A Potential-based Framework for Online Multi-class Learning with Partial Feedback

Shijun Wang
Radiology and Imaging Sciences
National Institutes of Health
wangshi@cc.nih.gov

Rong Jin
Computer Science and Engineering
Michigan State University
rongjin@cse.msu.edu

Hamed Valizadegan
Computer Science and Engineering
Michigan State University
valizade@cse.msu.edu

Abstract

We study the problem of online multi-class learning with partial feedback: in each trial of online learning, instead of providing the true class label for a given instance, the oracle will only reveal to the learner if the predicted class label is correct. We present a general framework for online multi-class learning with partial feedback that adapts the potential-based gradient descent approaches (Cesa-Bianchi & Lugosi, 2006). The generality of the proposed framework is verified by the fact that Banditron (Kakade et al., 2008) is indeed a special case of our work if the potential function is set to be the squared L_2 norm of the weight vector. We propose an exponential gradient algorithm for online multi-class learning with partial feedback. Compared to the Banditron algorithm, the exponential gradient algorithm is advantageous in that its mistake bound is independent from the dimension of data, making it suitable for classifying high dimensional data. Our empirical study with four data sets show that the proposed algorithm for online learning with partial feedback is comparable to the Banditron algorithm.

1 Introduction

We study the problem of online multi-class learning with partial feedback. In particular, we assume that in each trial of online learning, instead of providing the true class label for a given instance, the oracle only reveals to the learner if the predicted class label is correct. Although online multi-class learning with full feedback had been extensively studied, the problem of online multi-class learning with partial feedback is only studied recently (Kakade et al., 2008; Langford & Tong, 2007). In this paper, we propose a general framework to address the challenge of partial feedback in the setup of online multi-class learning. This general framework adapts the potential-based gradient descent approaches for online learning (Cesa-Bianchi & Lugosi, 2006) to the scenario of partial feedback. The generality of the proposed framework is verified by the fact that banditron is indeed a special case of our work if the potential function is set to be the squared L_2 norm of the weight vector. Besides the general framework, we further propose an exponential gradient algorithm for online multi-class learning with partial feedback. Compared to the Banditron algorithm, the exponential gradient algorithm is advantageous in that its mistake bound is independent from the dimension of data, making it suitable for classifying high dimensional data. We verify the efficacy of the proposed algorithm for online learning with partial feedback by an extensive empirical study.

This paper is organized as follows: Section 2 reviews the related work. Section 3 presents the general framework for online multi-class learning with partial feedback, and the exponential gradient algorithm. Section 4 verifies the efficacy of the proposed algorithm empirically. Section 5 concludes this work.

2 Related Work

Although introduced very recently and there is only a few work directly related, the problem of online multi-class learning with bandit feedback can be traced back to online multi-class classification with full feedback and multi-armed bandit learning. Several additive and multiplicative online multi-class learning algorithms have been introduced in the literature (Crammer et al., 2003). Perceptron (Rosenblatt, 1958) and Winnow (Littlestone, 1988) are two such algo-

Appearing in Proceedings of the 13^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2010, Chia Laguna Resort, Sardinia, Italy. Volume 9 of JMLR: W&CP 9.

rithms. Kivinen and Warmuth developed potential functions that can be used to analyze different online algorithms (Kivinen & Warmuth, 1995). Grove et al. (Grove et al., 2001) showed that polynomial potential can be considered as a parameterized interpolation between additive and multiplicative algorithms.

Multi-armed bandit problem refers to the problem of choosing an action from a list of actions to maximize reward given that the feedback is (bandit) partial (Auer et al., 2003; Even-Dar et al., 2006). The algorithms developed for this problem usually utilize the exploitation vs. exploitation tradeoff strategy to handle the challenge with partial feedback (Even-Dar et al., 2002; Mannor & Tsitsiklis, 2004).

Multi-class learning with bandit feedback can be considered as a multi-armed bandit problem with side information. Langford et al. in (Langford & Tong, 2007) extended the multi-armed setting to the case where some side information is provided. Their setting has a high level of abstraction and its application to the multi-class bandit learning is not straightforward. Banditron, which can be considered as a special case of our framework, is a direct generalization of Perceptron to the case of partial feedback and uses exploration vs. exploitation tradeoff strategy to handle partial feedback (Kakade et al., 2008). Potential function and exploration vs. exploitation tradeoff techniques are the main tools used to develop the framework in this paper.

3 A Potential-based Framework for Multi-class Learning with Partial Feedback

We first present the problem of multi-class learning with partial feedback, followed by the presentation of potential-based framework and algorithms for automatically tuning the parameters.

3.1 Multi-class Learning with Partial Feedback

We denote by K the number of classes, and by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$ the sequence of training examples received over trials, where $\mathbf{x}_i \in \mathbb{R}^d$ and T is the number of received training instances. In each trial, we denote by $\widetilde{y}_i \in \{1, \ldots, K\}$ the predicted class label. Unlike the classical setup of online learning where an oracle provides the true class label $y_i \in \{1, \ldots, K\}$ to the learner, in the case of partial feedback, the oracle only tells the learner if the predicted class label is correct, i.e., $[y_t = \widetilde{y}_t]$. This partial feedback makes it difficult to learn a multi-class classification model.

