Learning the Truth From Only One Side of the Story

Heinrich Jiang
Google Research

Qijia Jiang
Stanford University

Aldo Pacchiano
UC Berkeley

Abstract

Learning under one-sided feedback (i.e., where we only observe the labels for examples we predicted positively on) is a fundamental problem in machine learning – applications include lending and recommendation systems. Despite this, there has been surprisingly little progress made in ways to mitigate the effects of the sampling bias that arises. We focus on generalized linear models and show that without adjusting for this sampling bias, the model may converge suboptimally or even fail to converge to the optimal solution. We propose an adaptive approach that comes with theoretical guarantees and show that it outperforms several existing methods empirically. Our method leverages variance estimation techniques to efficiently learn under uncertainty, offering a more principled alternative compared to existing approaches.

1 INTRODUCTION

Machine learning is deployed in a wide range of critical scenarios where the feedback is one-sided, including bank lending (Tsai and Chen, 2010; Kou et al., 2014; Tiwari, 2018), criminal recidivism prediction (Tollenaar and Van der Heijden, 2013; Wang et al., 2010; Berk, 2017), credit card fraud (Chan et al., 1999; Srivastava et al., 2008), spam detection (Jindal and Liu, 2007; Sculley, 2007), self-driving motion planning (Paden et al., 2016; Lee et al., 2014), and recommendation systems (Pazzani and Billsus, 2007; Covington et al., 2016; He et al., 2014). These applications can often times be modeled as one-sided feedback in that the true labels are only observed for the positively predicted examples and the learner is simultaneously making predictions and actively learning a better model. For example, in bank loans, the learner only observes whether the loan was repaid if it was approved. In criminal recidivism prediction, the decision maker only observes any re-offences for inmates who were released.

Incidentally, this problem can be viewed as a variation on the classical active learning problem in the streaming setting (Bordes et al., 2005; Chu et al., 2011; Lu et al., 2016), where unlabeled examples arrive in a sequential manner and the learner must decide whether to query for its label for a fixed cost in order to build a better model. Here, the goal is similar, with the difference that the labels being queried are the ones with positive predictions. There is a tension between making the correct predictions and choosing the right examples to query for labels – a cost is associated with querying negative examples on one hand, and on the other we seek to learn a better model for improved future performance. As we show later, the key difficulty of this problem lies in understanding this trade-off and exactly pinpointing when to make a positive prediction in the face of uncertainty. In the case of bank lending, for example, assessing the confidence for the prediction on applicant's chance of repayment is of great importance. Decision needs to be made on balancing the risk of default if granted the loan, which comes with a high cost, and the benefit of the additionally gathered data our model can learn from.

One often overlooked aspect is that the samples used to train the model, which prescribes which data points we should act upon next, are inherently biased by its own past predictions. In practical applications, there is a common belief that the main issue caused by such one-sided sampling is label imbalance (He et al., 2014), as the number of positive examples will be expected to be much higher than overall for the population. Indeed, this biasing of the labels leading to label imbalance can be a challenge, motivating much of the vast literature on label imbalance. However, the challenges go beyond label imbalance. We show that without accounting for such potential myopia caused by biased sampling, it is possible that we under-sample in regions where the model makes false negative predictions, and

Proceedings of the 24[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, San Diego, California, USA. PMLR: Volume 130.

even with continual feedback, the model never ends up correcting itself. In the bank loan example, such under-sampling may systematically put minority group in a disadvantaged position, as reflected by the error being disproportionally attributed across groups, if we content ourselves with a point estimator that doesn't take into consideration the error bar that's associated.

In this paper, we take a data-driven approach to guide intervention efforts on correcting for the bias – uncertainty quantification tools are used for striking the balance between short-term desideratum (i.e., low error rate on current sample) and long-term welfare (i.e., information collection for designing optimal policy). More concretely, we focus on generalized linear models, borrowing assumptions from a popular framework of Filippi et al. (2010). Our contributions can be summarized as follows.

- In Section 3, we propose an objective, *one-sided loss*, to capture the one-sided learner's goals for the model under consideration.
- In Section 4, we show that without leveraging active learning where the model is continuously updated upon seeing new labeled examples, a model may need to be trained on a sub-optimal amount of data to achieve a desirable performance on the objective.
- In Section 5, we show that the greedy active approach (i.e., updating the model only on examples with positive predictions at each timestep) in general will not exhibit asymptotically vanishing loss.
- In Section 6, we give a strategy that adaptively adjusts the model decision by incorporating the uncertainty of the prediction and show an improved rate of convergence on the objective.
- In Section 7, we explore the option of using iterative methods for learning the optimal model parameters while maintaining small misclassification rate under this partial feedback setting. The proposed SGD variant of the adaptive method complements our main results which focus on models fully optimized on all of the labeled examples observed so far.
- In Section 8, we provide an extensive experimental analysis on linear and logistic regression on various benchmark datasets showing that our method outperforms a number of baselines widely used in practice.

To the best of our knowledge, we give the most detailed analysis in the ways in which passive or greedy learners are sub-optimal in the one-sided feedback setting and we present a practical algorithm that comes with rigorous theoretical guarantees which outperforms existing methods empirically.

2 RELATED WORK

Despite the importance and ubiquity of this active learning problem with one-sided feedback, there has been surprisingly little work done in studying the effects of such biased sampling and how to mitigate it. Learning with partial feedback was first studied by Helmbold et al. (2000) under the name "apple tasting" who suggest to transform any learning procedure into an apple tasting one by randomly flipping some of the negative predictions into positive ones with probability decaying over time. They give upper and lower bounds on the number of mistakes made by the procedure in this setting. Sculley (2007) studies the one-sided feedback setting for the application of email spam filtering and show that the approach of Helmbold et al. (2000) was less effective than a simple greedy strategy. Cesa-Bianchi et al. (2006a) propose an active learning method for linear models to query an example's label randomly with probability based on the model's prediction score for that example. Bechavod et al. (2019) consider the problem of one-sided learning in the group-based fairness context with the goal of satisfying equal opportunity (Hardt et al., 2016) at every round. They consider convex combinations over a finite set of classifiers and arrive at a solution which is a randomized mixture of at most two of these classifiers.

