Bandit Multiclass Linear Classification: Efficient Algorithms for the Separable Case

Alina Beygelzimer[*1] **Dávid Pál**[*1] **Balázs Szörényi**[*1] **Devanathan Thiruvenkatachari**[*2] **Chen-Yu Wei**[*3] **Chicheng Zhang**[*4]

Abstract

We study the problem of efficient online multiclass linear classification with bandit feedback, where all examples belong to one of K classes and lie in the d-dimensional Euclidean space. Previous works have left open the challenge of designing efficient algorithms with finite mistake bounds when the data is linearly separable by a margin γ. In this work, we take a first step towards this problem. We consider two notions of linear separability, *strong* and *weak*.

1. Under the strong linear separability condition, we design an efficient algorithm that achieves a near-optimal mistake bound of $O\left(K/\gamma^2\right)$.
2. Under the more challenging weak linear separability condition, we design an efficient algorithm with a mistake bound of $\min(2^{\widetilde{O}(K\log^2(1/\gamma))}, 2^{\widetilde{O}(\sqrt{1/\gamma}\log K)})$.[1] Our algorithm is based on kernel Perceptron and is inspired by the work of Klivans & Servedio (2008) on improperly learning intersection of halfspaces.

1. Introduction

We study the problem of ONLINE MULTICLASS LINEAR CLASSIFICATION WITH BANDIT FEEDBACK (Kakade et al., 2008). The problem can be viewed as a repeated game between a learner and an adversary. At each time step t, the adversary chooses a labeled example (x_t, y_t) and reveals the feature vector x_t to the learner. Upon receiving x_t, the learner makes a prediction $\widehat{y}_t$ and receives feedback. In contrast with the standard full-information setting, where the feedback given is the correct label y_t, here the feedback is only a binary indicator of whether the prediction was correct or not. The protocol of the problem is formally stated below.

Protocol 1 ONLINE MULTICLASS LINEAR CLASSIFICATION WITH BANDIT FEEDBACK

Require: Number of classes K, number of rounds T.
Require: Inner product space $(V, \langle\cdot,\cdot\rangle)$.
for $t = 1, 2, \ldots, T$ **do**
  Adversary chooses example $(x_t, y_t) \in V \times \{1, 2, \ldots, K\}$ where x_t is revealed to the learner.
  Predict class label $\widehat{y}_t \in \{1, 2, \ldots, K\}$.
  Observe feedback $z_t = \mathbb{1}\left[\widehat{y}_t \neq y_t\right] \in \{0, 1\}$.

The performance of the learner is measured by its cumulative number of mistakes $\sum_{t=1}^{T} z_t = \sum_{t=1}^{T} \mathbb{1}\left[\widehat{y}_t \neq y_t\right]$, where $\mathbb{1}$ denotes the indicator function.

In this paper, we focus on the special case when the examples chosen by the adversary lie in $\mathbb{R}^d$ and are linearly separable with a margin. We introduce two notions of linear separability, *weak* and *strong*, formally stated in Definition 1. The standard notion of multiclass linear separability (Crammer & Singer, 2003) corresponds to the weak linear separability. For multiclass classification with K classes, weak linear separability requires that all examples from the same class lie in an intersection of $K-1$ halfspaces and all other examples lie in the complement of the intersection of the halfspaces. Strong linear separability means that examples from each class are separated from the remaining examples by a *single* hyperplane.

In the full-information feedback setting, it is well known (Crammer & Singer, 2003) that if all examples have norm at most R and are weakly linearly separable with a margin γ, then the MULTICLASS PERCEPTRON algorithm makes at most $\lfloor 2(R/\gamma)^2 \rfloor$ mistakes. It is also known that any (possibly randomized) algorithm must make $\frac{1}{2}\left\lfloor (R/\gamma)^2 \right\rfloor$ mistakes in the worst case. The MULTICLASS PERCEPTRON achieves an information-theoretically optimal mis-

[*]The authors are listed in alphabetical order. [1]Yahoo Research, New York, NY, USA [2]New York University, New York, NY, USA [3]University of Southern California, Los Angeles, CA, USA [4]Microsoft Research, New York, NY, USA. Correspondence to: Dávid Pál <davidko.pal@gmail.com>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

[1]We use the notation $\widetilde{O}(f(\cdot)) = O(f(\cdot)\,\mathrm{polylog}(f(\cdot)))$.

take bound, while being time and memory efficient.[2] [3]

The bandit feedback setting, however, is much more challenging. For the strongly linearly separable case, we are not aware of any prior efficient algorithm with a finite mistake bound. [4] We design a simple and efficient algorithm (Algorithm 1) that makes at most $O(K(R/\gamma)^2)$ mistakes in expectation. Its memory complexity and per-round time complexity are both $O(dK)$. The algorithm can be viewed as running K copies of the BINARY PERCEPTRON algorithm, one copy for each class. We prove that any (possibly randomized) algorithm must make $\Omega(K(R/\gamma)^2)$ mistakes in the worst case. The extra $O(K)$ multiplicative factor in the mistake bound, as compared to the full-information setting, is the price we pay for the bandit feedback, or more precisely, the lack of full-information feedback.

For the case when the examples are weakly linearly separable, it was open for a long time whether there exist *efficient* algorithms with finite mistake bound (Kakade et al., 2008; Beygelzimer et al., 2017). Furthermore, Kakade et al. (2008) ask the question: Is there *any* algorithm with a finite mistake bound that has no explicit dependence on the dimensionality of the feature vectors? We answer both questions affirmatively by providing an efficient algorithm with finite dimensionless mistake bound (Algorithm 2).[5]

The strategy used in Algorithm 2 is to construct a non-linear feature mapping ϕ and associated positive definite kernel $k(x, x')$ that makes the examples *strongly* linearly separable in a higher-dimensional space. We then use the kernelized version of Algorithm 1 for the strongly separable case. The kernel $k(x, x')$ corresponding to the feature mapping ϕ has a simple explicit formula and can be computed in $O(d)$ time, making Algorithm 2 computationally efficient. For details on kernel methods see e.g. (Schölkopf & Smola, 2002) or (Shawe-Taylor & Cristianini, 2004).

The number of mistakes of the kernelized algorithm depends on the margin in the corresponding feature space. We analyze how the mapping ϕ transforms the margin parameter of weak separability in the original space $\mathbb{R}^d$ into a margin parameter of strong separability in the new feature space. This problem is related to the problem of learning intersection of halfspaces and has been studied previously by Klivans & Servedio (2008). As a side result, we improve on the results of Klivans & Servedio (2008) by removing the dependency on the original dimension d.