In our study, we assume a linear classifier for each class, denoted by $W = (\mathbf{w}_1, \ldots, \mathbf{w}_K) \in \mathbb{R}^{d \times K}$, although the extension to nonlinear classifiers using kernel trick is straightforward. Given a training example $(\mathbf{x}, y)$, we measure its loss by $\ell\left(\max_{k \neq y} \mathbf{w}_k^\top \mathbf{x} - \mathbf{w}_y^\top \mathbf{x}\right)$ where $\ell(z) = \max(0, z + 1)$ is a hinge loss. We denote by $W^1, \ldots, W^T$ a sequence of linear classifiers generated by an online learning algorithm over the trials. Our objective is to bound the number of mistakes made by the online learning algorithm. Since the proposed framework is a stochastic algorithm, we will focus on the expectation of the mistake bound. As will be shown later, the expectation of the mistake bound is often written in the form

$$\alpha \Phi(U) + \beta \sum_{t=1}^{T} \ell\left(\max_{k \neq y_t} \mathbf{x}_t^\top \mathbf{u}_k - \mathbf{x}_t^\top \mathbf{u}_{y_t}\right)$$

where $U = (\mathbf{u}_1, \ldots, \mathbf{u}_K)$ is the linear classifier, $\Phi(W) : \mathbb{R}^{d \times K} \mapsto \mathbb{R}$ is a strictly convex function that measures the complexity of the linear classifiers, and α and β are weight constants for the complexity term and the classification errors. Note that the Banditron algorithm is a special case of the above framework where it measures the complexity of W by its Frobenius norm, i.e., $\Phi(W) = \frac{1}{2}|W|_F^2$. In this work, we design a general approach for online learning with partial feedback that is adapted to any complexity measure $\Phi(W)$. Finally, for the convenience of presentation, we define

$$\ell_t(W) = \ell\left(\max_{k \neq y_t} \mathbf{x}_t^\top \mathbf{w}_k^t - \mathbf{x}_t^\top \mathbf{w}_{y_t}^t\right) \tag{1}$$

3.2 Potential-based Online Learning for Partial Feedback

Our framework, depicted in Algorithm 1, generalizes the Banditron algorithm (Kakade et al., 2008) by considering any complexity measure $\Phi(W)$ that is strictly convex. In this algorithm, we introduce $\theta \in \mathbb{R}^{d \times K}$, the dual representation of the linear classifiers W. In each iteration, we first update θ_t based on the partial feedback $[y_t = \widetilde{y}_t]$, and compute the linear classifier W_t via the mapping $\nabla \Phi^*(\theta)$, where $\Phi^*(\theta)$ is the Lagendre conjugate of $\Phi(W)$. Similar to most online learning with partial feedback (Cesa-Bianchi & Lugosi, 2006), a stochastic approach is used to predict class assignment, in which parameter $\gamma > 0$ is introduced to ensure sufficient exploration (**?**).

In the following, we show the regret bound for the proposed algorithm. For the convenience of discussion, we define vector $\tau_t \in \mathbb{R}^d$ as

$$\tau_t = \mathbf{1}_{\widehat{y}_t} - \mathbf{1}_{y_t} \tag{4}$$

Proposition 1. *For $\ell_t(W) \geq 1$, we have*

$$\nabla \ell_t(W) = \left\langle W^{t-1}, \mathbf{x}_t \tau_t^\top \right\rangle, \quad \textit{and} \qquad \mathrm{E}_t[\delta_t] = \tau_t \tag{5}$$

where $\mathrm{E}_t[\cdot]$ is the expectation over $\widetilde{y}_t$ and δ_t is defined in (3).

Algorithm 1 Online Learning Algorithm for Multiclass Bandit Problem

1: Parameters:
- Smoothing parameter: $\gamma \in (0, 0.5)$
- Step size: $\eta > 0$
- Potential function: $\Phi : \mathbb{R}^{d\times K} \mapsto \mathbb{R}$ and its Legendre conjugate $\Phi^* : \mathbb{R}^{d\times K} \mapsto \mathbb{R}$

2: Set $\mathbf{w}_k^0 = \mathbf{0}, k = 1, \ldots, K$ and $\theta^0 = \nabla\Phi^*(W^0)$
3: **for** $t = 1, \ldots, T$ **do**
4: Receive $\mathbf{x}_t \in \mathbb{R}^d$
5: Compute

$$\widehat{y}_t = \arg\max_{1\le k\le K} \mathbf{x}_t^\top \mathbf{w}_k^{t-1} \tag{2}$$

6: Set $p_k = (1-\gamma)[k = \widehat{y}_t] + \gamma/K, k = 1, \ldots, K$
7: Randomly sample $\widetilde{y}_t$ according to the distribution $\mathbf{p} = (p_1, \ldots, p_K)$.
8: Predict $\widetilde{y}_t$ and receive feedback $[y_t = \widetilde{y}_t]$
9: Compute

$$\delta_t = \mathbf{1}_{\widehat{y}_t} - \mathbf{1}_{\widetilde{y}_t} \frac{[y_t = \widetilde{y}_t]}{p_{\widehat{y}_t}} \tag{3}$$

where $\mathbf{1}_k$ stands for the vector with all its elements being zero except its kth element is 1.
10: Compute $\theta^t = \theta^{t-1} - \eta\mathbf{x}_t\delta_t^\top$
11: Compute $W^t = \nabla\Phi(\theta^t)$ where $\theta^t = (\theta_1^t, \ldots, \theta_K^t)$
12: **end for**

We denote by $D_\Phi(A, B)$ the Bregman distance function for a given convex function Φ , which is defined as follows

$$D_\Phi(A, B) = \Phi(A) - \Phi(B) - \langle A - B, \nabla\Phi(B)\rangle \tag{6}$$

The following classical result in convex analysis summarizes useful properties of Bregman distance.

