HU Berlin

kreditech

Shallow Self-Learning for Reject Inference in Credit Scoring

Nikita Kozodoi, Panagiotis Katsas,
Stefan Lessmann, Luis Moreira-Matias, Konstantinos Papakonstantinou

nikita.kozodoi@hu-berlin.de

Credit scoring models are supervised binary classifiers that predict the probability of default. Scoring models are trained on data from previously granted credit applications with the observed repayment behavior. This creates sample bias: a model is trained on the accepted cases only.

Figure 1a demonstrates the acceptance cycle. **Figure 1b** illustrates the bias after running the acceptance cycle for 300 iterations with synthetic data. Since labels of rejects are not missing at random, the bias degrades model performance when applying it to new applications as shown in **Figure 1c**.

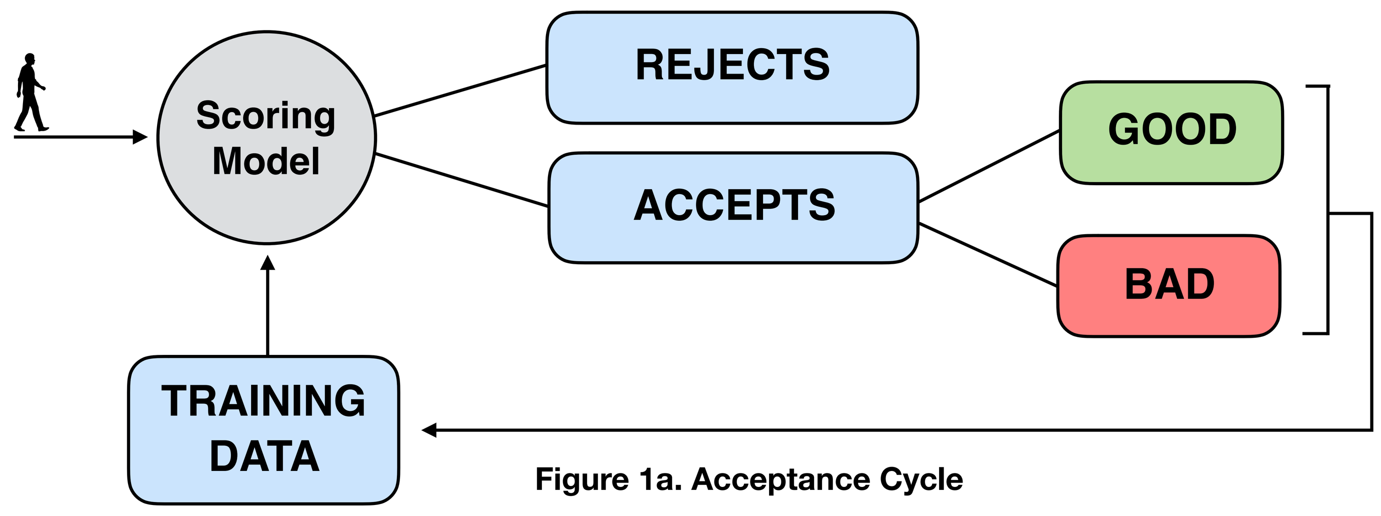

Figure 1a. Acceptance Cycle

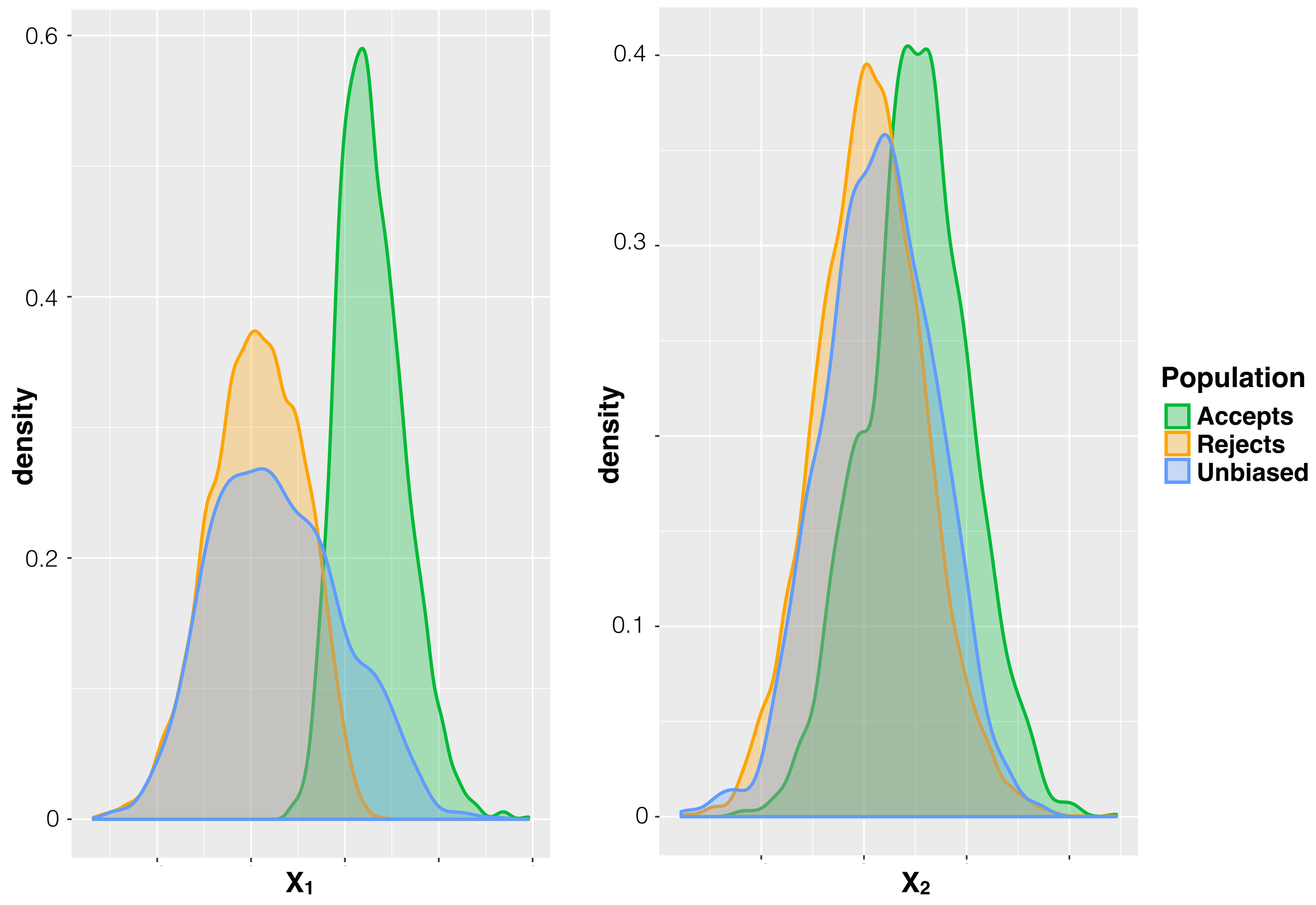

Figure 1b. Sample Bias

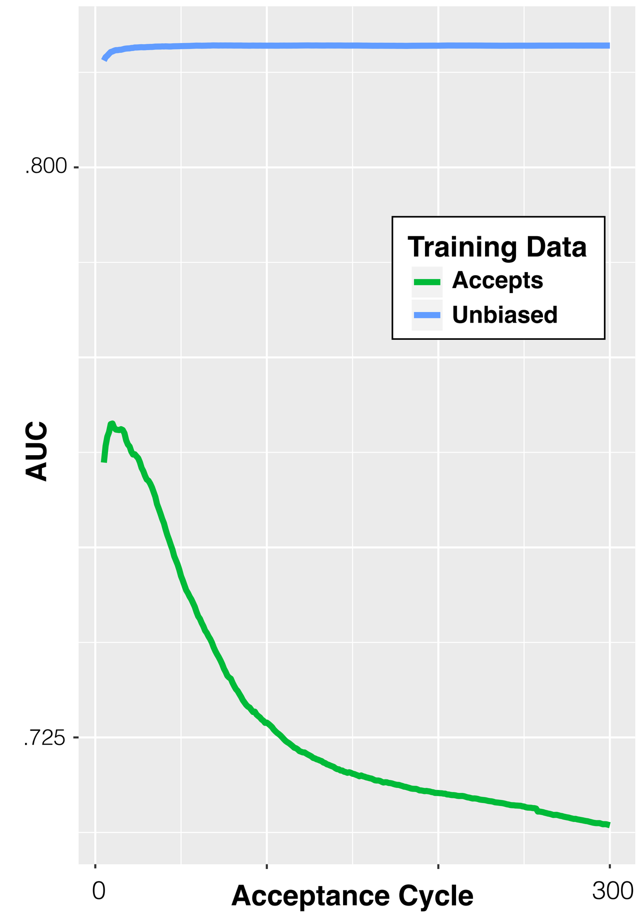

Figure 1c. Loss due to Bias

We make two contributions: **1) Shallow Self-Learning** to label rejected cases and mitigate sample bias; **2) Kickout Metric** to improve model selection when true labels of rejects are unknown.

Shallow Self-Learning

Stage I filtering

ACCEPTS → train → F → score → REJECTS → FILTERED REJECTS

isolation forest

Goal: removing rejects whose distribution is most different from the accepts

Stage II labeling

ACCEPTS → train → L → score → FILTERED REJECTS

L1 classifier

label & append the most confident cases

- **using weak learner** (L1) to get well-calibrated probabilities
- **imbalance parameter** θ to account for higher BAD rate among rejects
- **early stopping** to reduce the risk of error propagation

Goal: labeling rejected cases that are classified with high confidence

Stage III scoring

ACCEPTS + LABELED REJECTS → TRAINING DATA → train → S → score → NEW APPLICANTS

XGB classifier

Goal: training a scoring model on the extended data with smaller bias

Figure 2a depicts performance gains from using shallow self-learning on synthetic data.

Compared to the scoring model trained on the unbiased sample, accepts-based model suffers from sample bias resulting in a 0.1 drop in AUC. Reject inference with shallow self-learning increases AUC by 0.05, recovering about 50% of the loss due to bias.

