

# CS236r Project Update

## Scoring Systems for Predicting Credit Risk

Alex Lin

Melissa Yu

## 1 Introduction

In our final project, we apply the Supersparse Linear Integer Model (SLIM) scoring system to make interpretable predictions of loan default risk. We begin by examining a simple neural network baseline model, which achieves reasonably high accuracies on our dataset, but falls short in terms of both interpretability and fairness, two important requirements for any credit scoring model. We then describe TODO

## 2 Baseline

### 2.1 Data

We use the GMC dataset, which was provided by a financial institution for the “Give me some credit” Kaggle competition (<https://www.kaggle.com/c/GiveMeSomeCredit>). This comprises historical data on 250,000 borrowers, each quantified along 10 dimensions, including age and monthly income. Notably, this dataset does not include information on the borrowers’ race or gender. We are considering using the German credit (GC) dataset from the UCI Library (Lichman, 2013) for future work, since it does include these features and may allow us to quantify fairness more accurately.

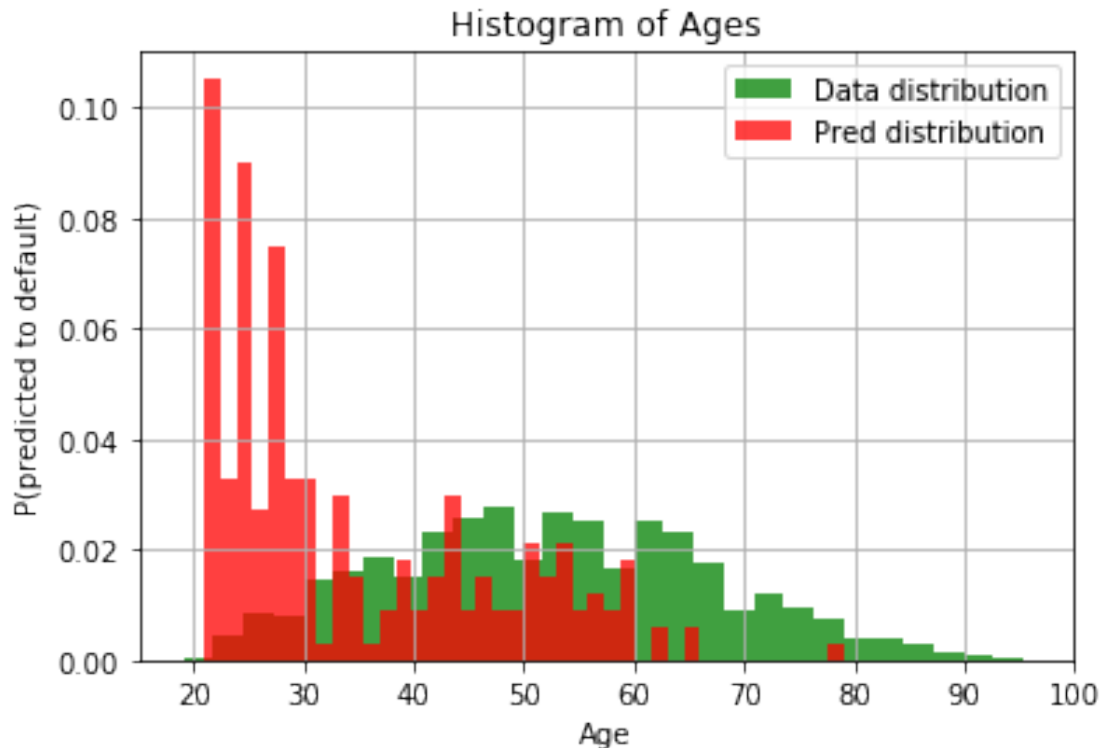
### 2.2 Neural Network Model

Benchmark studies of classification algorithms for credit scoring discuss these models in the framework of both individual (e.g., CART, Linear support vector machine, Logistic Regression, Multilayer perceptron artificial neural network) and ensemble methods (e.g, Bagged decision trees, Simple average ensemble). To gain a preliminary gauge of the performance of a simple model on our dataset, we explore the use of Multilayer perceptron artificial neural networks.

We used PyTorch to implement a single-layer, linear, feedforward neural net with hidden layer size 30 and ReLU activations. The model was trained using the Adam optimizer with learning rate 0.0001 over 30 epochs. Our vanilla model achieved an AUC score of 0.801831 on the Kaggle test set. However, while this model has reasonable predictive power, is not very interpretable to consumers: The model’s classification process cannot be easily explained to or understood by consumers, and neither can the most important predictive factors can be intuitively identified to aid loan seekers in improving their profile.

This baseline also has low fairness: As discussed in our project proposal, the model must comply with the Equal Credit Opportunity Act (ECOA) and Fair Credit Reporting Act (FCRA). Thus, it cannot discriminate on the basis of race, age, education, employment history, gender, marital status, or wealth. We perform a rudimentary evaluation of the group fairness of our classification

model by visualizing the population age distribution and the age distribution of borrowers predicted to default on the same histogram. The two distributions diverge quite a bit, with the percentage of young people predicted to default being much higher than the percentage of young people in the population as a whole. This indicates our algorithm may not be fair when it comes to age.



### 3 Research Proposal

Our next steps are to implement the RiskSLIM algorithm to create scoring rules for improving fairness and interpretability in the credit scoring problem, while hopefully not losing too much accuracy. We will compare our results against those of the baseline. We chose to focus on RiskSLIM instead of SLIM, because we would prefer to have our scoring rules to output scores that correlate with the risk of defaulting on a loan. If time allows, we may additionally implement SLIM to see if there is a tradeoff between encoding risk and having a higher AUC.

To use RiskSLIM, we will formulate a Mixed Integer Nonlinear Program (MINLP) and using the lattice plane cutting algorithm (LPCA) to optimize solutions. Ustun and Rudin indicate that there are five constraint types that are available to encode in the MINLP. To enforce fairness, we will use the following constraints that have to do with the 'Probability' type: Let there be a sensitive attribute  $X$  (e.g. race) that can take on  $k$  possible values  $x_1, \dots, x_k$  for all the individuals in a populations. Let the proportion of individuals in the population with  $X = x_i$  be  $p_i$  for  $i = 1, \dots, k$ . Then, we will make sure that out of the people who are classified to successfully return their loan, around  $p_i$  of them have  $X = x_i$ . That is, our scoring rule will predict  $Pr(\text{return loan} = \text{True}) \approx p_i$  when  $X = x_i$ . Such a constraint will be easy to build into the algorithm using the existing framework provided by Ustun and Rudin.