Lemma 1. *Let $\Phi(W)$ be a strictly convex function with constant ρ with respect to norm $\|\cdot\|$, i.e., for any W and W' we have*

$$\langle W - W', \nabla\Phi(W) - \nabla\Phi(W')\rangle \ge \rho\|W - W'\|^2.$$

We have the following inequality for any θ and θ'

$$\langle \theta - \theta', \nabla\Phi^*(\theta) - \nabla\Phi * (\theta')\rangle \le \frac{1}{\rho}\|\theta - \theta'\|_*^2$$

where $\Phi^*(\theta)$ is the Legendre conjugate of $\Phi(W)$, and $\|\cdot\|_*$ is dual of norm $\|\cdot\|$. Furthermore, we have the following equality for any W and W'

$$D_\Phi(W, W') = D_{\Phi^*}(\theta, \theta'),$$

where $\theta = \nabla\Phi(W)$ and $\theta' = \nabla\Phi(W')$.

Proposition 2. *For any linear classifier $U \in \mathbb{R}^{d\times K}$, We have the following inequality hold for two consecutive classifier W^{t-1} and W^t generated by Algorithm 1*

$$\begin{aligned} & D_{\Phi^*}(U, W^{t-1}) - D_{\Phi^*}(U, W^t) + D_{\Phi^*}(W^{t-1}, W^t) \\ = \; & -\langle U - W^{t-1}, \eta\mathbf{x}_t\delta_t^\top\rangle \end{aligned} \tag{7}$$

Proof. Using the property of Bregman distance function (see for example Chapter 11.2 in (Cesa-Bianchi & Lugosi, 2006)), we have

$$\begin{aligned} & D_{\Phi^*}(U, W^{t-1}) - D_{\Phi^*}(U, W^t) + D_{\Phi^*}(W^{t-1}, W^t) \\ = \; & \langle U - W^{t-1}, \nabla\Phi^*(W^t) - \nabla\Phi^*(W^{t-1})\rangle \\ = \; & \langle U - W^{t-1}, \theta^t - \theta^{t-1}\rangle = -\left\langle U - W^{t-1}, \eta\mathbf{x}_t\delta_t^\top\right\rangle \end{aligned}$$

The second step follows the property $\theta_t = \nabla\Phi^*(W^t)$, and the last step uses the updating rule of Algorithm 1. □

Proposition 3. *For any $s > 0$, we have*

$$\mathrm{E}[|\delta_t|_s^2] \le \frac{\gamma}{1-\gamma} + [\widehat{y}_t \ne y_t]\left\{1 - \gamma + \frac{\gamma}{K}\left(1 + \left[\frac{K}{\gamma}\right]^s\right)^{2/s}\right\}$$

Proof.

$$\begin{aligned} \mathrm{E}_t[|\delta_t|_s^2] &= \mathrm{E}_t\left[\left|\mathbf{1}_{\widehat{y}_t} - \mathbf{1}_{\widetilde{y}_t}\frac{[y_t = \widetilde{y}_t]}{p_{\widetilde{y}_t}}\right|_s^2\right] \\ &= [\widehat{y}_t = y_t]\mathrm{E}_t\left[\left|1 - \frac{[\widehat{y}_t = \widetilde{y}_t]}{p_{\widetilde{y}_t}}\right|^2\right] \\ &\quad + [\widehat{y}_t \ne y_t]\mathrm{E}_t\left[\left|\mathbf{1}_{\widehat{y}_t} - \mathbf{1}_{\widetilde{y}_t}\frac{[y_t = \widetilde{y}_t]}{p_{\widetilde{y}_t}}\right|_s^2 \middle| \widehat{y}_t \ne y_t\right] \\ &= [\widehat{y}_t = y_t]\frac{\gamma(K-1)/K}{1-\gamma+\gamma/K} + \\ &\quad [\widehat{y}_t \ne y_t]\left\{1 - \frac{\gamma}{K} + \frac{\gamma}{K}\left(1 + \left[\frac{K}{\gamma}\right]^s\right)^{2/s}\right\} \\ &\le \frac{\gamma}{1-\gamma} + [\widehat{y}_t \ne y_t]\left\{1 - \gamma + \frac{\gamma}{K}\left(1 + \left[\frac{K}{\gamma}\right]^s\right)^{2/s}\right\} \end{aligned}$$

□

We use $|W|_{p,s}$ to measure the norm of matrix $W \in \mathbb{R}^{d\times K}$ with $p \ge 1$ and $s \ge 1$. It is defined as

$$|W|_{p,s} = \max_{|\mathbf{u}|_p\le 1, |\mathbf{v}|_s\le 1} \langle \mathbf{u}, W\mathbf{v}\rangle \tag{8}$$

where $\mathbf{u} \in \mathbb{R}^d$, $\mathbf{v} \in \mathbb{R}^K$, and $|\mathbf{u}|_q$ and $|\mathbf{v}|_t$ are L_q and L_t norm of vector $\mathbf{u}$ and $\mathbf{v}$, respectively. Evidently, the dual norm of $|\cdot|_{p,s}$ is $|\cdot|_{q,t}$, with $p^{-1} + q^{-1} = 1$ and $s^{-1} + t^{-1} = 1$. The theorem below shows the regret bound for Algorithm 1.