Cesa-Bianchi et al. (2006b) studies a setting which generalizes the one-sided feedback, called *partial monitoring*, through considering repeated two-player games in which the player receives a feedback generated by the combined choice of the player and the environment. They propose a randomized solution. Antos et al. (2013) provides a classification of such two-player games in terms of the regret rates attained and Bartók and Szepesvári (2012) study a variant of the problem with side information. Our approach does not rely on randomization that is typically required to solve such two-player games. There has also been work studying the effects of distributional shift caused by biased sampling (Perdomo et al., 2020). Ensign et al. (2017) studies the one-sided feedback setting through the problems of predictive policing and recidivism prediction. They show a reduction to the partial monitoring setting and provide corresponding regret guarantees.

Filippi et al. (2010) propose a generalized linear model framework for the multi-armed bandit problem, where for arm a, the reward is of the form $\mu(a^\top \beta^*) + \epsilon$ where β^* is unknown to the learner, ϵ is additive noise, and $\mu(\cdot)$ is a link function. Our work borrows ideas from

this framework as well as proof techniques. Their notion of regret is based on the difference between the expected reward of the chosen arm and that of an optimal arm. One of our core contributions is showing that, surprisingly, modifications to the GLM-UCB algorithm leads to a procedure that minimizes a very different objective under a disparate feedback model.

3 PROBLEM SETUP

We assume that data pairs $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ are streaming in and the learner interacts with the data in sequential rounds: at time step t we are presented with a batch of N samples $(x_1^t, , \cdots, x_N^t)$, and for the data points we decide to observe, we are further shown the corresponding labels y_i^t, while no feedback is provided for the unobserved ones. We make the following assumptions.

Assumption 1 (GLM Model)**.** *There exists $\beta^* \in \mathbb{R}^d$ (unknown to the learner) and link function $\mu : \mathbb{R} \mapsto \mathbb{R}$ (known to the learner) such that y is drawn according to an additive noise model $y = \mu(x^\top \beta^*) + \epsilon$. The link function $\mu(\cdot)$ is continuously differentiable and strictly monotonically increasing, with Lipschitz constant L, i.e., $0 < \mu'(z) \leq L \; \forall z \in \mathbb{R}$. Moreover, $\mu(0) \leq \gamma$.*

Assumption 2 (Bounded Covariate)**.** *There exists some $B > 0$ such that $\|x_i^t\|_2 \leq B$ for all $i \in [N], t \geq 0$.*

Assumption 3 (Parameter Diameter)**.** *The unknown parameter β^* satisfies $\|\beta^*\|_2 \leq M$.*

Assumption 4 (Subgaussian Noise)**.** *The noise residuals $\epsilon_i^t := y_i^t - \mu(x_i^{t\top} \beta^*)$ are mutually independent, conditionally zero-mean and conditionally ϕ-subgaussian. That is, $\forall i \in [N], t \geq 1, \tau \in \mathbb{R}$,*

$$\mathbb{E}[\epsilon_i^t | \{x_i^t\}_i, \{\epsilon_i^{t-1}\}_i, \cdots, \{x_i^0\}_i, \{\epsilon_i^0\}_i] = 0,$$

$$\mathbb{E}[\exp(\tau \epsilon_i^t) | \{x_i^t\}_i, \{\epsilon_i^{t-1}\}_i, \cdots, \{x_i^0\}_i, \{\epsilon_i^0\}_i] \leq \exp(\phi^2 \tau^2).$$

Remark. Taking $\mu(z) = z$ gives a linear model and $\mu(z) = (1 + e^{-z})^{-1}$ gives a logistic model. Also note that the assumptions imply there exists $\eta > 0$ such that $\mu'(x^\top \beta) \geq \eta$ for all $x, \beta \in \mathbb{R}^d$ satisfying $\|x\|_2 \leq B$ and $\|\beta\|_2 \leq M$ (see Lemma 4 in Appendix C for a short proof).

We are interested in learning a strategy that can identify all the feature vectors $x \in \mathbb{R}^d$ that have response y above some pre-specified cutoff c, while making as few mistakes as possible along the sequential learning process compared to the Bayes-optimal oracle that knows β^* (i.e., the classifier $x \mapsto \mathbb{1}\{\mu(x^\top \beta^*) \geq c\}$). It is worth noting that we don't make any distributional assumption on the feature vectors $x \in \mathbb{R}^d$. Thus, our adaptive algorithm works in both the adversarial setting and the stochastic setting where the features are drawn i.i.d. from some unknown underlying distribution.

Our goal is to minimize the objective formally defined in Definition 1, which penalizes exactly when the model performs an incorrect prediction compared to the Bayes-optimal decision rule, and the penalty is the distance of the expected response value for that example to the desired cutoff c.

Definition 1 (One-Sided Loss)**.** *For feature-action pairs $(x_i^t, a_i^t)_{i=1}^N \in \mathbb{R}^d \times \{0, 1\}$, the one-sided loss incurred at time t on a batch of size N with cutoff at c is the following:*

$$r_t := \sum_{i=1}^{N} |\mu(x_i^{t\top} \beta^*) - c| \cdot \mathbb{1}\Big\{\mathbb{1}\{\mu(x_i^{t\top} \beta^*) > c\} \neq a_i^t\Big\}. \tag{1}$$

We give an illustrative example of how this objective naturally arises in practice. Suppose that a company is looking to hire job applicants, where each applicant will contribute some variable amount of revenue to the company and the cost of hiring an applicant is a fixed cost of c. If the company makes the *correct* decision on each applicant, it will incur no loss, where correct means that it hired exactly the applicants whose expected revenue contribution to the company is at least c. The company incurs loss whenever it makes an incorrect decision: if it hires an applicant whose expected revenue is below c, it is penalized on the difference. Likewise, if it doesn't hire an applicant whose expected revenue is above c, it is also penalized for the expected profit that could have been made. Moreover, this definition of loss promotes a notion of *individual fairness* because it encourages the decision maker to not hire an unqualified applicant over a qualified one. While our setup captures scenarios beyond fairness applications, this aspect of individual fairness in one-sided learning may be of independent interest.