The resulting kernelized algorithm runs in time polynomial in the original dimension of the feature vectors d, the number of classes K, and the number of rounds T. We prove that if the examples lie in the unit ball of $\mathbb{R}^d$ and are weakly linearly separable with margin γ, Algorithm 2 makes at most $\min(2^{\widetilde{O}(K \log^2(1/\gamma))}, 2^{\widetilde{O}(\sqrt{1/\gamma} \log K)})$ mistakes.

In Appendix G, we propose and analyze a very different algorithm for weakly linearly separable data. The algorithm is based on the obvious idea that two points that are close enough must have the same label.

Finally, we study two questions related to the computational and information-theoretic hardness of the problem. Any algorithm for the bandit setting collects information in the form of so called *strongly labeled* and *weakly labeled* examples. Strongly labeled examples are those for which we know the class label. Weakly labeled example is an example for which we know that class label can be anything except for one particular class. In Appendix H, we show that the offline problem of finding a multiclass linear classifier consistent with a set of strongly and weakly labeled examples is NP-hard. In Appendix I, we prove a lower bound on the number of mistakes of any algorithm that uses only strongly-labeled examples and ignores weakly labeled examples.

2. Related work

The problem of online bandit multiclass learning was initially formulated in the pioneering work of Auer & Long (1999) under the name of "weak reinforcement model". They showed that if all examples agree with some classifier from a prespecified hypothesis class $\mathcal{H}$, then the optimal mistake bound in the bandit setting can be upper bounded by the optimal mistake bound in the full information setting, times a factor of $(2.01 + o(1))K \ln K$. Long (2017) later improved the factor to $(1 + o(1))K \ln K$ and showed its near-optimality. Daniely & Helbertal (2013) extended the results to the setting where the performance of the algorithm is measured by its regret, i.e. the difference between the number of mistakes made by the algorithm and the number of mistakes made by the best classifier in $\mathcal{H}$ in hindsight. We remark that all algorithms developed in this context are computationally inefficient.

The linear classification version of this problem is initially studied by Kakade et al. (2008). They proposed two computationally inefficient algorithms that work in the weakly linearly separable setting, one with a mistake bound of $O(K^2 d \ln(d/\gamma))$, the other with a mistake bound of

[2] We call an algorithm computationally efficient, if its running time is polynomial in K, d, $1/\gamma$ and T.

[3] For completeness, we present these folklore results along with their proofs in Appendix A in the supplementary material.

[4] Although Chen et al. (2009) claimed that their Conservative OVA algorithm with PA-I update has a finite mistake bound under the strong linear separability condition, their Theorem 2 is incorrect: first, their Lemma 1 (with $C = +\infty$) along with their Theorem 1 implies a mistake upper bound of $(\frac{R}{\gamma})^2$, which contradicts the lower bound in our Theorem 3; second, their Lemma 1 cannot be directly applied to the bandit feedback setting.

[5] An inefficient algorithm was given by (Daniely & Helbertal, 2013).

$\widetilde{O}((K^2/\gamma^2)\ln T)$. The latter result was later improved by Daniely & Helbertal (2013), which gives a computationally inefficient algorithm with a mistake upper bound of $\widetilde{O}(K/\gamma^2)$. In addition, Kakade et al. (2008) propose the BANDITRON algorithm, a computationally efficient algorithm that has a $O(T^{2/3})$ regret against the multiclass hinge loss in the general setting, and has a $O(\sqrt{T})$ mistake bound in the γ-weakly linearly separable setting. In contrast to mild dependencies on the time horizon for mistake bounds of computationally inefficient algorithms, the polynomial dependence of BANDITRON's mistake bound on the time horizon is undesirable for problems with a long time horizon, in the weakly linearly separable setting. One key open question left by Kakade et al. (2008) is whether one can design computationally efficient algorithms that achieve mistake bounds that match or improve over those of inefficient algorithms. In this paper, we take a step towards answering this question, showing that efficient algorithms with mistake bounds quasipolynomial in $1/\gamma$ (for constant K) and quasipolynomial in K (for constant γ) can be obtained.

The general problem of linear bandit multiclass learning has received considerable attention (Abernethy & Rakhlin, 2009; Wang et al., 2010; Crammer & Gentile, 2013; Hazan & Kale, 2011; Beygelzimer et al., 2017; Foster et al., 2018). Chen et al. (2014); Zhang et al. (2018) study online bandit multiclass boosting under bandit feedback, where one can view boosting as linear classification by treating each base hypothesis as a separate feature. In the weakly linearly separable setting, however, these algorithms can only guarantee a mistake bound of $O(\sqrt{T})$ at best.

The problem considered here is a special case of the contextual bandit problem (Auer et al., 2003; Langford & Zhang, 2008). In this general problem, there is a hidden cost vector c_t associated with every prediction in round t. Upon receiving x_t and predicting $\widehat{y}_t \in \{1,\ldots,K\}$, the learner gets to observe the incurred cost $c_t(\widehat{y}_t)$. The goal of the learner is to minimize its regret with respect to the best predictor in some predefined policy class Π, given by $\sum_{t=1}^T c_t(\widehat{y}_t) - \min_{\pi\in\Pi}\sum_{t=1}^T c_t(\pi(x_t))$. Bandit multiclass learning is a special case where the cost $c_t(i)$ is the classification error $\mathbb{1}[i \neq y_t]$ and the policy class is the set of linear classifiers $\left\{x \mapsto \operatorname{argmax}_y (Wx)_y : W \in \mathbb{R}^{K\times d}\right\}$. There has been significant progress on the general contextual bandit problem assuming access to an optimization oracle that returns a policy in Π with the smallest total cost on any given set of cost-sensitive examples (Dudík et al., 2011; Agarwal et al., 2014; Rakhlin & Sridharan, 2016; Syrgkanis et al., 2016a;b). However, such an oracle abstracting efficient search through Π is generally not available in our setting due to computational hardness results (Arora et al., 1997).

Recently, Foster & Krishnamurthy (2018) developed a rich theory of contextual bandits with surrogate losses, focusing on regrets of the form $\sum_{t=1}^T c_t(\widehat{y}_t) - \min_{f\in\mathcal{F}}\sum_{t=1}^T \frac{1}{K}\sum_{i=1}^K c_t(i)\phi(f_i(x_t))$, where $\mathcal{F}$ contains score functions $f = (f_1,\ldots,f_K)$ such that $\sum_{i=1}^K f_i(\cdot) \equiv 0$, and $\phi(s) = \max(1-\frac{s}{\gamma},0)$ or $\min(1,\max(1-\frac{s}{\gamma},0))$. On one hand, it gives information-theoretic regret upper bounds for various settings of $\mathcal{F}$. On the other hand, it gives an efficient algorithm with an $O(\sqrt{T})$ regret against the benchmark of $\mathcal{F} = \left\{x \mapsto Wx : W \in \mathbb{R}^{K\times d}, \mathbb{1}^T W = 0\right\}$. A direct application of this result to ONLINE BANDIT MULTICLASS LINEAR CLASSIFICATION gives an algorithm with $O(\sqrt{T})$ mistake bound in the strongly linearly separable case.