Theorem 1. *Assume that for the sequence of examples, $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_T, y_T)$, we have, for all t, $\mathbf{x}_t \in \mathbb{R}^d$, $\|\mathbf{x}\|_p \le 1$ and the number of classes is K. Let*

$U = (\mathbf{u}_1, \ldots, \mathbf{u}_K) \in \mathbb{R}^{d\times K}$ be any matrix, and $\Phi^* : \mathbb{R}^{d\times K} \mapsto \mathbb{R}$ be a strictly convex function with constant ρ with respect to norm $|\cdot|_{p,s}$. The expectation of the number of mistakes made by by Algorithm 1, denoted by $\mathrm{E}[M]$, is bounded as follows

$$\mathrm{E}[M] \le \frac{1}{\eta\kappa} D_{\Phi^*}(U) + \frac{1}{\kappa}\sum_{t=1}^{T} \ell_t(U) + \frac{\eta\gamma T}{2\rho\kappa(1-\gamma)} + \gamma T$$

where

$$\kappa = 1 - \frac{\eta}{2\rho}\left\{1 - \gamma + \frac{\gamma}{K}\left(1 + \left[\frac{K}{\gamma}\right]^s\right)^{2/s}\right\}$$

Proof. We take the expectation of both sides of the equality in (7) with respect to $\widetilde{y}_t$, denoted by $\mathrm{E}_t[\cdot]$, and have

$$\begin{aligned}&\mathrm{E}_t\left[D_{\Phi^*}(U, W^{t-1}) - D_{\Phi^*}(U, W^t) + D_{\Phi^*}(W^{t-1}, W^t)\right]\\ &= \langle W^{t-1} - U, \eta\mathbf{x}_t\tau_t^\top\rangle\end{aligned}$$

We define $\hat{M}_t = [\widehat{y}_t \neq y_t]$. Since $\widehat{y}_t \neq y_t$ implies $\nabla\ell_t(W^{t-1}) = \mathbf{x}_t\tau_t^\top$, using the convexity of the loss function, we have

$$\begin{aligned}(\ell_t(W^{t-1}) - \ell_t(U))\hat{M}_t &\le \langle W^{t-1} - U, \nabla\ell_t(W^{t-1})\rangle\\ &= \langle W^{t-1} - U, \mathbf{x}_t\tau_t^\top\rangle\end{aligned}$$

We thus have

$$\begin{aligned}&\frac{1}{\eta}\mathrm{E}_t\left[D_{\Phi^*}(U, W^{t-1}) - D_{\Phi^*}(U, W^t) + D_{\Phi^*}(W^{t-1}, W^t)\right]\\ &= \langle W^{t-1} - U, \eta\mathbf{x}_t\tau_t^\top\rangle \ge (\ell_t(W^{t-1}) - \ell_t(U))\hat{M}_t\end{aligned}$$

By adding the above inequalities of all trials, we have

$$\begin{aligned}&\mathrm{E}\left[\sum_{t=1}^{T}\ell_t(W^{t-1})\right] - \left\{\frac{1}{\eta}D_{\Phi^*}(U) + \sum_{t=1}^{T}\ell_t(U)\right\}\\ &\le \frac{1}{\eta}\sum_{t=1}^{T}\mathrm{E}\left[D_{\Phi^*}(W^{t-1}, W^t)\right] = \sum_{t=1}^{T}\frac{1}{\eta}\mathrm{E}\left[D_{\Phi}(\theta^{t-1}, \theta^t)\right]\end{aligned}$$

The last step uses the property of Bregman distance in Lemma 1. Since Φ^* is a strictly convex function with constant ρ with respect to $\|\cdot\|_{p,s}$, according to Lemma 1, we have

$$D_\Phi(A, B) \le \frac{1}{2\rho}\|A - B\|_{q,t}^2$$

where $p^{-1} + q^{-1} = 1$ and $s^{-1} + t^{-1} = 1$. Hence,

$$\begin{aligned}\mathrm{E}[D_\Phi(\theta^{t-1}, \theta^t)] &\le \frac{\eta^2}{2\rho}\mathrm{E}[|\mathbf{x}_t\delta_t^\top|_{q,t}^2]\\ &\le \frac{\eta^2}{2\rho}\mathrm{E}[|\delta_t|_s^2|\mathbf{x}_t|_p^2] \le \frac{\eta^2}{2\rho}\mathrm{E}[|\delta_t|_s^2]\end{aligned}$$

where the second inequality is due to Holder's inequality. Using the result in Proposition 3 and the fact $\sum_{t=1}^{T}\ell_t(W^{t-1})\hat{M}_t \ge \sum_{t=1}^{T}\hat{M}_t$, and that $\mathrm{E}[M] \le \mathrm{E}[\hat{M}] + \gamma T$ we have the result in the theorem. □

Notice that the Banditron algorithm is a special case of the general framework with $\Phi^*(W) = \frac{1}{2}|W|_F^2$ and $|\cdot|_{p,s} = |\cdot|_{2,2} = |\cdot|_F$. The Banditron bound is specifically obtained through approximations $\gamma/(1-\gamma) \le 2\gamma$ and $1 + k/\gamma \le 2k/\gamma$ in summarizing the terms in κ.

