l - J] - E[(X_{ij}^l)^T W E[X_{ij}^l]]^2]
\]

Computationally, adopt the first moment approximation and choose \(a=b=1\).

Setup

Current Fraud Detection System
- Buyer models: generate buyer risk using buyer data including buyer funding instrument
- Seller models: generate seller risk using merchant data including the associated seller account
- Rule-based model: decide total risk given buyer risk, seller risk, the current channel (that processes transaction).

Data

PayPal 2018 transaction data.
- Buyer data: \((X, Y) = (buyer features, buyer fraud tag)\)
- Feature size: 625.

Stratification of Environments

- Training data: transactions before Thanksgiving;
- Test data: transactions after Thanksgiving.

Experiments

Evaluation metric: Logistic loss

\[-Y \log(p) - (1 - Y) \log(1 - p)\]

Compared models:
- Vanilla model: logistic model
- Balanced model (ours): logistic model + batch balancing

Acknowledgements

We are grateful to generous support from PayPal Innovation Fellowship.

References