CS236r Project Proposal: Scoring Systems for Predicting Credit Risk

Alex Lin, Melissa Yu

1 Introduction

In this proposal, we apply the Supersparse Linear Integer Model (SLIM) scoring system to make interpretable predictions of loan default risk. We motivate this research by describing how the proposed mechanism can...

2 Background

Consumer spending is among the most important determinants of macroeconomic conditions and systemic risk, accounting for around 70% of the U.S. GDP between 2001 and 2010 [2]. As the credit industry has rapidly expanded over the last few decades, the risk associated with consumer lending has multiplied as well. Today, it is more important than ever that financial institutions leverage accurate, interpretable prediction models and algorithms to make the many lending decisions they must process on a daily basis, instead of relying on potentially biased, inaccurate, and un-scalable human discretion. These predictive algorithms leverage characteristics pulled from the consumer credit files collected by credit bureau agencies to discern the riskiness of a loan.

2.1 Credit scoring models

Credit scoring models are quantitative models widely used by financial institutions to determine the probability of delinquency or default for loan applicants [3, 6]. These models have as their goal accurately classifying loan applicants into one of two groups: "good" customers, who are likely to repay their debt, and "bad" customers, who are denied credit on the basis that they are likely to default.

In order for lenders to leverage scoring algorithms to make smart lending decisions, the model they use must satisfy several requirements [1]:

1. **Predictive power**: The model should make accurate predictions of customer type – good or bad.

2. **Fairness**: The model must comply with the Equal Credit Opportunity Act (ECOA) and Fair Credit Reporting Act (FCRA). It cannot discriminate on the basis of race, age, education, employment history, gender, marital status, or wealth.

3. **Broad coverage**: The model is adaptable enough to apply across large datasets of people.

4. **Data transparency**: The data used in the model is verifiable and correctable by consumers.

5. **Interpretability (consumer-focused)**: The model's classification process can be easily explained to and understood by consumers, and the most important predictive factors can be intuitively identified to aid loan seekers in improving their profile.

2.2 Scoring Systems

Scoring systems are linear classification models that only require users to add, subtract and multiply a few small numbers in order to make a prediction... [4, 5]

3 Research Proposal

The previously described attributes of the credit risk prediction problem motivate us to suggest a scoring system-based solution.

References

[1] FICO. 5 building blocks of a great scoring system, 2015.

[2] A. E. Khandani, A. J. Kim, and A. W. Lo. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11):2767–2787, 2010.

[3] C.-S. Ong, J.-J. Huang, and G.-H. Tzeng. Building credit scoring models using genetic programming. *Expert Systems with Applications*, 29(1):41–47, 2005.

[4] B. Ustun and C. Rudin. Supersparse linear integer models for optimized medical scoring systems. *Machine Learning*, 102(3):349–391, 2016.

[5] B. Ustun and C. Rudin. Optimized risk scores. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1125–1134. ACM, 2017.

[6] D. West. Neural network credit scoring models. *Computers & Operations Research*, 27(11-12):1131–1152, 2000.