3. Notions of linear separability

Let $[n] = \{1,2,\ldots,n\}$. We define two notions of linear separability for multiclass classification. The first notion is the standard notion of linear separability used in the proof of the mistake bound for the MULTICLASS PERCEPTRON algorithm (see e.g. Crammer & Singer, 2003). The second notion is stronger, i.e. more restrictive.

Definition 1 (Linear separability). *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space, K be a positive integer, and γ be a positive real number. We say that labeled examples $(x_1,y_1),(x_2,y_2),\ldots,(x_T,y_T) \in V\times[K]$ are*

weakly linearly separable with a margin γ *if there exist vectors $w_1,w_2,\ldots,w_K \in V$ such that*

$$\sum_{i=1}^K \|w_i\|^2 \leq 1, \tag{1}$$

$$\langle x_t, w_{y_t}\rangle \geq \langle x_t, w_i\rangle + \gamma \quad \forall t\in[T]\ \forall i \in [K]\setminus\{y_t\}, \tag{2}$$

and strongly linearly separable with a margin γ *if there exist vectors $w_1,w_2,\ldots,w_K \in V$ such that*

$$\sum_{i=1}^K \|w_i\|^2 \leq 1, \tag{3}$$

$$\langle x_t, w_{y_t}\rangle \geq \gamma/2 \quad \forall t\in[T], \tag{4}$$

$$\langle x_t, w_i\rangle \leq -\gamma/2 \quad \forall t\in[T]\ \forall i\in[K]\setminus\{y_t\}. \tag{5}$$

The notion of strong linear separability has appeared in the literature; see e.g. (Chen et al., 2009). Intuitively, strong linear separability means that, for each class i, the set of examples belonging to class i and the set of examples belonging to the remaining $K-1$ classes are separated by a linear classifier w_i with margin $\frac{\gamma}{2}$.

It is easy to see that if a set of labeled examples is strongly linearly separable with margin γ, then it is also weakly linearly separable with the same margin (or larger). Indeed, if

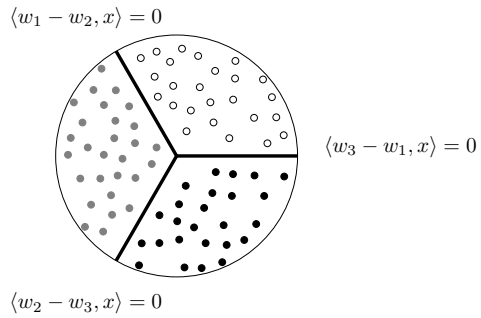


Figure 1. A set of labeled examples in $\mathbb{R}^2$. The examples belong to $K = 3$ classes colored white, gray and black respectively. Each class lies in a 120° wedge. In other words, each class lies in an intersection of two halfspaces. While the examples are weakly linearly separable with a positive margin γ, they are *not* strongly linearly separable with any positive margin γ. For instance, there does *not* exist a linear separator that separates the examples belonging to the gray class from the examples belonging to the remaining two classes.

$w_1, w_2, \ldots, w_K \in V$ satisfy (3), (4), (5) then they satisfy (1) and (2).

In the special case of $K = 2$, if a set of labeled examples is weakly linearly separable with a margin γ, then it is also strongly linearly separable with the same margin. Indeed, if w_1, w_2 satisfy (1) and (2) then $w_1' = \frac{w_1 - w_2}{2}$, $w_2' = \frac{w_2 - w_1}{2}$ satisfy (3), (4), (5). Equation (3) follows from $\left\|w_i'\right\|^2 \leq (\frac{1}{2}\|w_1\| + \frac{1}{2}\|w_2\|)^2 \leq \frac{1}{2}\|w_1\|^2 + \frac{1}{2}\|w_2\|^2 \leq \frac{1}{2}$ for $i = 1, 2$. Equations (4) and (5) follow from the fact that $w_1' - w_2' = w_1 - w_2$.

However, for any $K \geq 3$ and any inner product space of dimension at least 2, there exists a set of labeled examples that is weakly linearly separable with a positive margin γ but is not strongly linearly separable with any positive margin. Figure 1 shows one such set of labeled examples.

4. Algorithm for strongly linearly separable data

In this section, we consider the case when the examples are strongly linearly separable. We present an algorithm for this setting (Algorithm 1) and give an upper bound on its number of mistakes, stated as Theorem 2 below. The proof of the theorem can be found in Appendix B.

The idea behind Algorithm 1 is to use K copies of the BINARY PERCEPTRON algorithm, one copy per class; see e.g. (Shalev-Shwartz, 2012, Section 3.3.1). Upon seeing each example x_t, copy i predicts whether or not x_t belongs to class i. Multiclass predictions are done by evaluating all K binary predictors and outputting any class with a positive prediction. If all binary predictions are negative, the algorithm chooses a prediction uniformly at random from $\{1, 2, \ldots, K\}$.

Algorithm 1 BANDIT ALGORITHM FOR STRONGLY LINEARLY SEPARABLE EXAMPLES

Require: Number of classes K, number of rounds T.
Require: Inner product space $(V, \langle\cdot,\cdot\rangle)$.
Initialize $w_1^{(1)} = w_2^{(1)} = \cdots = w_K^{(1)} = 0$
for $t = 1, 2, \ldots, T$ **do**
  Observe feature vector $x_t \in V$
  Compute $S_t = \left\{ i \,:\, 1 \leq i \leq K,\ \left\langle w_i^{(t)}, x_t \right\rangle \geq 0 \right\}$
  if $S_t = \emptyset$ **then**
    Predict $\widehat{y}_t \sim \text{Uniform}(\{1, 2, \ldots, K\})$
    Observe feedback $z_t = \mathbb{1}\left[\widehat{y}_t \neq y_t\right]$
    if $z_t = 1$ **then**
      Set $w_i^{(t+1)} = w_i^{(t)}, \forall i \in \{1, 2, \ldots, K\}$
    else
      Set $w_i^{(t+1)} = w_i^{(t)}, \forall i \in \{1, 2, \ldots, K\} \setminus \{\widehat{y}_t\}$
      Update $w_{\widehat{y}_t}^{(t+1)} = w_{\widehat{y}_t}^{(t)} + x_t$
  else
    Predict $\widehat{y}_t \in S_t$ chosen arbitrarily
    Observe feedback $z_t = \mathbb{1}\left[\widehat{y}_t \neq y_t\right]$
    if $z_t = 1$ **then**
      Set $w_i^{(t+1)} = w_i^{(t)}, \forall i \in \{1, 2, \ldots, K\} \setminus \{\widehat{y}_t\}$
      Update $w_{\widehat{y}_t}^{(t+1)} = w_{\widehat{y}_t}^{(t)} - x_t$
    else
      Set $w_i^{(t+1)} = w_i^{(t)}, \forall i \in \{1, 2, \ldots, K\}$