4 PASSIVE LEARNER HAS SLOW RATE

In this section, we show that under the stronger i.i.d data generation assumption, in order to achieve asymptotically vanishing loss, one could leverage an "offline" algorithm that learns on an initial training set only, but at the cost of having a slower rate for the one-sided loss we are interested in. Our passive learner (Algorithm 1) proceeds by predicting positively on the first $K + S$ samples to collect the labeled examples to fit on, where the first K samples are used to obtain a finite set of models which represent all possible binary decision combinations on these K samples that could have been made by the GLM model. The entire observed $K + S$ labeled examples are then used to pick the best model from this finite set to be used for the remaining rounds without further updating.

Algorithm 1 PASSIVE LEARNER

Inputs: Discretization sample size K, Exploration sample size S, cutoff c, Time horizon T

Initialization: Choose to observe pairs of $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$ for $K+S$ rounds, set the action $a_i = 1$.

1. Construct discretized strategy class $\hat{\Pi}$ using the first K samples, containing one representative $\hat{\beta}^k \in \mathbb{R}^d$ for each element of the set $\{(\pi(x_1), \cdots, \pi(x_K)) : \pi \in \Pi\}$.
2. Find the best strategy on the observed $K+S$ data pairs as:

$$\hat{\pi}_K^{\hat{\beta}^*} = \arg\min_{\pi \in \hat{\Pi}} \sum_{t=1}^{K+S} u_t(x_t, \pi(x_t))$$

for $t = K+S+1, \cdots, T$ **do**

Output $a_t = \hat{\pi}_K^{\hat{\beta}^*}(x_t) = \mathbb{1}\{\mu(x_t^\top \hat{\beta}^*) \geq c\}$ as decision on x_t, observe y_t if $a_t = 1$

Output: $\hat{\beta}^*, \{a_t\}_t$

More formally, we work with the setting where the feature-utility pairs $(x_t, u_t) \sim \mathcal{P}$ are generated i.i.d in each round. Let the class of strategies be $\Pi = \{\pi^\beta : \|\beta\|_2 \leq M\}$, where $\pi^\beta(x) := \mathbb{1}\{\mu(x^\top \beta) \geq c\}$ is the threshold rule corresponding to parameter β. Moreover, let the utility for covariate x_t with action $a_t \in \{0, 1\}$ be

$$u_t(x_t, a_t) := |y_t - c| \cdot \mathbb{1}\Big\{\mathbb{1}\{y_t > c\} \neq a_t\Big\}.$$

The initial discretization of the strategy class is used for a covering argument, the size of which is bounded with VC dimension. Using Hoeffding's inequality and a union bound over $|\hat{\Pi}|$, one can easily obtain a high-probability deviation on the quantity

$$\left| \frac{1}{K+S} \sum_{t=1}^{K+S} u_t(x_t, \hat{\pi}(x_t)) - \mathbb{E}_{\mathcal{P}(u,x)}[u(x, \hat{\pi}(x))] \right|$$

uniformly over all $\hat{\pi} \in \hat{\Pi}$, after which the optimality of $\hat{\pi}_K$ is invoked for reaching the final conclusion. We show that with optimal choices of K and S, Algorithm 1 has suboptimal guarantees – needing as many as $\widetilde{\mathcal{O}}(1/\epsilon^3)$ rounds in order to attain an average one-sided loss of at most ϵ, whereas our adaptive algorithm to be introduced later will only need $\widetilde{\mathcal{O}}(1/\epsilon^2)$ rounds. This suggests the importance of having the algorithm *actively* engaging throughout the data streaming process, beyond working with large collection of observational data only, for efficient learning. We give the guarantee in the proposition below. The proof is in Appendix A.

Proposition 1 (Bound for Algorithm 1)**.** *Under Assumption 1-4 and the additional assumption that the feature-utility pairs $(x_t, u_t) \sim \mathcal{P}$ are drawn i.i.d in each round, we have that picking $K = \mathcal{O}(T^{1/3})$, $S = \mathcal{O}(T^{2/3})$ in Algorithm 1, for $C_{T,\delta} = LBM + \gamma + c + \phi\sqrt{\log(2T/\delta)}$, with probability at least $1 - 2\delta$,*

$$\sum_{t=1}^{T} \mathbb{E}_{\mathcal{P}}[u(x, a_t)] \leq \min_{\pi \in \Pi} \sum_{t=1}^{T} \mathbb{E}_{\mathcal{P}}[u(x, \pi(x))] + \mathcal{O}\Big(C_{T,\delta} T^{2/3} d \log\Big(\frac{T}{d\delta}\Big)\Big).$$

This in turn gives the following one-sided loss bound with the same probability:

$$\mathbb{E}\Big[\sum_{t=1}^{T} r_t\Big] \leq \mathcal{O}\Big(C_{T,\delta} T^{2/3} d \log\Big(\frac{T}{d\delta}\Big)\Big).$$

5 GREEDY ACTIVE LEARNER MAY NOT CONVERGE

In this section, we show that the greedy active learner, which updates the model after each round on the received labeled examples without regards for one-sided feedback, can fail to find the optimal decision rule, even under the i.i.d data assumption. More specifically, the greedy learner fits parameter $\hat{\beta}$ that minimizes the empirical loss $\sum_{(x_t, y_t):\, a_t = 1} \ell(x_t, y_t; \beta)$ on the datapoints whose labels it has observed so far at each time step. For example in the case $\mu(z) = z$ we use $\ell(x, y; \beta) = (x^\top \beta - y)^2$, the squared loss; when $\mu(z) = (1 + e^{-z})^{-1}$ we instead use $\ell(x, y; \beta) = -y \log(\mu(x^\top \beta)) - (1 - y) \log(1 - \mu(x^\top \beta))$, the cross-entropy loss. An alternative definition of the greedy learner can utilize the decision rule mandated by the $\hat{\beta}$ that minimizes the one-sided loss (Definition 1) on the datapoints predicted positive thus far. In our setup this is possible because whenever a datapoint label is revealed, the loss incurred by the decision can be estimated. As it turns out, these two methods share similar behavior, and we refer to reader to Appendix B for the discussion of this alternative method.

