JMLR: Workshop and Conference Proceedings 83:1–28, 2018 ALT 2018

The k-Nearest Neighbour UCB Algorithm for Multi-Armed Bandits with Covariates

Henry WJ Reeve
HENRY.REEVE@MANCHESTER.AC.UK
School of Computer Science,
The University of Manchester,
Manchester, UK

Joe Mellor
JOE.MELLOR@ED.AC.UK
Usher Institute of Population Health Sciences and Informatics
The University of Edinburgh,
Edinburgh, UK

Gavin Brown
GAVIN.BROWN@MANCHESTER.AC.UK
School of Computer Science,
The University of Manchester,
Manchester, UK

Editor: Mehryar Mohri & Karthik Sridharan

Abstract

In this paper we propose and explore the k-Nearest Neighbour UCB algorithm for multi-armed bandits with covariates. We focus on a setting where the covariates are supported on a metric space of low intrinsic dimension, such as a manifold embedded within a high dimensional ambient feature space. The algorithm is conceptually simple and straightforward to implement. The k-Nearest Neighbour UCB algorithm does not require prior knowledge of the either the intrinsic dimension of the marginal distribution or the time horizon. We prove a regret bound for the k-Nearest Neighbour UCB algorithm which is minimax optimal up to logarithmic factors. In particular, the algorithm automatically takes advantage of both low intrinsic dimensionality of the marginal distribution over the covariates and low noise in the data, expressed as a margin condition. In addition, focusing on the case of bounded rewards, we give corresponding regret bounds for the k-Nearest Neighbour KL-UCB algorithm, which is an analogue of the KL-UCB algorithm adapted to the setting of multi-armed bandits with covariates. Finally, we present empirical results which demonstrate the ability of both the k-Nearest Neighbour UCB and k-Nearest Neighbour KL-UCB to take advantage of situations where the data is supported on an unknown sub-manifold of a high-dimensional feature space.

© 2018 H.W. Reeve, J. Mellor & G. Brown.

1. Introduction

The multi-armed bandit is a simple model which exemplifies the exploitation-exploration trade-off in reinforcement learning. Solutions to this problem have numerous practical applications from sequential clinical trials to web-page ad placement (Bubeck et al. (2012)). We focus upon the stochastic setting in which an agent is given access to a collection of unknown reward distributions (arms); the agent sequentially selects a reward distribution to sample from, so as to maximise their cumulative reward. One of the most widely used strategies for stochastic multi-armed bandits is the Upper Confidence Bound (UCB) algorithm, which is based on the principle of optimism in the face of uncertainty (Lai and Robbins (1985); Agrawal (1995); Auer et al. (2002)). Garivier and Cappé's KL-UCB algorithm utilises tighter upper confidence bounds to provide an algorithm with sharper regret bounds and a superior empirical performance (Garivier and Cappé (2011)).

Multi-armed bandits with covariates extend this simple model by allowing the reward distributions to depend upon observable side information (Bubeck et al., 2012, Section 4.3). For example, in sequential clinical trials the agent might have access to a patient's MRI scan or genome sequence; in web-page ad placement side-information might include a particular user's preferences and purchasing history. Owing to their widespread applicability, multi-armed bandits with covariates have been extensively studied (Beygelzimer et al. (2011); Kakade et al. (2008); Langford and Zhang (2008); Perchet et al. (2013); Qian and Yang (2016); Rigollet and Zeevi (2010); Seldin et al. (2011); Slivkins (2011); Wang et al. (2005a,b); Yang et al. (2002)). In this paper we shall consider the non-parametric setting in which the relationship between reward distribution and side-information is assumed to satisfy smoothness conditions, without specifying a particular parametric form. Yang and Zhu proved strong-consistency for an epsilon-greedy approach to this problem, using either nearest neighbour or histogram based methods to model the functional dependency of the reward distribution upon the covariate (Yang et al. (2002)). Rigollet and Zeevi introduced the UCBogram which partitions the covariate space into cubes and runs the UCB locally on each member of the partition (Rigollet and Zeevi (2010)). Rigollet and Zeevi prove a regret bound with exponents depending upon distributional assumptions including a natural extension of the Tysbakov margin condition (Tsybakov (2004)). Unfortunately, the regret bound is sub-optimal when the margin parameter is greater than one. Later Perchet and Rigollet developed the Adaptively Binned Successive Elimination algorithm (ABSE) which runs the Successive Elimination algorithm locally on increasingly refined partitions of the covariate space (Perchet et al. (2013)). Perchet and Rigollet demonstrated that the ABSE algorithm achieves minimax optimal regret guarantees for all values of the margin parameter (Perchet et al. (2013)). Hence, the adaptive refinement of the partition of the covariate space enables the ABSE algorithm to take advantage of low-noise conditions, expressed as a margin condition.