Theorem 2 (Mistake upper bound)**.** *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space, K be a positive integer, γ be a positive real number, R be a non-negative real number. If the examples $(x_1, y_1), \ldots, (x_T, y_T) \in V \times \{1, 2, \ldots, K\}$ are strongly linearly separable with margin γ and $\|x_1\|, \|x_2\|, \ldots, \|x_T\| \leq R$ then the expected number of mistakes that Algorithm 1 makes is at most $(K-1)\lfloor 4(R/\gamma)^2 \rfloor$.*

The upper bound $(K-1)\lfloor 4(R/\gamma)^2 \rfloor$ on the expected number of mistakes of Algorithm 1 is optimal up to a constant factor, as long as the number of classes K is at most $O((R/\gamma)^2)$. This lower bound is stated as Theorem 3 below. The proof of the theorem can be found in Appendix B. Daniely & Helbertal (2013) provide a lower bound under the assumption of weak linear separability, which does not immediately imply a lower bound under the stronger notion.

Theorem 3 (Mistake lower bound)**.** *Let γ be a positive real number, R be a non-negative real number and let $K \leq (R/\gamma)^2$ be a positive integer. Any (possibly randomized) algorithm makes at least $((K-1)/2)\left\lfloor (R/\gamma)^2/4 \right\rfloor$ mistakes in expectation on some sequence of labeled examples $(x_1, y_1), (x_2, y_2), \ldots, (x_T, y_T) \in V \times \{1, 2, \ldots, K\}$*

for some inner product space $(V, \langle\cdot,\cdot\rangle)$ such that the examples are strongly linearly separable with margin γ and satisfy $\|x_1\|, \|x_2\|, \ldots, \|x_T\| \leq R$.

Remark. If $\gamma \leq R$ then, irrespective of any other conditions on K, R, and γ, a trivial lower bound on the expected number of mistakes of any randomized algorithm is $(K-1)/2$. To see this, note that the adversary can choose an example (Re_1, y), where e_1 is some arbitrary unit vector in V and y is a label chosen uniformly from $\{1, 2, \ldots, K\}$, and show this example K times. The sequence of examples trivially satisfies the strong linear separability condition, and the $(K-1)/2$ expected mistake lower bound follows from (Daniely & Helbertal, 2013, Claim 2).

Algorithm 1 can be extended to nonlinear classification using *positive definite kernels* (or *kernels*, for short), which are functions of the form $k : X \times X \to \mathbb{R}$ for some set X such that the matrix $\big(k(x_i, x_j)\big)_{i,j=1}^{m}$ is a symmetric positive semidefinite for any positive integer m and $x_1, x_2, \ldots, x_m \in X$ (Schölkopf & Smola, 2002, Definition 2.5).[6] As opposed to explicitly maintaining the weight vector for each class, the algorithm maintains the set of example-scalar pairs corresponding to the updates of the non-kernelized algorithm. As a direct consequence of Theorem 2 we get a mistake bound for the kernelized algorithm.

Theorem 4 (Mistake upper bound for kernelized algorithm)**.** *Let X be a non-empty set, let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. Let $\phi : X \to V$ be a feature map and let $k : X \times X \to \mathbb{R}$, $k(x, x') = \langle\phi(x), \phi(x')\rangle$ be the associated positive definite kernel. Let K be a positive integer, γ be a positive real number, R be a non-negative real number. If $(x_1, y_1), (x_2, y_2), \ldots, (x_T, y_T) \in X \times \{1, 2, \ldots, K\}$ are labeled examples such that:*

1. *the mapped examples $(\phi(x_1), y_1), \ldots, (\phi(x_T), y_T)$ are strongly linearly separable with margin γ,*
2. *$k(x_1, x_1), k(x_2, x_2), \ldots, k(x_T, x_T) \leq R^2$,*

then the expected number of mistakes that Algorithm 2 makes is at most $(K-1)\lfloor 4(R/\gamma)^2 \rfloor$.

5. From weak separability to strong separability

In this section, we consider the case when the examples are weakly linearly separable. Throughout this section, we assume without loss of generality that all examples lie in the unit ball $\mathrm{B}(\mathbf{0}, 1) \subseteq \mathbb{R}^d$.[7] Note that Algorithm 1 alone does not guarantee a finite mistake bound in this setting, as weak linear separability does not imply strong linear separability.

We use a positive definite kernel function $k(\cdot,\cdot)$, namely a *rational kernel* (Shalev-Shwartz et al., 2011) whose corresponding feature map $\phi(\cdot)$ transforms any sequence of *weakly* linearly separable examples to a *strongly* linearly separable sequence of examples. Specifically, ϕ has the property that if a set of labeled examples in $\mathrm{B}(\mathbf{0}, 1)$ is weakly linearly separable with a margin γ, then after applying ϕ the examples become strongly linearly separable with a margin γ' and their squared norms are bounded by 2. [8] The parameter γ' is a function of the old margin γ and the number of classes K, and is specified in Theorem 5 below.

Algorithm 2 KERNELIZED BANDIT ALGORITHM

Require: Number of classes K, number of rounds T.
Require: Kernel function $k(\cdot,\cdot)$.
Initialize $J_1^{(1)} = J_2^{(1)} = \cdots = J_K^{(1)} = \emptyset$
for $t = 1, 2, \ldots, T$ **do**
  Observe feature vector x_t.
  Compute
  $S_t = \Big\{ i \,:\, 1 \leq i \leq K,\ \sum_{(x,y)\in J_i^{(t)}} y k(x, x_t) \geq 0 \Big\}$
  if $S_t = \emptyset$ **then**
    Predict $\widehat{y}_t \sim \mathrm{Uniform}(\{1, 2, \ldots, K\})$
    Observe feedback $z_t = \mathbb{1}\left[\widehat{y}_t \neq y_t\right]$
    if $z_t = 1$ **then**
      Set $J_i^{(t+1)} = J_i^{(t)}$ for all $i \in \{1, 2, \ldots, K\}$
    else
      Set $J_i^{(t+1)} = J_i^{(t)}, \forall i \in \{1, 2, \ldots, K\} \setminus \{\widehat{y}_t\}$
      Update $J_{\widehat{y}_t}^{(t+1)} = J_{\widehat{y}_t}^{(t)} \cup \big\{(x_t, +1)\big\}$
  else
    Predict $\widehat{y}_t \in S_t$ chosen arbitrarily
    Observe feedback $z_t = \mathbb{1}\left[\widehat{y}_t \neq y_t\right]$
    if $z_t = 1$ **then**
      Set $J_i^{(t+1)} = J_i^{(t)}, \forall i \in \{1, 2, \ldots, K\} \setminus \{\widehat{y}_t\}$
      Update $J_{\widehat{y}_t}^{(t+1)} = J_{\widehat{y}_t}^{(t)} \cup \big\{(x_t, -1)\big\}$
    else
      Set $J_i^{(t+1)} = J_i^{(t)}$ for all $i \in \{1, 2, \ldots, K\}$

[6] For every kernel there exists an associated feature map $\phi : X \to V$ into some inner product space $(V, \langle\cdot,\cdot\rangle)$ such that $k(x, x') = \langle\phi(x), \phi(x')\rangle$.