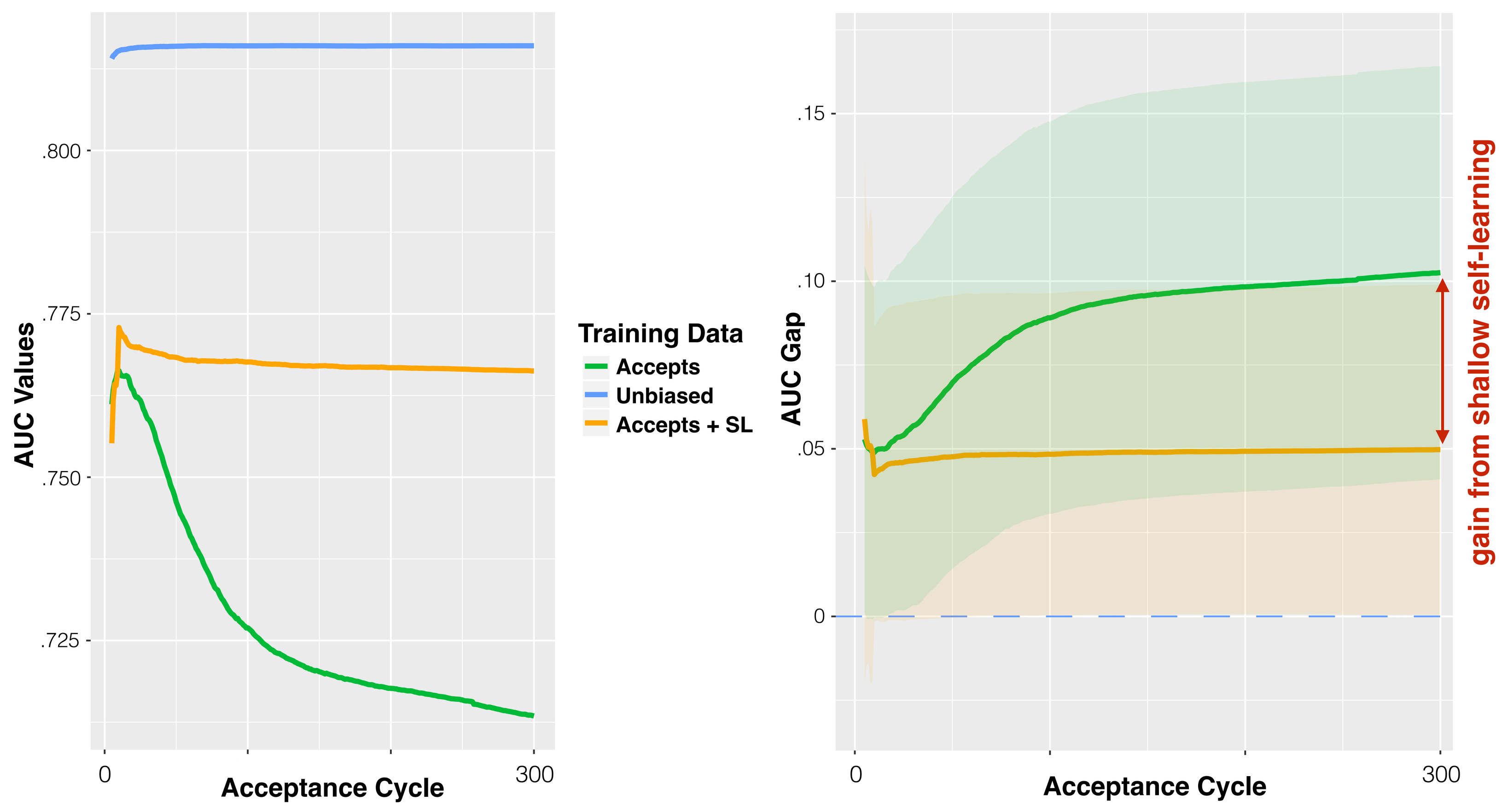

Figure 2a. Performance Gains: Reject Inference with Shallow SL

Kickout Metric

Intuition:

compare applicants accepted by two models: **before** and **after** reject inference

Calculation:

- train initial model on **TRAIN** and score cases in **VALID**
- split rejected cases into two subsets: **R1** and **R2**
- use reject inference to label **R1**
- append labeled **R1** to **TRAIN**; append **R2** without labels to **VALID**
- train new model on **(TRAIN + R1)** and score cases in **(VALID + R2)**
- count **GOODs** and **BADs** that are "kicked out" - rejected after reject inference

Kicked-out cases are replaced with **rejects with unknown labels:**

- kicking out a **BAD** case has a positive expected value
- kicking out a **GOOD** case has a negative expected value
- hence, reject inference should aim at kicking out more **BADs** and less **GOODs**

$$kickout = \frac{\frac{K_B}{p(B)} - \frac{K_G}{1-p(B)}}{\frac{S_B}{p(B)}}$$

- K_B, K_G - kicked-out **BADs** and **GOODs**
- **p(B)** - probability of selecting **BAD** example
- S_B - number of selected **BAD** examples

Figure 2b shows performance gains from using the kickout metric for model selection.

Selecting a scoring model based on its AUC on the accepted cases leads to suboptimal performance on the unbiased holdout sample (AUC of 0.79). Selecting a model based on kickout improves AUC on the unbiased sample by 0.01.

Overall, ranking models based on their AUC on the unbiased sample better correlates with the kickout metric (r = 0.30) than with the AUC on the accepted cases (r = 0.12).