We illustrate in Theorem 1 below that even when allowing warm starting with full-rank randomly drawn i.i.d samples, there are settings where the greedy learner will fail to converge. More specifically, if the underlying data distribution produces with constant probability a vector v with the rest of the mass concentrated on the orthogonal subspace, under Gaussian noise assumption, the prediction $\mu(v^\top \hat{\beta})$ has Gaussian distribution centered at the true prediction $\mu(v^\top \beta^*)$. Using the Gaussian anti-concentration inequality from Lemma 3 provided in Appendix B, we can show that if $\mu(v^\top \beta^*)$ is too close to the decision boundary c, there is a constant probability that the model will predict $\mu(v^\top \hat{\beta}) < c$, and therefore the model may never gather more information

in direction v for updating its prediction since no more observation will be made on v's label from this point on. This situation can arise for instance when dealing with a population consisting of two subgroups having small overlap between their features.

Theorem 1 (Non-Convergence for Greedy Learner)**.** *Let $y = \mu(x^\top \beta^*) + \epsilon$ with $\epsilon \sim \mathcal{N}(0,1)$ and independent of x. Moreover, for $v \in \mathbb{R}^d$, let P be a distribution such that $P(v) = 1/10$ and for all other vectors $v' \sim P$, it holds that $v'^\top v = 0$. Consider an MLE fit using $\ell(x, y; \beta)$ with n pairs of i.i.d. samples from P for warm starting the greedy learner. Under the additional assumption that $x_1, \cdots, x_n$ span all of $\mathbb{R}^d$, if $\mu(v^\top \beta^*) = c + \tau$, with $\tau \leq 1/\sqrt{n'}$ (where n' is the number of samples among $\{(x_i, y_i)\}_{i=1}^n$ with $x_i = v$), the loss after round T is lower bounded as:*

$$\mathbb{E}\left[\sum_{t=1}^{T} r_t\right] \geq \Omega((T-n)\tau)\,.$$

6 ADAPTIVE ALGORITHM

We propose Algorithm 2 with the goal of minimizing the cumulative one-sided loss at time horizon T, $R_T := \sum_{t=1}^T r_t$, independent of the data distribution at each round. The algorithm proceeds by first training a model on an initial labeled sample with the assumption that after initialization, the empirical covariance matrix A is invertible with the smallest eigenvalue $\lambda_0 > 0$. At each time step, we solve for the MLE fit $\hat{\beta}_t$ on the examples observed so far, using e.g., Newton's method. If $\|\hat{\beta}_t\|_2$ is too large, we perform a projection step – this step is only required as an analysis artifact to ensure that $\mu'(\cdot) > 0$ whenever it is evaluated in the algorithm. The model then produces point estimate $\mu(x^\top \beta_t)$ for each example x in the current batch.

From here, we adopt an adaptive approach based on the point-wise uncertainty in the prediction, which for data point x is proportional to $\sqrt{x^\top A^{-1} x}$ (where A is the covariance matrix of the labeled examples the model is fit on thus far). This choice is justified by showing that for any $X \in \mathbb{R}^{N \times (d+1)}$, whose rows consist of either $[0; x_i^t]$ or $[1; 0_d]\,, \forall i \in [N]$, we have with high probability

$$|1^\top \mu(X\tilde{\beta}^*) - 1^\top \mu(X\tilde{\beta}_t)| \leq \rho_t(\delta) \cdot \sum_{i=1}^{N} \sqrt{\bar{x}_i^\top A_{t-1}^{-1} \bar{x}_i}$$

for $\tilde{\beta}^* := [\mu^{-1}(c); \beta^*]$ the parameter of the optimal predictor and $\tilde{\beta}_t := [\mu^{-1}(c); \beta_t]$ our current best guess, where $\bar{x}_i$ is the last d coordinates of the i-th row of the matrix X. With this on hand, a short calculation reveals that the loss incurred at all time step $t \leq T$, with probability at least $1 - \delta$, is upper bounded as

$$r_t \leq 2\rho_t(\delta/2T) \cdot \sum_i \sqrt{x_{t,i}^\top A_{t-1}^{-1} x_{t,i}}$$

Algorithm 2 Adaptive One-sided Batch Alg.

Inputs: Batch size N, initialization sample size $K \geq d+1$ and eigenvalue $\lambda_0 > 0$, cutoff c
Inputs: Lipschitz constant L, norm bounds M, B, ϕ, η, γ, time horizon T, confidence level $\delta \in (0, 1 \wedge d/e)$
Initialization: Choose to observe K pairs of $\{(x_i^0, y_i^0)\}_{i=1}^K \in \mathbb{R}^d \times \mathbb{R}$, set $A \leftarrow \sum_{i=1}^K x_i^0 x_i^{0\top}$
Set $\kappa = \sqrt{3 + 2\log(1 + 2NB^2/\lambda_0)}$
for $t = 1, \cdots, T$ **do**
  Solve for $\hat{\beta}_t \in \mathbb{R}^d$ such that

$$\sum_{i=0}^{t-1} X_i^\top (y_i - \mu(X_i \hat{\beta}_t)) = 0_d \tag{2}$$

  if $\|\hat{\beta}_t\|_2 \leq M$ **then** $\beta_t \leftarrow \hat{\beta}_t$
  else Perform projection step on $\hat{\beta}_t$ as

$$\beta_t = \operatorname*{argmin}_{\|\beta\|_2 \leq M} \left\| \sum_{i=1}^{t-1} X_i^\top \mu(X_i\beta) - \sum_{i=1}^{t-1} X_i^\top \mu(X_i \hat{\beta}_t) \right\|_{A^{-1}}$$

  Set $\rho_t(\delta) = \frac{2L}{\eta} \kappa C_{T,\delta} \sqrt{2d \log t} \sqrt{\log(2dT/\delta)}$
  Initialize $X_t, y_t = \emptyset$
  for $j = 1, \cdots, N$ **do**
    if $\mu(x_j^{t\top} \beta_t) - c + \rho_t(\delta) \sqrt{x_j^{t\top} A^{-1} x_j^t} > 0$ **then**
      Choose to observe y_j^t and set $a_j^t = 1$
      Update $X_t \leftarrow [X_t; x_j^t]\,, y_t = [y_t; y_j^t]$
      Let $A \leftarrow A + x_j^t x_j^{t\top}$

Output: $\beta_T, \{a_j^t\}$

for $x_{t,i}$ the i-th row of X_t. It only remains to upper bound $\sum_i \|x_{t,i}\|_{A_{t-1}^{-1}}$, for which matrix determinant lemma is invoked for volume computation of matrices under low-rank updates.