[7] Instead of working with feature vector x_t we can work with normalized feature vectors $\widehat{x}_t = \frac{x_t}{\|x_t\|}$. It can be easily checked that if $(x_1, y_1), (x_2, y_2), \ldots, (x_T, y_T)$ are weakly linearly separable with margin γ and $\|x_t\| \leq R$ for all t, then the normalized examples $(\widehat{x}_1, y_1), (\widehat{x}_2, y_2), \ldots, (\widehat{x}_T, y_T)$ are weakly linearly separable with margin γ/R.

[8] Other kernels, such as the polynomial kernel $k(x, x') = (1 + \langle x, x'\rangle)^d$, or the multinomial kernel (Goel & Klivans, 2017) $k(x, x') = \sum_{i=0}^{d} (\langle x, x'\rangle)^i$, will have similar properties for large enough d.

The rational kernel $k : \mathrm{B}(\mathbf{0},1) \times \mathrm{B}(\mathbf{0},1) \to \mathbb{R}$ is defined as

$$k(x,x') = \frac{1}{1 - \frac{1}{2}\langle x, x' \rangle_{\mathbb{R}^d}} \,. \tag{6}$$

Note that $k(x,x')$ can be evaluated in $O(d)$ time.

Consider the classical real separable Hilbert space $\ell_2 = \{x \in \mathbb{R}^\infty \,:\, \sum_{i=1}^\infty x_i^2 < +\infty\}$ equipped with the standard inner product $\langle x, x' \rangle_{\ell_2} = \sum_{i=1}^\infty x_i x_i'$. If we index the coordinates of ℓ_2 by d-tuples $(\alpha_1, \alpha_2, \ldots, \alpha_d)$ of non-negative integers, the feature map that corresponds to k is $\phi : \mathrm{B}(\mathbf{0},1) \to \ell_2$,

$$\big(\phi(x_1, x_2, \ldots, x_d)\big)_{(\alpha_1,\alpha_2,\ldots,\alpha_d)} = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_d^{\alpha_d} \cdot \sqrt{2^{-(\alpha_1+\alpha_2+\cdots+\alpha_d)} \binom{\alpha_1+\alpha_2+\cdots+\alpha_d}{\alpha_1, \alpha_2, \ldots, \alpha_d}} \tag{7}$$

where $\binom{\alpha_1+\alpha_2+\cdots+\alpha_d}{\alpha_1,\alpha_2,\ldots,\alpha_d} = \frac{(\alpha_1+\alpha_2+\cdots+\alpha_d)!}{\alpha_1!\alpha_2!\ldots\alpha_d!}$ is the multinomial coefficient. It can be easily checked that

$$k(x,x') = \langle \phi(x), \phi(x') \rangle_{\ell_2} \,.$$

The last equality together with the formula for k implies that $k(x,x) < +\infty$ for any x in $\mathrm{B}(\mathbf{0},1)$ and thus in particular implies that $\phi(x)$ indeed lies in ℓ_2.

The following theorem is our main technical result in this section. We defer its proof to Section 5.1.

Theorem 5 (Margin transformation)**.** *Let $(x_1,y_1), (x_2,y_2), \ldots, (x_T,y_T) \in \mathrm{B}(\mathbf{0},1) \times \{1,2,\ldots,K\}$ be a sequence of labeled examples that is weakly linearly separable with margin $\gamma > 0$. Let ϕ be as defined in equation (7) and let*

$$\gamma_1 = \frac{\left\lceil 376 \lceil \log_2(2K-2) \rceil \cdot \left\lceil \sqrt{\frac{2}{\gamma}} \right\rceil \right\rceil^{\frac{-\lceil \log_2(2K-2) \rceil \cdot \lceil \sqrt{2/\gamma} \rceil}{2}}}{2\sqrt{K}},$$

$$\gamma_2 = \frac{\big(2^{s+1} r (K-1)(4s+2)\big)^{-(s+1/2)r(K-1)}}{4\sqrt{K}(4K-5)2^{K-1}} \,,$$

where $r = 2\left\lceil \frac{1}{4}\log_2(4K-3) \right\rceil + 1$ and $s = \lceil \log_2(2/\gamma) \rceil$. Then, the sequence of labeled examples transformed by ϕ, namely $(\phi(x_1),y_1), (\phi(x_2),y_2), \ldots, (\phi(x_T),y_T)$, is strongly linearly separable with margin $\gamma' = \max\{\gamma_1,\gamma_2\}$. In addition, for all t in $\{1,\ldots,T\}$, $k(x_t,x_t) \le 2$.

Using this theorem we derive a mistake bound for Algorithm 2 with kernel (6) under the weak linear separability assumption.

Corollary 6 (Mistake upper bound)**.** *Let K be a positive integer and let γ be a positive real number. If $(x_1,y_1), (x_2,y_2), \ldots, (x_T,y_T) \in \mathrm{B}(\mathbf{0},1) \times \{1,2,\ldots,K\}$ is a sequence of weakly separable labeled examples with margin $\gamma > 0$, then the expected number of mistakes made by Algorithm 2 with kernel $k(x,x')$ defined by (6) is at most $\min(2^{\widetilde{O}(K \log^2(1/\gamma))}, 2^{\widetilde{O}(\sqrt{1/\gamma}\log K)})$.*

This corollary follows directly from Theorems 4 and 5. We remark that under the weakly linearly separable setting, (Daniely & Helbertal, 2013) gives a mistake lower bound of $\Omega(\frac{K}{\gamma^2})$ for *any algorithm* (see also Theorem 3). We leave the possibility of designing efficient algorithms that have mistakes bounds matching this lower bound as an important open question.