3.3 Exponential Gradient for Online Multi-class Learning with Partial Feedback

In this section, we extend the exponent gradient algorithm to online multi-class learning with partial feedback. A straightforward approach is to use the result in Theorem 1 by setting

$$\Phi(\theta) = \sum_{k=1}^{K}\sum_{i=1}^{d}\exp(\theta_{i,k}) \quad (9)$$

$$\Phi^*(W) = \sum_{k=1}^{K}\sum_{i=1}^{d}W_{i,k}(\ln W_{i,k} - 1) \quad (10)$$

where each $\mathbf{w}_k$ is a probability distribution. Following the general framework presented in Algorithm 1, Algorithm 2 summarizes the exponential gradient algorithm for online multi-class learning with partial feedback. Since $\Phi^*(W)$ is strictly convex with constant 1 with respect to $|\cdot|_F$, we have following mistake bound for the exponential gradient algorithm.

Theorem 2. *Assume that for the sequence of examples, $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_T, y_T)$, we have, for all t, $\mathbf{x}_t \in \mathbb{R}^d$, $\|\mathbf{x}\|_2 \le 1$ and the number of classes is K. Let $U = (\mathbf{u}_1, \ldots, \mathbf{u}_K) \in \mathbb{R}^{d\times K}$ where each $\mathbf{u}_k$ is a distribution. The expectation of the number of mistakes made by by Algorithm 2 is bounded as follows*

$$\mathrm{E}[M] \le \frac{K\ln K}{\kappa\eta} + \frac{1}{\kappa}\sum_{t=1}^{T}\ell_t(U) + \frac{\eta\gamma T}{2\rho\kappa(1-\gamma)} + \gamma T$$

where $\kappa = 1 - \frac{\eta}{2\rho}\left(1 - \gamma + \frac{\gamma}{K} + \frac{K}{\gamma}\right)$.

By minimizing the mistake bound in the above theorem, we choose step size η as follows

$$\eta = \sqrt{\frac{K(1-\gamma)\ln K}{T\gamma}} \quad (11)$$

For the high dimensional data, we can improve the result in Theorem 2 by using the following lemma.

Lemma 2. *$\Phi(W)$ and $\Phi^*(W)$ defined in (9) and (10) satisfies the following properties*

$$\begin{aligned}\langle W - W', \nabla\Phi^*(W) - \nabla\Phi^*(W')\rangle &\ge \sum_{k=1}^{K}|\mathbf{w}_k - \mathbf{w}'_k|_1^2\\ \langle\theta - \theta', \nabla\Phi(\theta) - \nabla\Phi^*(\theta')\rangle &\le \sum_{k=1}^{K}|\theta_{*,k} - \theta'_{*,k}|_\infty^2\end{aligned}$$

where $\theta_{*,k} = (\theta_{1,k}, \ldots, \theta_{d,k})$.

Algorithm 2 Exponential Gradient Algorithm for Online Multi-class Learning with Partial Feedback

1: Parameters:
- Smoothing parameter: $\gamma \in (0, 0.5)$
- Step size: $\eta > 0$

2: Set $\theta^0 = \mathbf{1}\mathbf{1}^\top / d$
3: **for** $t = 1, \ldots, T$ **do**
4: Compute $W^t_{i,k} = \exp(\theta^t_{i,k})/Z^t_k$ where $Z^t_k = \sum_{i=1}^d \exp(\theta^t_{i,k})$.
5: Receive $\mathbf{x}_t \in \mathbb{R}^d$
6: Compute

$$\widehat{y}_t = \arg\max_{1 \le k \le K} \mathbf{x}_t^\top \mathbf{w}_k^{t-1} \tag{12}$$

7: Set $p_k = (1-\gamma)[k = \widehat{y}_t] + \gamma/K, k = 1, \ldots, K$
8: Randomly sample $\widetilde{y}_t$ according to the distribution $\mathbf{p} = (p_1, \ldots, p_K)$.
9: Predict $\widetilde{y}_t$ and receive feedback $[y_t = \widetilde{y}_t]$
10: Compute

$$\delta_t = \mathbf{1}_{\widehat{y}_t} - \mathbf{1}_{\widetilde{y}_t} \frac{[y_t = \widetilde{y}_t]}{p_{\widehat{y}_t}} \tag{13}$$

where $\mathbf{1}_k$ stands for the vector of all elements being zero except that its kth element is 1.
11: Compute $\theta^t = \theta^{t-1} - \eta \mathbf{x}_t \delta_t^\top$
12: **end for**

Proof.

$$\begin{aligned}
& \langle W - W', \nabla\Phi^*(W) - \nabla\Phi^*(W') \rangle \\
= & \sum_{k=1}^K \sum_{i=1}^d (W_{i,k} - W'_{i,k})(\ln W_{i,k} - \ln W'_{i,k}) \\
= & \sum_{k=1}^K \sum_{i=1}^d \frac{(W_{i,k} - W'_{i,k})^2}{\overline{W}_{i,k}} \\
\ge & \sum_{k=1}^K \frac{\left(\sum_{i=1}^d |W_{i,k} - W'_{i,k}|\right)^2}{\sum_{i=1}^d \overline{W}_{i,k}}
\end{aligned}$$