Intuitively, the algorithm chooses to observe the samples for which either we can't yet make a confident decision, by which collecting the sample would greatly reduce the uncertainty in that corresponding subspace (as manifested by reduction in $\sum_i \|x_{t,i}\|_{A_{t-1}^{-1}}$ for future rounds after updating the model at the end of the batch); or we are confident that the response of the sample is above c (for which current decision would incur small loss). We give the following result whose proof can be found in Appendix C.

Theorem 2 (Guarantee for Algorithm 2)**.** *Suppose that Assumption 1-4 hold. Given a batch size N, we have that for all $T \geq 1$,*

$$R_T \leq \widetilde{\mathcal{O}}\Big(C_{T,\delta} K + \frac{L}{\eta} C_{T,\delta} \sqrt{Ts} dN\Big)$$

with probability at least $1 - 2\delta$ for $0 < \delta < \min\{1, d/e\}$,

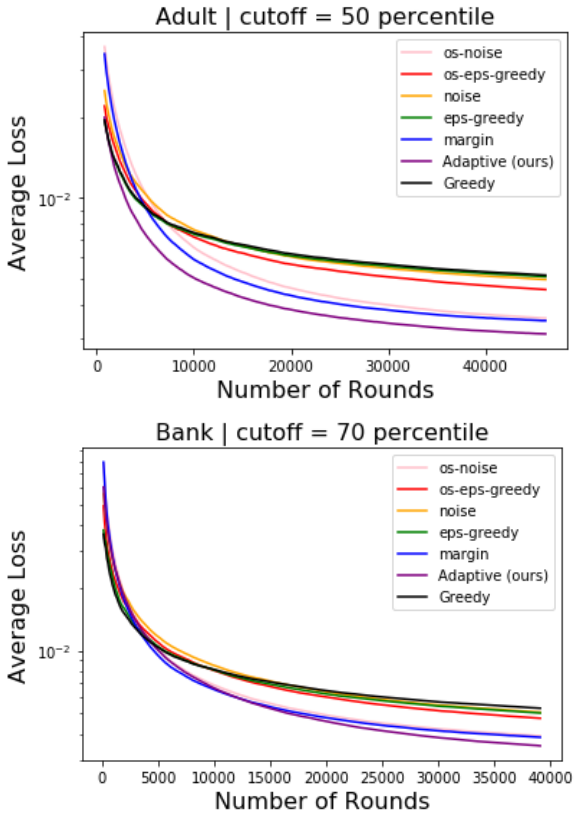

Figure 1: Average one-sided loss R_t/t for OLS. Each round consists of presenting a batch of 1 example. All methods are under optimal tuning averaged across 10 runs. The rest of the charts are in Appendix E.

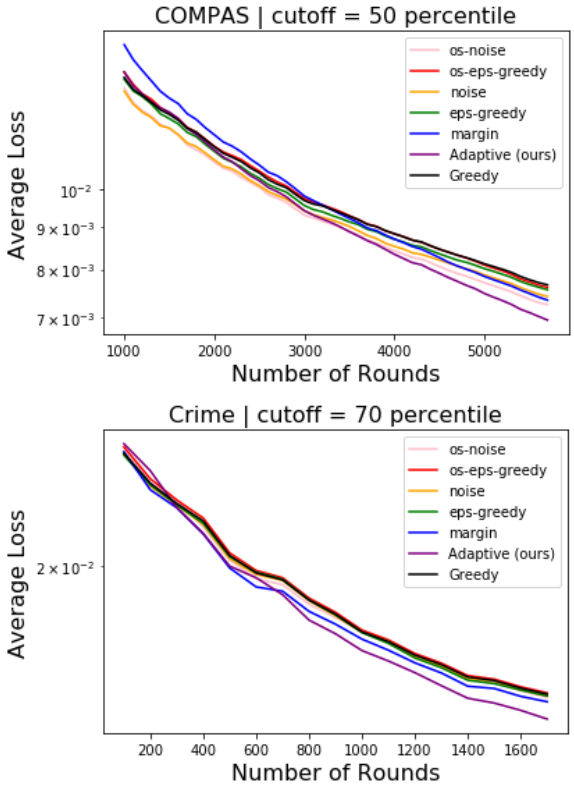

Figure 2: Average one-sided loss R_t/t for Logistic. Each round consists of presenting batch of 100 samples. All methods are under optimal tuning averaged across 10 runs. The rest of the charts are in Appendix E.

where $s = \min(N, d)$, $C_{T,\delta} = LBM + \gamma + c + \phi\sqrt{\log(2T/\delta)}$ *and* $\widetilde{\mathcal{O}}$ *hides poly-logarithmic factors in* $T, \delta^{-1}, d, N, B, \lambda_0$.

7 SGD UNDER ONE-SIDED FEEDBACK

In this section, we explore learning the parameter β^* with iterative updates under one-sided feedback. We consider running projected SGD with the following gradient update on (x_t, y_t) at time step t for the GLM model:

$$\beta_{t+1} = \mathcal{P}_\Omega\big(\beta_t - \eta \cdot (-y_t x_t + \mu(x_t^\top \beta_t)x_t) \cdot \mathbb{1}\{\mu(x_t^\top \beta_t) + s_t \geq c\}\big)$$

for some exploration bonus s_t to be specified later, where the projection assures that $\mu'(\cdot) \geq \gamma$ for some $\gamma > 0$ throughout the execution of the algorithm. For example in logistic regression, we project onto the convex set $\Omega := \{\beta \colon |x_t^\top \beta| \leq r\}$ at each step t to maintain this. In words, we perform prediction on x_t with the current parameter β_t, and take a stochastic projected gradient step on the sample if $\mu(x_t^\top \beta_t) + s_t \geq c$. Since the noise ϵ_t is assumed to be zero-mean and independent of β_t and x_t, this implies that in expectation (condition on β_t), we have