Despite the strong theoretical merits of the ABSE algorithm, there are several limitations owing to its dependency upon a partition of the feature space into dyadic hyper-cubes. Firstly, there are many applications in which it is natural to construct a metric between data points which cannot be embedded in a Euclidean space without significant distortion. Examples include the Wasserstein distance between images and the edit distance on graphs,

Frogner et al. (2015); Luxburg and Bousquet (2004). However, neither the UBogram nor the ABSE algorithm can be applied to non-Euclidean metric spaces.

Secondly, the regret bounds for the ABSE algorithm require that the marginal distribution μ be Lebesgue absolutely continuous with a density bounded from below on the unit hyper-cube $[0,1]^D$. Whilst this condition is not entirely necessary for the analysis, the proof does depend crucially upon the existence of constants $C_d, d > 0$ such that the following holds. For every dyadic hyper-cube $B \subset [0,1]^D$ of the form $B = 2^{-q} \cdot \prod_{i=1}^{D} [z_i, z_i + 1]$ with $z_1, \cdots, z_D, q \in \mathbb{N} \cup \{0\}$, we have either $\mu(B) \geq C_d \cdot \mathrm{diam}(B)^d$ or $\mu(B) = 0$. However, this condition does not hold for many well-behaved measures on Euclidean space (see Appendix H for a simple example).

Thirdly, the construction of the partitions in both the UCBogram and the ABSE algorithm requires prior knowledge of the intrinsic dimensionality of the covariate space as an input parameter. In the case of the UCBogram the dimension d is used to choose the optimal partition size (Rigollet and Zeevi, 2010, Theorem 3.1). In the case of the ABSE algorithm, a partition element B is refined after l_B rounds, where l_B is a number which depends upon the dimension d (Perchet et al., 2013, Equation (5.2)). Moreover, if covariates are supported on a low-dimensional sub-manifold, then the intrinsic dimensionality of the sub-manifold is unlikely to be known in advance. The aim of the current paper is to address these three limitations.

The k-nearest neighbour method is amongst the simplest approaches to supervised learning. In addition, it has strong theoretical guarantees. Kpotufe has shown that the k-nearest neighbour regression algorithm attains distribution dependent minimax optimal rates, without prior knowledge of the intrinsic dimensionality of the data (Kpotufe (2011)). Chaudhuri and Dasgupta have shown the k-nearest neighbour method attains distribution dependent minimax optimal rates in the supervised classification setting (Chaudhuri and Dasgupta (2014)). In particular, the k-nearest neighbour classifier automatically takes advantage of low noise in the data, expressed as a margin condition. In light of these theoretical strengths, it is natural to apply the k-nearest neighbour method to problem of multi-armed bandits with covariates.

We propose the k-nearest neighbour UCB algorithm (k-NN UCB), a conceptually simple procedure for multi-armed bandits with covariates which combines the UCB algorithm with k-nearest neighbour regression. The algorithm does not require prior knowledge of the intrinsic dimensionality of the data. It is also naturally anytime, without resorting to the doubling trick. We prove a regret bound for the k-NN UCB algorithm which is minimax optimal up to logarithmic factors. In particular, the algorithm automatically takes advantage of both low intrinsic dimensionality of the marginal distribution over the covariates and low noise conditions, expressed as a margin condition. In addition, focusing on the case of bounded rewards, we give corresponding regret bounds for the k-nearest neighbour KL-UCB algorithm (k-NN KL-UCB), which is an analogue of the KL-UCB algorithm (Garivier and Cappé (2011)) adapted to the setting of multi-armed bandits with covariates. Finally, we present empirical results which demonstrate the ability of both k-NN UCB and k-NN KL-UCB to take advantage of situations where the data is supported on an unknown sub-manifold of a high-dimensional feature space.

2. Bandits on a metric space

In this section we shall introduce some notation and background.

2.1. Notation

We consider the problem of bandits with covariates on metric spaces. Suppose we have a metric space $(\mathcal{X}, \rho)$. Given $x \in \mathcal{X}$ and $r > 0$ we let $B(x, r)$ denote the open metric ball of radius r, centred at x. Given $q \in \mathbb{N}$ we let $[q] = \{1, \cdots, q\}$. Given a collection of A arms, we let $\mathbb{P}$ denote a distribution over random variables (X, Y) with $X \in \mathcal{X}$ and $Y = (Y^a)_{a \in [A]} \in \mathbb{R}^A$, where Y^a denotes the value of arm a. We let μ denote the marginal of $\mathbb{P}$ over $X \in \mathcal{X}$ and let $\text{supp}(\mu)$ denote its support. For each $a \in [A]$ we define a function $f^a : \mathcal{X} \to [0, 1]$ by $f^a(x) = \mathbb{E}\left[Y^a | X = x\right]$.