The second equality is due to mean value theorem and uses the Taylor expansion of log function where $\overline{W} = \lambda W + (1-\lambda)W'$ with $\lambda \in [0, 1]$. Since

$$\frac{\left(\sum_{i=1}^d |W_{i,k} - W'_{i,k}|\right)^2}{\sum_{i=1}^d \overline{W}_{i,k}} \sum_{i=1}^d \overline{W}_{i,k} \ge \left(\sum_{i=1}^d |W_{i,k} - W'_{i,k}|\right)^2$$

and $\sum_{i=1}^d \overline{W}_{i,k} = 1$, we have

$$\langle W - W', \nabla\Phi^*(W) - \nabla\Phi^*(W') \rangle \ge \sum_{k=1}^K |\mathbf{w}_k - \mathbf{w}'_k|_1^2$$

Using the property of Bregman distance in Lemma 1 and the fact the dual norm of L_1 is L_∞, we have the result for $\Phi(\theta)$. □

Using the above lemma, we have the following theorem that updates the result in Theorem 2

Theorem 3. *Same as the setup of Theorem 2 except that $|\mathbf{x}_i|_\infty \le 1$. The expectation of the number of mistakes made by by Algorithm 2 is bounded as follows*

$$\mathrm{E}[M] \le \frac{K \ln K}{\kappa\eta} + \frac{1}{\kappa} \sum_{t=1}^T \ell_t(U) + \frac{\eta\gamma T}{2\rho\kappa(1-\gamma)} + \gamma T$$

where $\kappa = 1 - \frac{\eta}{2\rho}\left(2 - 2\gamma - \frac{4\gamma}{K}\right)$.

Proof. The proof is the same as the proof of Theorem 1 except that we have

$$\mathrm{E}[D_\Phi(\theta^{t-1}, \theta^t)] \le \frac{\eta^2}{2\rho} \mathrm{E}\left[\sum_{k=1}^K |\delta_{t,k}||\mathbf{x}_t|_\infty^2\right] \le \frac{\eta^2}{2\rho} \mathrm{E}[|\delta_t|_1]$$

A simple computation shows that $\mathrm{E}[|\delta_t|_1] = 2 - 2\gamma - 4\gamma/K$. By combining these results, we have the theorem. □

The major difference between Theorem 2 and 3 is the constraint on $\mathbf{x}$: L_2 is used in Theorem 2 and L_∞ is used in Theorem 3. Therefore, Theorem 3 shows that the exponential gradient algorithm is essentially independent from dimensionality d, making it suitable for handling high dimensional data.

4 Experiments

To study the performance of the proposed framework, we follow the experimental setup in (Kakade et al., 2008) and test the exponential gradient algorithm for partial feedback on both synthesized and real-world data sets. For easy reference, we refer to our algorithm as **exp_grad**. We compare **exp_grad** to Banditron (Kakade et al., 2008), a state-of-the-art approach for online multi-class learning with partial feedback.

4.1 Data set descriptions

The following data sets are used in our study:

Synthetic data sets: Following (Kakade et al., 2008), we create two synthesized data sets, i.e., *separable synthetic data set* and *non-separable synthetic data set*.

- *separable synthetic data set.* This data set was used to simulate text categorization and consists of 10,000 synthesized documents, with each document represented by a binary vector of 400 dimensions. To simulate the documents, we first generate 9 exemplar documents, and each exemplar document is viewed as a profile of a different topic. For each exemplar document, we randomly

turn on 20 to 40 bits in its first 120 dimensions. The synthesized documents are generated by using these exemplars as follows: we first randomly select one exemplar document and randomly turn off 5 bits in its first 120 dimensions and turn on 20 bits in the last 280 dimensions which serve as noisy factors. All the synthesized documents generated by the same exemplar document are assigned to the same class.

- *Non-separable synthetic data set.* The synthetic data set introduced above is linearly separable. We furthermore introduced 5% label noise to this data set to makes the documents of different classes non-separable. In particular, for each generated sample, we change its original label to one of the other eight classes with probability 5%.

Real-world data sets: The two real-world data sets are *RCV1 document collection for text categorization* and *MNIST data set for hand written digits*:

- *RCV1 document collection for text categorization.* The RCV1 collection consists of $804,414$ documents. Each document is a news report from Reuters (Lewis et al., 2004). The vocabulary size of RCV1 is $47,236$, implying each document is represented by a vector of $47,236$ dimension. Since the focus of our study is online multi-class learning, we will consider the documents with a single class assignment. In particular, we follow (Kakade et al., 2008) and only use the documents in the following Reuters topic categories: CCAT, ECAT, GCAT and MCAT, which leads to $665,265$ documents.
- *MNIST data set for hand written digits.* It is consisted of images of hand written digits. All the digit images have been normalized into the size 28×28. Each image is represented by gray scale of its pixels, leading to a vector representation of 784 dimension. The original data set contains $60,000$ training samples and $10,000$ test samples. Only the training images are used in this study(`http://yann.lecun.com/exdb/mnist/`).

4.2 Experimental settings

We compared the classification performance of the proposed exponential gradient algorithm to the Banditron algorithm on the aforementioned four data sets. For each tested algorithm and for each data set, we conduct 10 independent runs with different seeds for randomization. We evaluate the performance of online learning by the accumulate error rate, which is computed as the ratio of the number of misclassified samples and the number of samples received so far during the online learning process.