$$\begin{aligned}\beta_{t+1} - \beta^* &= \mathcal{P}_\Omega\big(\beta_t - \beta^* - \eta \cdot (-\mu(x_t^\top \beta^*)x_t \\ &\quad + \mu(x_t^\top \beta_t)x_t) \cdot \mathbb{1}\{\mu(x_t^\top \beta_t) + s_t \geq c\}\big) \\ &= \mathcal{P}_\Omega\big(\beta_t - \beta^* - \eta \cdot \mu'(z) x_t^\top (\beta_t - \beta^*)x_t \\ &\quad \cdot \mathbb{1}\{\mu(x_t^\top \beta_t) + s_t \geq c\}\big)\end{aligned}$$

where we used mean value theorem for some $z \in [x_t^\top \beta^*, x_t^\top \beta_t]$. Taking norms on both sides and using the fact that convex projection is a contractive mapping, we have at step t, the expected progress as:

$$\begin{aligned}\|\beta_{t+1} - \beta^*\|_2^2 &\leq \|\beta_t - \beta^* - \eta \cdot \mu'(z) x_t^\top (\beta_t - \beta^*)x_t\|_2^2 \\ &= \|\beta_t - \beta^*\|_2^2 + \big[\eta^2 \cdot \mu'(z)^2 \|x_t\|_2^2 \\ &\quad - 2\eta \cdot \mu'(z)\big](x_t^\top (\beta_t - \beta^*))^2\end{aligned} \tag{3}$$

if $\mu(x_t^\top \beta_t) + s_t \geq c$; and contraction ratio of 1 (i.e., no update on β) if $\mu(x_t^\top \beta_t) + s_t < c$. This suggests that in the case where we choose to accept, either $|x_t^\top (\beta_t - \beta^*)|$

Dataset	cutoff	greedy	ϵ-grdy	os-ϵ-grdy	noise	os-noise	margin	ours
Adult	50%	239.45	236.34	211.74	230.77	165.77	162.31	**144.92**
	70%	134.74	134.18	133.8	131.66	132.39	132.67	**129.81**
Bank	50%	164.23	162.67	117.86	136.0	88.49	86.26	**74.64**
	70%	207.6	197.0	185.9	198.66	153.3	150.75	**137.24**
COMPAS	50%	41.56	36.67	36.93	36.93	28.09	28.12	**26.01**
	70%	41.66	39.16	39.61	39.87	38.03	36.98	**34.07**
Crime	50%	15.77	15.77	15.5	15.66	14.93	14.73	**13.95**
	70%	22.0	21.75	21.99	20.33	20.63	20.1	**19.19**
German	50%	14.7	14.51	14.12	13.62	11.12	10.52	**9.63**
	70%	15.89	15.53	15.93	15.41	14.09	14.52	**13.07**
Blood	50%	2.06	2.06	2.06	2.06	1.92	1.72	**1.52**
	70%	3.7	2.78	3.04	**2.38**	3.13	3.06	2.65
Diabetes	50%	4.17	4.16	4.23	3.94	3.81	3.95	**3.61**
	70%	6.05	5.56	6.14	6.05	5.6	5.39	**5.33**
EEG Eye	50%	256.47	200.04	175.8	173.52	106.26	**96.85**	119.7
	70%	175.71	167.94	168.73	157.68	167.52	160.76	**155.79**
Australian	50%	3.74	3.74	3.77	3.63	3.0	2.79	**2.65**
	70%	6.77	6.77	6.77	6.66	5.09	5.26	**4.65**
Churn	50%	46.98	43.65	30.65	36.64	21.24	18.83	**14.89**
	70%	49.99	47.84	47.91	49.89	41.18	36.17	**35.27**

Table 1: Experimental results for cumulative one-sided loss for Linear Regression.

is small, in which case the probability of making a mistake on this sample is small already; or if large we make sufficient progress in this direction by performing the update. This is formalized in Algorithm 3 and the corresponding Proposition 2 below, whose proof we defer to Appendix D. In order to have any hope of making progress towards β^* (i.e., observing y_t with non-trivial probability), however, we make the following assumption on the feature vectors.

Assumption 5 (Subgaussian i.i.d Features)**.** *The feature vectors x_t at each time step t are drawn i.i.d with independent σ-sub-gaussian coordinates. This in turn implies that since $\mu(x^\top\beta^*)$ is a univariate L-lipschitz function of $\|\beta^*\|_2\sigma$-subgaussian random variable, $\mu(x_t^\top\beta^*) - \mathbb{E}_x[\mu(x^\top\beta^*)]$ is itself $CL\|\beta^*\|\sigma$-subgaussian for some numerical constant C.*

Proposition 2. *Under Assumption 1 and 5, given $\rho \in (0,1)$, we have with probability at least $1-\rho$, for cutoff $c = \mathbb{E}_x[\mu(x^\top\theta^*)] - \zeta$ with $\zeta \geq \sqrt{2L\|\beta^*\|^2\sigma^2\log(\rho^{-1})}$, at iteration t of Algorithm 3, either*

$$\mathbb{E}[\|\beta_{t+1} - \beta^*\|^2] \leq \mathbb{E}[\|\beta_t - \beta^*\|_2^2] - \frac{\alpha^2}{\|x_t\|_2^2 L^2},$$

or

$$\left|\mu(x_t^\top\beta^*) - \mu(x_t^\top\beta_t)\right| \leq L\gamma^{-1}(\alpha + 2B)$$

if picking $\delta^{-1} = \rho - e^{-\frac{\zeta^2}{2L^2\|\beta^*\|^2\sigma^2}}$. Moreover, the probability of making a misclassification error at time step t satisfies

$$\mathbb{P}\left(\mathbb{1}\{\mu(x_t^\top\beta^*) \geq c\} \neq \mathbb{1}\{\mu(x_t^\top\beta_t) + s_t \geq c\}\right) \leq \rho.$$