5.1. Proof of Theorem 5

Overview. The idea behind the construction and analysis of the mapping ϕ is polynomial approximation. Specifically, we construct K multivariate polynomials $p_1, p_2, \ldots, p_K$ such that

$$\forall t \in \{1,2,\ldots,T\}\,, \qquad p_{y_t}(x_t) \ge \frac{\gamma'}{2}\,, \tag{8}$$

$$\forall t \in \{1,2,\ldots,T\}\ \forall i \in \{1,2,\ldots,K\} \setminus \{y_t\}\,, \quad p_i(x_t) \le -\frac{\gamma'}{2}\,. \tag{9}$$

We then show (Lemma 9) that each polynomial p_i can be expressed as $\langle c_i, \phi(x) \rangle_{\ell_2}$ for some $c_i \in \ell_2$. This immediately implies that the examples $(\phi(x_1),y_1), \ldots, (\phi(x_T),y_T)$ are strongly linearly separable with a positive margin.

The conditions (8) and (9) are equivalent to that

$$\forall t \in \{1,2,\ldots,T\}, y_t = i \quad \Rightarrow \quad p_i(x_t) \ge \frac{\gamma'}{2}\,, \tag{10}$$

$$\forall t \in \{1,2,\ldots,T\}, y_t \ne i \quad \Rightarrow \quad p_i(x_t) \le -\frac{\gamma'}{2}\,. \tag{11}$$

hold for all $i \in \{1,2,\ldots,K\}$. We can thus fix i and focus on construction of one particular polynomial p_i.

Since examples $(x_1,y_1), (x_2,y_2), \ldots, (x_T,y_T)$ are weakly linearly separable, all examples from class i lie in

$$R_i^+ = \bigcap_{j \in \{1,2,\ldots,K\} \setminus \{i\}} \left\{ x \in \mathrm{B}(\mathbf{0},1) \,:\, \left\langle w_i^* - w_j^*, x \right\rangle \ge \gamma \right\},$$

and all examples from the remaining classes lie in

$$R_i^- = \bigcup_{j \in \{1,2,\ldots,K\} \setminus \{i\}} \left\{ x \in \mathrm{B}(\mathbf{0},1) \,:\, \left\langle w_i^* - w_j^*, x \right\rangle \le -\gamma \right\}.$$

Therefore, to satisfy conditions (10) and (11), it suffices to

construct p_i such that

$$x \in R_i^+ \quad \Longrightarrow \quad p_i(x) \geq \frac{\gamma'}{2}\,, \tag{12}$$

$$x \in R_i^- \quad \Longrightarrow \quad p_i(x) \leq -\frac{\gamma'}{2}\,. \tag{13}$$

According to the well known Stone-Weierstrass theorem (see e.g. Davidson & Donsig, 2010, Section 10.10), on a compact set, multivariate polynomials uniformly approximate any continuous function. Roughly speaking, the conditions (12) and (13) mean that p_i approximates on $\mathrm{B}(\mathbf{0},1)$ a scalar multiple of the indicator function of the intersection of $K-1$ halfspaces $\bigcap_{j\in\{1,2,\ldots,K\}\setminus\{i\}}\left\{x : \left\langle w_i^* - w_j^*, x\right\rangle \geq 0\right\}$ while within margin γ along the decision boundary, the polynomial is allowed to attain arbitrary values. It is thus clear such a polynomial exists.

We give two explicit constructions for such polynomial in Theorems 7 and 8. Our constructions are based on Klivans & Servedio (2008) which in turn uses the constructions from Beigel et al. (1995). More importantly, the theorems quantify certain parameters of the polynomial, which allows us to upper bound the transformed margin γ'.

Before we state the theorems, recall that a polynomial of d variables is a function $p : \mathbb{R}^d \to \mathbb{R}$ of the form

$$\begin{aligned} p(x) &= p(x_1, x_2, \ldots, x_d) \\ &= \sum_{\alpha_1,\alpha_2,\ldots,\alpha_d} c_{\alpha_1,\alpha_2,\ldots,\alpha_d} x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_d^{\alpha_d} \end{aligned}$$

where the sum ranges over a finite set of d-tuples $(\alpha_1, \alpha_2, \ldots, \alpha_d)$ of non-negative integers and $c_{\alpha_1,\alpha_2,\ldots,\alpha_d}$'s are real coefficients. The *degree* of a polynomial p, denoted by $\deg(p)$, is the largest value of $\alpha_1 + \alpha_2 + \cdots + \alpha_d$ for which the coefficient $c_{\alpha_1,\alpha_2,\ldots,\alpha_d}$ is non-zero. Following the terminology of Klivans & Servedio (2008), the *norm of a polynomial* p is defined as

$$\|p\| = \sqrt{\sum_{\alpha_1,\alpha_2,\ldots,\alpha_d} \left(c_{\alpha_1,\alpha_2,\ldots,\alpha_d}\right)^2}\,.$$

It is easy see that this is indeed a norm, since we can interpret it as the Euclidean norm of the vector of the coefficients of the polynomial.

Theorem 7 (Polynomial approximation of intersection of halfspaces I)**.** *Let $v_1, v_2, \ldots, v_m \in \mathbb{R}^d$ be vectors such that $\|v_1\|, \|v_2\|, \ldots, \|v_m\| \leq 1$. Let $\gamma \in (0,1)$. There exists a multivariate polynomial $p : \mathbb{R}^d \to \mathbb{R}$ such that*

1. $\underset{m}{p(x)} \geq 1/2$ *for all* $x \in R^+ = \bigcap_{i=1}^{m} \left\{x \in \mathrm{B}(\mathbf{0},1) \,:\, \langle v_i, x\rangle \geq \gamma\right\}$,
2. $\underset{m}{p(x)} \leq -1/2$ *for all* $x \in R^- = \bigcup_{i=1}^{m} \left\{x \in \mathrm{B}(\mathbf{0},1) \,:\, \langle v_i, x\rangle \leq -\gamma\right\}$,
3. $\deg(p) = \left\lceil \log_2(2m)\right\rceil \cdot \left\lceil \sqrt{1/\gamma}\right\rceil$,
4. $\|p\| \leq \left[188\left\lceil \log_2(2m)\right\rceil \cdot \left\lceil \sqrt{1/\gamma}\right\rceil\right]^{\frac{\left\lceil \log_2(2m)\right\rceil \cdot \left\lceil \sqrt{1/\gamma}\right\rceil}{2}}$.

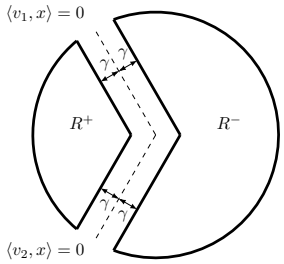


Figure 2. The figure shows the two regions R^+ and R^- used in parts 1 and 2 of Theorems 7 and 8 for the case $m = d = 2$ and a particular choice of vectors v_1, v_2 and margin parameter γ. The separating hyperplanes $\langle v_1, x\rangle = 0$ and $\langle v_2, x\rangle = 0$ are shown as dashed lines.