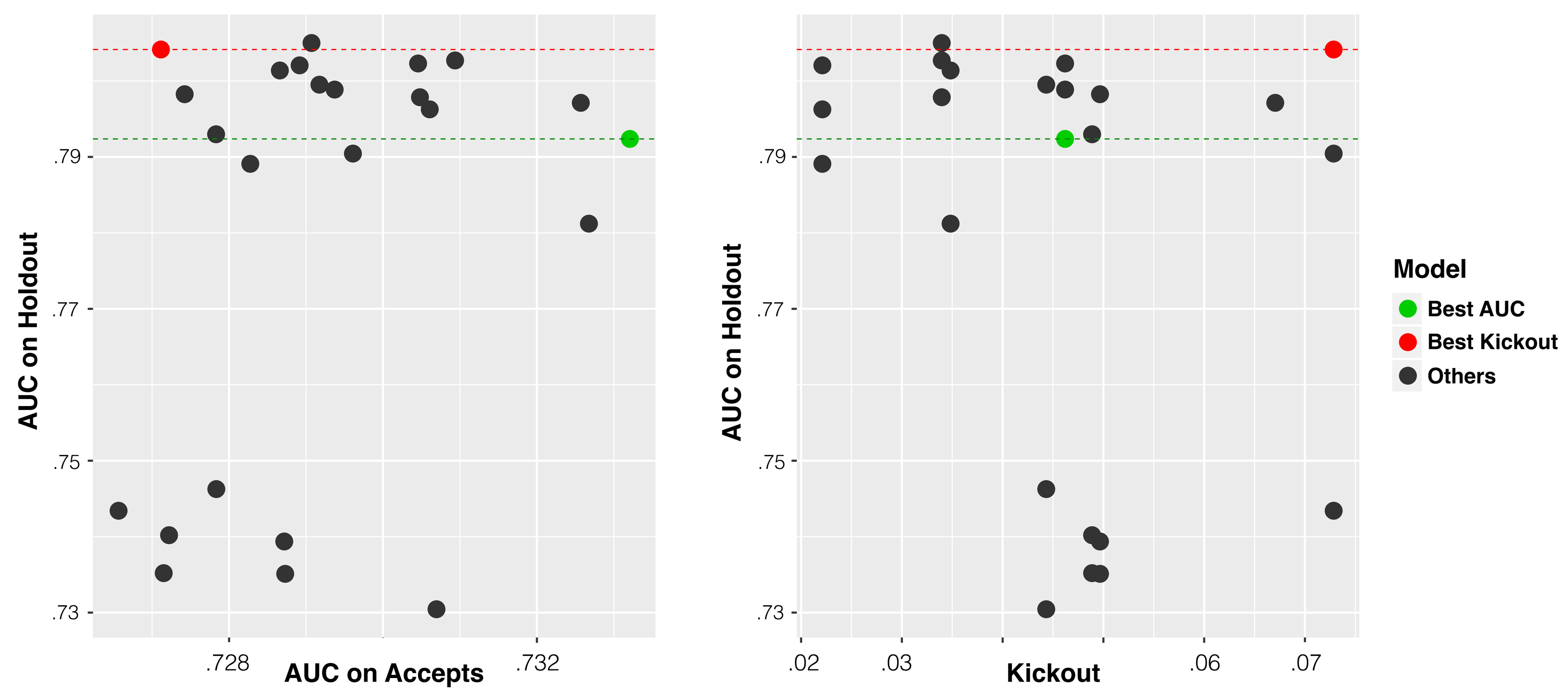

Figure 2b. Performance Gains: Model Selection with Kickout Metric

Real-World Data

Data description:

- consumer micro-loans provided by kreditech
- contains data on **accepted** and **rejected** applicants
- features **unbiased sample:** randomly accepted loans

Characteristic	Accepts	Rejects	Unbiased
Number of cases	39,579	18,047	1,967
Number of features	2,410	2,410	2,410
BAD rate	0.39	-	0.66

Performance Gains

Method	Mean AUC
Ignore rejects	**0.8007**
Label all rejects as BAD	0.6797
Bureau score based inference	0.7911
Hard cutoff augmentation	0.7994
Parceling	0.8041
Shallow Self-Learning + Kickout	**0.8072**

Monetary Gains

Assumptions:

- acceptance rate = **30%** cases with the lowest score
- average loan size = **$17,100**, interest rate = **10.36%**
- average loss given default = **25%**

Measure	Ignore rejects	SSL + Kickout	Difference
Precision at 30%	0.7936	0.7996	**0.006**
BAD loans per 10,000 clients	2,064	2,004	**60**
Revenue loss	$9,737,725	$9,454,652	**$283,073**