Algorithm 3 SGD Under Partial Feedback

Inputs: Initial β_0 and d_0 such that $\|\beta_0 - \beta^*\| \leq d_0$
Inputs: Accuracy α, Lipschitz const L, param δ
Inputs: Bound B such that $|\epsilon_t| \leq B\ \forall t$
for $t = 0, \cdots, T$ **do**
  Set $s_t = L \cdot \delta \cdot d_t\|x_t\|_2$
  if $\mu(x_t^\top\beta_t) + s_t < c$ **then**
    Don't accept x_t, keep $\beta_{t+1} = \beta_t$ and $d_{t+1} = d_t$
  else Accept x_t and receive label y_t
    if $\left|y_t - \mu(x_t^\top\beta_t)\right| \leq \alpha + B$ **then**
      Set $\beta_{t+1} = \beta_t$ and $d_{t+1} = d_t$
    else Update as $\beta_{t+1} = \mathcal{P}_\Omega(\beta_t - (L\|x_t\|_2^2)^{-1} \cdot (-y_t x_t + \mu(x_t^\top\beta_t)x_t))$; set $d_{t+1}^2 = d_t^2 - \alpha^2\|x_t\|_2^{-2}L^{-2}$
Output: β_T

Remark. If we are interested in cutoff $c = \mathbb{E}[\mu(x^\top\beta^*)] + \zeta$ for some $\zeta > 0$, a similar argument shows that picking $s_t = -L \cdot \delta \cdot d_t\|x_t\|_2$ will give the same misclassification error probability ρ, with the exception of course being that we won't be able to get the high probability contraction ratio for $\|\beta_t - \beta^*\|_2$ due to the lack of observations on y_t.

8 EXPERIMENTS

To further support our theoretical findings and demonstrate the effectiveness of our algorithm in practice, we test our method on the following datasets:

	cutoff	greedy	ϵ-grdy	os-ϵ-grdy	noise	os-noise	margin	ours
Adult	50%	43.48	43.55	43.48	43.35	43.41	43.38	**42.63**
	70%	102.86	102.86	102.9	102.6	102.81	102.47	**100.06**
Bank	50%	23.22	23.26	**23.18**	23.3	23.33	23.2	23.23
	70%	85.72	85.94	85.67	85.51	**85.26**	85.27	85.75
COMPAS	50%	44.47	43.88	44.15	43.07	42.11	42.64	**40.34**
	70%	43.7	43.59	**43.41**	43.66	43.83	43.7	43.7
Crime	50%	11.04	10.83	11.04	10.85	10.33	10.44	**9.42**
	70%	26.05	25.93	26.13	25.94	25.84	25.55	**24.46**
German	50%	35.71	35.21	33.55	33.35	24.19	23.19	**20.33**
	70%	42.55	41.14	42.18	40.98	40.64	40.3	**37.12**
Blood	50%	5.05	5.05	4.87	4.83	4.71	4.53	**4.24**
	70%	13.04	13.04	13.03	13.04	10.84	12.14	**9.69**
Diabetes	50%	28.23	28.23	27.75	27.22	26.67	26.18	**25.16**
	70%	29.36	28.0	27.79	28.0	**27.4**	27.9	28.11
EEG Eye	50%	239.33	238.92	239.09	236.65	200.61	201.51	**187.28**
	70%	209.48	207.89	208.83	206.63	204.94	205.4	**199.04**
Australian	50%	21.88	21.88	21.87	21.21	21.76	20.81	**20.38**
	70%	17.47	17.29	17.46	**16.49**	17.24	17.46	17.43
Churn	50%	61.04	57.74	54.13	53.85	39.46	38.88	**34.89**
	70%	122.96	117.49	116.04	112.36	94.61	88.3	**82.23**

Table 2: Experimental results for cumulative one-sided loss for Logistic Regression.

1. **Adult** Lichman et al. (2013) (48842 examples). The task is to predict whether the person's income is more than 50k.
2. **Bank Marketing** Lichman et al. (2013) (45211 examples). Predict if someone will subscribe to a bank product.
3. **ProPublica's COMPAS** ProPublica (2018) (7918 examples). Recidivism data.
4. **Communities and Crime** Lichman et al. (2013) (1994 examples). Predict if community is high (>70%tile) crime.
5. **German Credit** Lichman et al. (2013) (1000 examples). Classify into good or bad credit risks.
6. **Blood Transfusion Service Center** Vanschoren et al. (2013) (784 examples). Predict if person donated blood.
7. **Diabetes** Vanschoren et al. (2013) (768 examples). Detect if patient shows signs of diabetes.
8. **EEG Eye State** Vanschoren et al. (2013) (14980 examples). Detect if eyes are open or closed based on EEG data.
9. **Australian Credit Approval** Vanschoren et al. (2013) (690 examples). Predict for credit card approvals.
10. **Churn** Vanschoren et al. (2013) (5000 examples). Determine whether or not the customer churned.

We compare against the following baselines:
1. **Greedy**, where we perform least-squares/logistic fit β_t on the collected data and predict positive/observe label if $\mu(x^\top \beta_t) > c$.
2. **ϵ-Greedy** Sutton and Barto (2018), which with probability $\alpha/\sqrt{t}$, we make a random decision on the prediction (with equal probability), otherwise we use the greedy approach.
3. **One-sided ϵ-Greedy**, which with probability $\alpha/\sqrt{t}$ we predict positively, otherwise we use the greedy approach. This baseline is inspired from ideas in the original apple tasting paper Helmbold et al. (2000).
4. **Noise**, which we add $\alpha u/\sqrt{t}$ to the prediction where u is drawn uniformly on $[-\frac{1}{2}, \frac{1}{2}]$.
5. **One-sided Noise**, which we add $\alpha u/\sqrt{t}$ to the prediction where u is drawn uniformly on $[0, 1]$.
6. **Margin**, which we add $\alpha/\sqrt{t}$ to the prediction. This can be seen as a non-adaptive version of our approach, since the quantity we add to the prediction for this baseline is uniform across all points.

For each dataset, we take all the examples and make a random stratified split so that 5% of the data is used to train the initial model and the rest is used for online learning. For the linear regression experiments, we used a batch size of 1 while for logistic regression we used a batch size of 1000 for Adult, Bank, EEG Eye State and 100 for the rest due to computational costs of retraining after each batch using `scikit-learn`'s implementation of logistic regression. We compute the loss based on using an estimated β^* obtained by fitting the respective model (either linear or logistic) on the entire dataset. Due to space limitation, we only show

the results for cutoff c chosen so that 50% and 70% of the data points are below the cutoff w.r.t. β^* in Table 1 for linear regression and Table 2 for logistic regression. Full results are in Appendix E. For each dataset and setting of c, we averaged the performance of each method across 10 different random splits of the dataset and tuned α over a grid of powers of 2 (except greedy).