Theorem 8 (Polynomial approximation of intersection of halfspaces II)**.** *Let $v_1, v_2, \ldots, v_m \in \mathbb{R}^d$ be vectors such that $\|v_1\|, \|v_2\|, \ldots, \|v_m\| \leq 1$. Let $\gamma \in (0,1)$. Define*

$$r = 2\left\lceil \frac{1}{4}\log_2(4m+1)\right\rceil + 1 \quad \textit{and} \quad s = \left\lceil \log_2(1/\gamma)\right\rceil\,.$$

Then, there exists a multivariate polynomial $p : \mathbb{R}^d \to \mathbb{R}$ such that

1. $\underset{m}{p(x)} \geq 1/2$ *for all* $x \in R^+ = \bigcap_{i=1}^{m} \left\{x \in \mathrm{B}(\mathbf{0},1) \,:\, \langle v_i, x\rangle \geq \gamma\right\}$,
2. $\underset{m}{p(x)} \leq -1/2$ *for all* $x \in R^- = \bigcup_{i=1}^{m} \left\{x \in \mathrm{B}(\mathbf{0},1) \,:\, \langle v_i, x\rangle \leq -\gamma\right\}$,
3. $\deg(p) \leq (2s+1)rm$,
4. $\|p\| \leq (4m-1)2^m \cdot \left(2^s rm(4s+2)\right)^{(s+1/2)rm}$.

The proofs of the theorems are in Appendix D. The geometric interpretation of the two regions R^+ and R^- in the theorems is explained in Figure 2. Similar but weaker results were proved by Klivans & Servedio (2008). Specifically, our bounds in parts 1, 2, 3, 4 of Theorems 7 and 8 are independent of the dimension d.

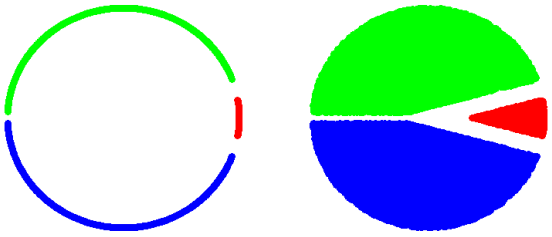
The following lemma establishes a correspondence between any multivariate polynomial in $\mathbb{R}^d$ and an element in ℓ_2, and gives an upper bound on its norm. Its proof follows from simple algebra, which we defer to Appendix C.

Lemma 9 (Norm bound). *Let $p : \mathbb{R}^d \to \mathbb{R}$ be a multivariate polynomial. There exists $c \in \ell_2$ such that $p(x) = \langle c, \phi(x) \rangle_{\ell_2}$ and $\|c\|_{\ell_2} \leq 2^{\deg(p)/2} \|p\|$.*

Using the lemma and the polynomial approximation theorems, we can prove that the mapping ϕ maps any set of weakly linearly separable examples to a strongly linearly separable set of examples. Due to space constraints, we defer the full proof of Theorem 5 to Appendix E.

6. Experiments

In this section, we provide an empirical evaluation on our algorithms, verifying their effectiveness on linearly separable datasets. We generated strongly and weakly linearly separable datasets with $K = 3$ classes in $\mathbb{R}^3$ i.i.d. from two data distributions. Figures 3a and 3b show visualizations of the two datasets, along with detailed descriptions of the distributions.



(a) Strongly separable case (b) Weakly separable case

Figure 3. Strongly and weakly linearly separable datasets in $\mathbb{R}^3$ with $K = 3$ classes and $T = 5 \times 10^6$ examples. Here we show projections of the examples onto their first two coordinates, which lie in the ball of radius $1/\sqrt{2}$ centered at the origin. The third coordinate is $1/\sqrt{2}$ for all examples. Class 1 is depicted red. Classes 2 and 3 are depicted green and blue, respectively. 80% of the examples belong to class 1, 10% belong to class 2 and 10% belong to class 3. Class 1 lies in the angle interval $[-15°, 15°]$, while classes 2 and 3 lie in the angle intervals $[15°, 180°]$ and $[-180°, -15°]$ respectively. The examples are strongly and weakly linearly separable with a margin of $\gamma = 0.05$, respectively. (Examples lying within margin γ of the linear separators were rejected during sampling.)

We implemented Algorithm 1, Algorithm 2 with rational kernel (6) and used implementation of BANDITRON algorithm by Orabona (2009). We evaluated these algorithms on the two datasets. BANDITRON has an exploration rate parameter, for which we tried values $0.02, 0.01, 0.005, 0.002, 0.001, 0.0005$. Since all three algorithms are randomized, we run each algorithm 20 times. The average cumulative number of mistakes up to round t as a function of t are shown in Figures 4 and 5.

We can see that there is a tradeoff in the setting of the exploration rate for BANDITRON. With large exploration parameter, BANDITRON suffers from over-exploration, whereas with small exploration parameter, its model cannot be updated quickly enough. As expected, Algorithm 1 has a small number of mistakes in the strongly linearly separable setting, while having a large number of mistakes in the weakly linearly separable setting, due to the limited representation power of linear classifiers. In contrast, Algorithm 2 with rational kernel has a small number of mistakes in both settings, exhibiting strong adaptivity guarantees. Appendix F shows the decision boundaries that each of the algorithms learns by the end of the last round.

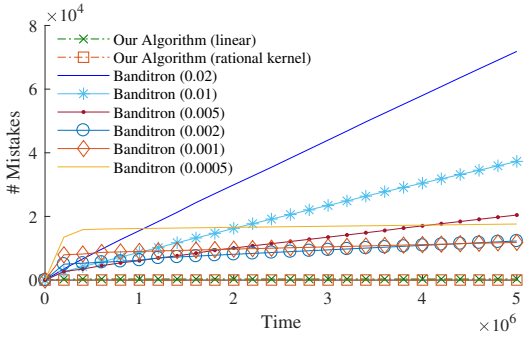


Figure 4. The average cumulative number of mistakes versus the number of rounds on the strongly linearly separable dataset in Figure 3a.

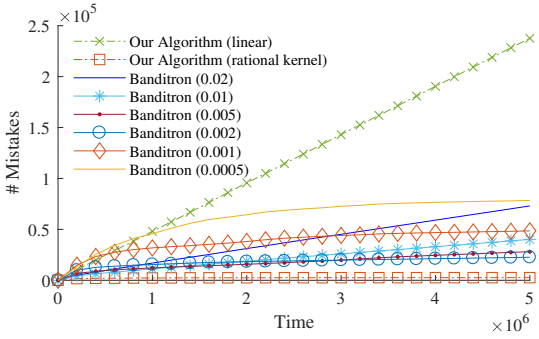


Figure 5. The average cumulative number of mistakes versus the number of rounds on the weakly linearly separable dataset in Figure 3b.