Since the exponential gradient algorithm assumes all the combination weights to be non-negative, in order to make fair comparison between the proposed approach and the Banditron algorithm, we run two sets of experiments as follows.

- **Setup 1:** *Restricted Banditron.* In this experiment, we compare basic exponential method with a modified version of Banditron, obtained as follows: in each iteration of Banditron, we project the learned weights into the positive orthants, which is equivalent to setting all the negative weights to be zero. It is easy to verify that the projection step does not change the theoretic properties of Banditron, in particular the mistake bound (of course only with respect to linear classifiers U in the positive orthants).
- **Setup 2:** *Balanced Exponential Algorithm.* In this experiment, we compare the the original Banditron, with an adapted version of exponential algorithm that uses each feature twice, one with positive sign and one with negative sign.

Since all the online learning algorithms rely on the parameter γ to control the tradeoff between exploration and exploitation, we examine the classification results of all the algorithms in comparison by varying γ. The step size η of online learning often play an important role in the final performance. For the proposed algorithm, we set the step size according to Eq. 11. Because the exponential function may exceed the upper bound of a real number with double precision type in a 64-bit computer, we further multiple the step size with a small factor (typically 10^{-5}) to avoid this issue.

In each experiment, we include the classification performance of online learning with full feedback using the Perceptron algorithm as a reference point.

4.3 Experimental results

Fig. 1 shows the classification performance of the two online multi-class learning with partial feedback on four data sets. The first column in this figure compares the average error rates of the online algorithms with varied γ values, and the second column shows the average error rates of the three online methods over the entire online process for the case of Setup 1 (only positive weights). The last column shows the average error rates of different methods for the case of Setup 2 (both positive and negative weights). For both Banditron and the proposed algorithm in the second and third column, we choose the optimal γ that results the lowest classification error rate for Banditron.

First, by examining the classification performance with varied γ, we clearly see that the exponential gradient algorithm shows comparable performance compared with the Banditron algorithm for online multi-class learning with limited feedback. In particular, we

observe that the proposed algorithm performs significantly better than the Banditron algorithm when γ is small. The difference between the online learning algorithms is outstanding particularly for the two synthetic data sets. The result indicates that the proposed algorithm is overall more reliable than the Banditron algorithm in terms of γ. Notice that for all data sets except for separable data set, we observe a significant gap between online learning with full feedback and online learning with partial feedback, which is due to the limited feedback from the oracle.

Second, we compare the learning rate of all three algorithms. We observe that the proposed algorithm overall exhibits a significantly better learning rate than the Banditron algorithm for Setup 1; i.e., for all four datasets, for most part of the online learning process, the proposed algorithm yield significantly lower classification error rates than the Banditron algorithm. Notice that the proposed algorithm is more suitable for the scenario when the number of trials is relatively small. For the case of Setup 2, the result of the proposed algorithm overall exhibits comparable learning rate compared with the Banditron algorithm, however is not as effective as the Setup 1 which is still an open question. This result indicates that the proposed online learning algorithm with partial feedback is generally effective in reducing the error rate.Similar to the overall classification error, we observe a significant gap in the learning rate between online learning with full feedback and online learning with partial feedback.

5 Conclusion

We present a general framework for online multi-class learning with partial feedback using the potential-based gradient descent approach. In addition, we propose an exponential gradient algorithm for online multi-class learning with partial feedback. Compared to the Banditron algorithm, the exponential gradient algorithm is advantageous in that its mistake bound is independent from the dimension of data, making it suitable for classifying high dimensional data. We verified the efficacy of the proposed algorithm by empirical studies with both synthetic and real-world data sets. Our experiments show the exponential gradient approach for online learning with partial feedback is more effective than Banditron in the choice of parameter γ and the learning rate, which makes it more suitable for the scenario when the number of training examples is relatively small. In future, we plan to examine the issue of how to automatically determine the optimal value for γ, a key parameter that controls the tradeoff between exploration and exploitation.

6 Acknowledgment

The work was supported in part by National Science Foundation (IIS-0643494) and Yahoo! Labs gift. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF and Yahoo! Labs.

References

Auer, P., Cesa-Bianchi, N., Freund, Y., & Schapire, R. E. (2003). The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, **32**(1), 48–77.

Cesa-Bianchi, N. & Lugosi, G. (2006). *Prediction, Learning, and Games*. Cambridge University Press.

Crammer, K., Singer, Y., & Warmuth, K. (2003). Ultraconservative online algorithms for multiclass problems. *JMLR*, **3**, 2003.

Even-Dar, E., Mannor, S., & Mansour, Y. (2002). Pac bounds for multi-armed bandit and markov decision processes. In *COLT '02*, pages 255–270, London, UK. Springer-Verlag.

Even-Dar, E., Mannor, S., & Mansour, Y. (2006). Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *JMLR*, **7**, 1079–1105.

Grove, A. J., Littlestone, N., & Schuurmans, D. (2001). General convergence results for linear discriminant updates. *Mach. Learn.*, **43**(3), 173–210.

Kakade, S. M., Shalev-Shwartz, S., & Tewari, A. (2008). Efficient bandit algorithms for online multiclass prediction. In *ICML '08*, pages 440–447, New York, NY, USA. ACM.

Kivinen, J. & Warmuth, M. K. (1995). Additive versus exponentiated gradient updates for linear prediction. In *STOC '95*, pages 209–218, New York, NY, USA. ACM.