9 DISCUSSION

Many machine learning systems learn under active one-sided feedback, where experimental design is intertwined with the decision making process. In such scenarios, the data collection is informed by past decisions and can be inherently biased. In this work, we show that without accounting for such biased sampling, the model could enter a feedback loop that only reinforce its past misjudgements, resulting in a strategy that may not align with the long term learning goal. Indeed, we demonstrate that the de facto default approach (i.e., greedy or passive learning) often yields suboptimal performance when viewed through this lens. In turn, we propose a natural objective for the one-sided learner and give a practical algorithm that can be used to avoid such undesirable downstream effects. Both the theoretical grounding and the empirical effectiveness of the proposed algorithm offer evidence that it serves as a much better alternative in such settings.

References

András Antos, Gábor Bartók, Dávid Pál, and Csaba Szepesvári. Toward a classification of finite partial-monitoring games. *Theoretical Computer Science*, 473:77–99, 2013.

Gábor Bartók and Csaba Szepesvári. Partial monitoring with side information. In *International Conference on Algorithmic Learning Theory*, pages 305–319. Springer, 2012.

Yahav Bechavod, Katrina Ligett, Aaron Roth, Bo Waggoner, and Steven Z Wu. Equal opportunity in online classification with partial feedback. In *Advances in Neural Information Processing Systems*, pages 8972–8982, 2019.

Richard Berk. An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. *Journal of Experimental Criminology*, 13 (2):193–216, 2017.

Alina Beygelzimer, John Langford, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandit algorithms with supervised learning guarantees. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 19–26, Fort Lauderdale, FL, USA, 2011. PMLR.

Antoine Bordes, Seyda Ertekin, Jason Weston, and Léon Bottou. Fast kernel classifiers with online and active learning. *Journal of Machine Learning Research*, 6(Sep):1579–1619, 2005.

Nicolo Cesa-Bianchi, Claudio Gentile, and Luca Zaniboni. Worst-case analysis of selective sampling for linear classification. *Journal of Machine Learning Research*, 7(Jul):1205–1230, 2006a.

Nicolo Cesa-Bianchi, Gábor Lugosi, and Gilles Stoltz. Regret minimization under partial monitoring. *Mathematics of Operations Research*, 31(3):562–580, 2006b.

Philip K Chan, Wei Fan, Andreas L Prodromidis, and Salvatore J Stolfo. Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems and Their Applications*, 14(6):67–74, 1999.

Wei Chu, Martin Zinkevich, Lihong Li, Achint Thomas, and Belle Tseng. Unbiased online active learning in data streams. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 195–203, 2011.

Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*, pages 191–198, 2016.

Danielle Ensign, Sorelle A Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. Decision making with limited feedback: Error bounds for recidivism prediction and predictive policing. 2017.

Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 586–594. 2010.

Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. In *Advances in neural information processing systems*, pages 3315–3323, 2016.

Xinran He, Junfeng Pan, Ou Jin, Tianbing Xu, Bo Liu, Tao Xu, Yanxin Shi, Antoine Atallah, Ralf Herbrich, Stuart Bowers, et al. Practical lessons from predicting clicks on ads at facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, pages 1–9, 2014.

David P Helmbold, Nicholas Littlestone, and Philip M Long. Apple tasting. *Information and Computation*, 161(2):85–139, 2000.

Nitin Jindal and Bing Liu. Review spam detection. In *Proceedings of the 16th international conference on World Wide Web*, pages 1189–1190, 2007.

Gang Kou, Yi Peng, and Chen Lu. Mcdm approach to evaluating bank loan default models. *Technological and Economic Development of Economy*, 20(2):292–311, 2014.

Unghui Lee, Sangyol Yoon, HyunChul Shim, Pascal Vasseur, and Cedric Demonceaux. Local path planning in a complex environment for self-driving car. In *The 4th Annual IEEE International Conference on Cyber Technology in Automation, Control and Intelligent*, pages 445–450. IEEE, 2014.

Moshe Lichman et al. Uci machine learning repository, 2013.

Jing Lu, Peilin Zhao, and Steven CH Hoi. Online passive-aggressive active learning. *Machine Learning*, 103(2):141–183, 2016.

Brian Paden, Michal Čáp, Sze Zheng Yong, Dmitry Yershov, and Emilio Frazzoli. A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Transactions on intelligent vehicles*, 1(1):33–55, 2016.

Michael J Pazzani and Daniel Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007.

Juan C Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. *arXiv preprint arXiv:2002.06673*, 2020.

ProPublica. Compas recidivism risk score data and analysis, Mar 2018. URL `https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis`.

D Sculley. Practical learning from one-sided feedback. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 609–618, 2007.

Abhinav Srivastava, Amlan Kundu, Shamik Sural, and Arun Majumdar. Credit card fraud detection using hidden markov model. *IEEE Transactions on dependable and secure computing*, 5(1):37–48, 2008.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

Abhishek Kumar Tiwari. Machine learning application in loan default prediction. *Machine Learning*, 4(5), 2018.

Nikolaj Tollenaar and PGM Van der Heijden. Which method predicts recidivism best?: a comparison of statistical, machine learning and data mining predictive models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2):565–584, 2013.

Chih-Fong Tsai and Ming-Lun Chen. Credit rating by hybrid machine learning techniques. *Applied soft computing*, 10(2):374–380, 2010.

Joaquin Vanschoren, Jan N. van Rijn, Bernd Bischl, and Luis Torgo. Openml: Networked science in machine learning. *SIGKDD Explorations*, 15(2):49–60, 2013. doi: 10.1145/2641190.2641198. URL `http://doi.acm.org/10.1145/2641190.2641198`.

Ping Wang, Rick Mathieu, Jie Ke, and HJ Cai. Predicting criminal recidivism with support vector machine. In *2010 International Conference on Management and Service Science*, pages 1–9. IEEE, 2010.