Acknowledgments

We thank Francesco Orabona and Wen Sun for helpful initial discussions, and thank Adam Klivans and Rocco Servedio for helpful discussions on (Klivans & Servedio, 2008) and pointing out the reference (Klivans & Servedio, 2004).

We also thank Dylan Foster, Akshay Krishnamurthy, and Haipeng Luo for providing a candidate solution to our problem.

References

Abernethy, J. and Rakhlin, A. An efficient bandit algorithm for $\sqrt{T}$-regret in online multiclass prediction? In *Proceedings of the 22nd Annual Conference on Learning Theory (COLT 2009)*, 2009.

Agarwal, A., Hsu, D., Kale, S., Langford, J., Li, L., and Schapire, R. Taming the monster: a fast and simple algorithm for contextual bandits. In *International Conference on Machine Learning (ICML 2014)*, 2014.

Arora, S., Babai, L., Stern, J., and Sweedyk, Z. The hardness of approximate optima in lattices, codes, and systems of linear equations. *Journal of Computer and System Sciences*, 54(2):317–331, 1997.

Auer, P. and Long, P. M. Structural results about online learning models with and without queries. *Machine Learning*, 36(3):147–181, 1999.

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

Beigel, R., Reingold, N., and Spielman, D. PP is closed under intersection. *Journal of Computer and System Sciences*, 50(2):191–202, 1995.

Beygelzimer, A., Orabona, F., and Zhang, C. Efficient online bandit multiclass learning with $\widetilde{O}(\sqrt{T})$ regret. In *International Conference on Machine Learning*, pp. 488–497, 2017.

Blum, A. L. and Rivest, R. L. Training a 3-node neural network is NP-complete. In *Machine learning: From theory to applications*, pp. 9–28. Springer, 1993.

Chen, G., Chen, G., Zhang, J., Chen, S., and Zhang, C. Beyond banditron: A conservative and efficient reduction for online multiclass prediction with bandit setting model. In *Ninth IEEE International Conference on Data Mining, 2009 (ICDM 2009)*, pp. 71–80. IEEE, 2009.

Chen, S.-T., Lin, H.-T., and Lu, C.-J. Boosting with online binary learners for the multiclass bandit problem. In Xing, E. P. and Jebara, T. (eds.), *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pp. 342–350, Bejing, China, 22–24 Jun 2014. PMLR.

Crammer, K. and Gentile, C. Multiclass classification with bandit feedback using adaptive regularization. *Machine learning*, 90(3):347–383, 2013.

Crammer, K. and Singer, Y. Ultraconservative online algorithms for multiclass problems. *Journal of Machine Learning Research*, 3(Jan):951–991, 2003.

Daniely, A. and Helbertal, T. The price of bandit information in multiclass online classification. In *Conference on Learning Theory*, pp. 93–104, 2013.

Davidson, K. R. and Donsig, A. P. *Real analysis and Applications*. Springer, 2010.

Dudík, M., Hsu, D., Kale, S., Karampatziakis, N., Langford, J., Reyzin, L., and Zhang, T. Efficient optimal learning for contextual bandits. In *UAI 2011*, pp. 169–178, 2011.

Foster, D. and Krishnamurthy, A. Contextual bandits with surrogate losses: Margin bounds and efficient algorithms. In *Advances in Neural Information Processing Systems*, 2018.

Foster, D. J., Kale, S., Luo, H., Mohri, M., and Sridharan, K. Logistic regression: The importance of being improper. In Bubeck, S., Perchet, V., and Rigollet, P. (eds.), *Proceedings of the 31st Conference On Learning Theory (COLT 2018)*, volume 75 of *Proceedings of Machine Learning Research*, pp. 167–208. PMLR, 06–09 Jul 2018.

Garey, M. R. and Johnson, D. S. *Computers and intractability: A guide to the theory of NP-completeness*. Freeman, 1979.

Goel, S. and Klivans, A. Learning depth-three neural networks in polynomial time. *arXiv preprint arXiv:1709.06010*, 2017.

Hazan, E. and Kale, S. Newtron: An efficient bandit algorithm for online multiclass prediction. In *Advances in neural information processing systems*, pp. 891–899, 2011.

Kakade, S. M., Shalev-Shwartz, S., and Tewari, A. Efficient bandit algorithms for online multiclass prediction. In *Proceedings of the 25th International Conference on Machine Learning*, pp. 440–447. ACM, 2008.

Klivans, A. R. and Servedio, R. A. Perceptron-like performance for learning intersections of halfspaces. *COLT*, Open problem, 2004.

Klivans, A. R. and Servedio, R. A. Learning intersections of halfspaces with a margin. *Journal of Computer and System Sciences*, 74(1):35–48, 2008.

Langford, J. and Zhang, T. The epoch-greedy algorithm for multi-armed bandits with side information. In *NIPS 20*, pp. 817–824, 2008.

Long, P. M. On the sample complexity of pac learning halfspaces against the uniform distribution. *IEEE Transactions on Neural Networks*, 6(6):1556–1559, 1995.

Long, P. M. New bounds on the price of bandit feedback for mistake-bounded online multiclass learning. In *International Conference on Algorithmic Learning Theory*, pp. 3–10, 2017.

Mason, J. C. and Handscomb, D. C. *Chebyshev polynomials*. Chapman and Hall/CRC, 2002.

Orabona, F. *DOGMA: a MATLAB toolbox for Online Learning*, 2009. Software available at `http://dogma.sourceforge.net`.

Rakhlin, A. and Sridharan, K. BISTRO: An efficient relaxation-based method for contextual bandits. In *International Conference on Machine Learning (ICML 2016)*, pp. 1977–1985, 2016.

Schölkopf, B. and Smola, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2002.

Shalev-Shwartz, S. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2012.

Shalev-Shwartz, S., Shamir, O., and Sridharan, K. Learning kernel-based halfspaces with the 0-1 loss. *SIAM Journal on Computing*, 40(6):1623–1646, 2011.

Shawe-Taylor, J. and Cristianini, N. *Kernel methods for pattern analysis*. Cambridge university press, 2004.

Syrgkanis, V., Krishnamurthy, A., and Schapire, R. Efficient algorithms for adversarial contextual learning. In *ICML*, pp. 2159–2168, 2016a.

Syrgkanis, V., Luo, H., Krishnamurthy, A., and Schapire, R. E. Improved regret bounds for oracle-based adversarial contextual bandits. In *NIPS*, pp. 3135–3143, 2016b.

Wang, S., Jin, R., and Valizadegan, H. A potential-based framework for online multi-class learning with partial feedback. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 900–907, 2010.

Zhang, D., Jung, Y. H., and Tewari, A. Online multiclass boosting with bandit feedback. *arXiv preprint arXiv:1810.05290*, 2018.