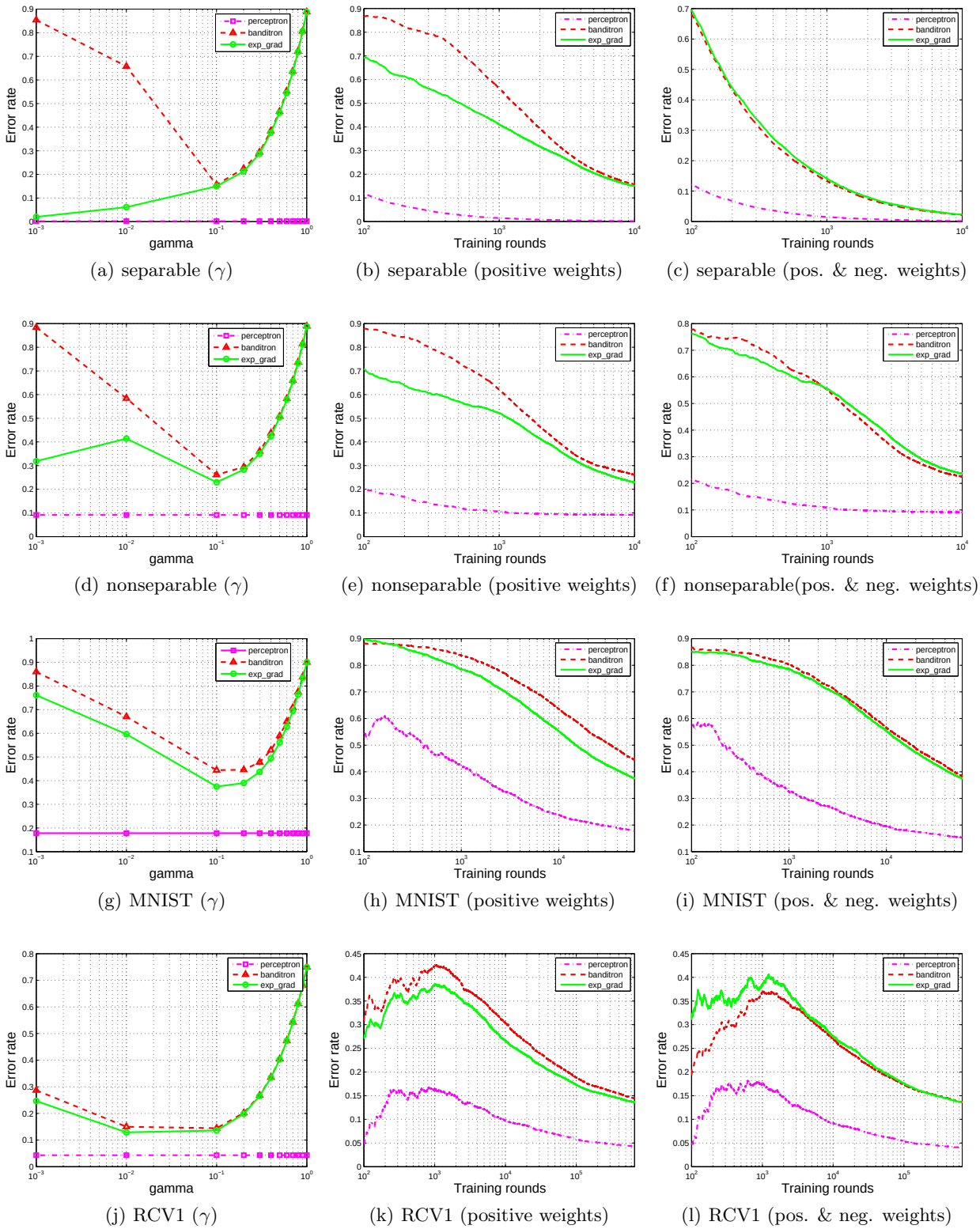
Langford, J. & Tong, Z. (2007). The epoch-greedy algorithm for contextual multi-armed bandits. In *NIPS '07*.

Lewis, D. D., Yang, Y., Rose, T., & Li, F. (2004). Rcv1: A new benchmark collection for text categorization research. *JMLR*, **5**, 361–397.

Littlestone, N. (1988). Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Mach. Learn.*, **2**(4), 285–318.

Mannor, S. & Tsitsiklis, J. N. (2004). The sample complexity of exploration in multi-armed bandit problem. *JMLR*, **5**, 623–648.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, **65**, 386–408.



(a) separable (γ) (b) separable (positive weights) (c) separable (pos. & neg. weights)

(d) nonseparable (γ) (e) nonseparable (positive weights) (f) nonseparable(pos. & neg. weights)

(g) MNIST (γ) (h) MNIST (positive weights) (i) MNIST (pos. & neg. weights)

(j) RCV1 (γ) (k) RCV1 (positive weights) (l) RCV1 (pos. & neg. weights)

Figure 1: Comparisons of the three methods with positive weights on the four data sets. **Left column** shows the error rates of Perceptron (with full feedback), Banditron and the proposed exponential gradient algorithm with varied γ, **Middle column** shows the error rates of the three methods over trials for Banditron and the proposed algorithm when the weights are limited to be positive, and **Right column** shows the error rates of the three methods over trials for Banditron and the proposed algorithm when there is no restriction on the weights. Each point on a curve is the average results of 10 random tests.