For each $t \in [n]$ a random sample (X_t, Y_t) is drawn i.i.d from $\mathbb{P}$. We are allowed to view the feature vector $X_t \sim \mu$ and we must choose an arm $a \in [A]$ and receive the stochastic reward Y_t^a. We are able to observe the value of our chosen arm, but not the value of the remaining arms. Our sequential choice of arms is given by a policy $\pi = \{\pi_t\}_{t \in [n]}$ consisting of functions $\pi_t : \mathcal{X} \to [A]$, where π_t is determined purely by the known reward history $\mathcal{D}_{t-1} = \left\{(X_s, \pi_s, Y_s^{\pi_s})\right\}_{s \in [t-1]}$. The goal is to choose π_t so as to maximise the cumulative reward $\sum_{t \in [n]} Y_t^{\pi_t}$. In order to quantify the quality of a policy π we compare its expected cumulative reward to the cumulative reward to that of an oracle policy $\pi^* = \{\pi_t^*\}_{t \in [n]}$ defined by $\pi_t^* \in \text{argmax}_{a \in [A]} \{f^a(X_t)\}$. We define the regret by $R_n(\pi) = \sum_{t \in [n]} \left(Y_t^{\pi_t^*} - Y_t^{\pi_t}\right)$.

2.2. Assumptions

We shall make the following assumptions:

Assumption 1 (Dimension assumption): There exists $C_d, d, R_{\mathcal{X}} > 0$ such that for all $x \in \text{supp}(\mu)$, $r \in (0, R_{\mathcal{X}})$ we have $\mu\left(B(x; r)\right) \geq C_d \cdot r^d$.

Assumption 1 holds for well-behaved measures μ which are absolutely continuous with respect to the Riemannian volume form $V_{\mathcal{M}}$ on a d-dimensional sub-manifold of Euclidean space (see Proposition 2, Appendix I). See Appendix H for an example where Assumption 1 holds whilst the measure of dyadic sub-cubes is not well behaved.

Assumption 2 (Lipschitz assumption): There exists a constant $\lambda > 0$ such that for all $a \in [A]$, $x_0, x_1 \in \mathcal{X}$ we have $\left|f^a(x_0) - f^a(x_1)\right| \leq \lambda \cdot \rho\left(x_0, x_1\right)$.

Assumption 2 quantifies the requirement that similar covariates should imply similar conditional reward expectations. Let $f^*(x) = \max_{a \in [A]} \{f^a(x)\}$. For each $a \in [A]$ let $\Delta^a(x) = f^*(x) - f^a(x)$, and define

$$\Delta(x) = \begin{cases} \min_{a \in [A]} \{\Delta^a(x) : \Delta^a(x) > 0\} & \text{if } \exists a \in [A] \;\; \Delta^a(x) > 0 \\ 0 \text{ otherwise.} \end{cases}$$

Assumption 3 (Margin assumption): There exists $\delta_\alpha, C_\alpha, \alpha > 0$ such that for all $\delta \in (0, \delta_\alpha)$ we have $\mu\left(\{x \in \mathcal{X} : 0 < \Delta(x) < \delta\}\right) \leq C_\alpha \cdot \delta^\alpha$.

Assumption 3 quantifies the difficulty of the problem. It is a natural analogue of Tysbakov's margin condition (Tsybakov (2004)) introduced by Rigollet and Zeevi (2010). Perchet and Rigollet showed that if $\mathcal{X}$ is a manifold and $\alpha > d$ then we must have $\eta(x) \neq \infty$ on the interior of $\text{supp}(\mu)$ (Perchet et al., 2013, Proposition 3.1). All of our theoretical results require assumptions 1, 2 and 3. We shall also use one of the following two assumptions.

Assumption 4 (Subgaussian noise assumption): For each $t \in [n]$ and $a \in [A]$ the arms Y_t^a have sub-gaussian noise ie. for all $x \in \mathcal{X}$ and $\theta \in \mathbb{R}$,

$$\mathbb{E}\left[\exp\left(\theta \cdot \left(Y_t^a - f^a(x)\right)\right) | X_t = x\right] \leq \exp\left(\theta^2/2\right).$$

Assumption 5 (Bounded rewards assumption): For all $t \in [n]$ & $a \in [A]$, $Y_t^a \in [0, 1]$.

3. Nearest neighbour algorithms

In this section we introduce a pair of nearest neighbour based UCB strategies. We begin by introducing a generalized k-nearest neighbours index strategy, of which the other strategies are special cases.

3.1. The generalized k-nearest neighbours index strategy

Suppose we are at a time step $t \in [n]$ and we have access to the reward history $\mathcal{D}_{t-1}$. For each $x \in \mathcal{X}$ we let $\{\tau_{t,q}(x)\}_{q \in [t-1]}$ be an enumeration of $[t-1]$ such that for each $q \leq t-2$,

$$\rho\left(x, X_{\tau_{t,q}(x)}\right) \leq \rho\left(x, X_{\tau_{t,q+1}(x)}\right).$$

Given $x \in \mathcal{X}$ and $k \in [t-1]$ we define $\Gamma_{t,k}(x) := \{\tau_{t,q}(x) : q \in [k]\} \subseteq [t-1]$ and let

$$r_{t,k}(x) = \max\left\{\rho\left(x, X_s\right) : s \in \Gamma_{t,k}(x)\right\} = \rho\left(x, X_{\tau_{t,k}(x)}\right).$$

We adopt the convention that $0/0 := 0$. For each $a \in [A]$ we define

$$N_{t,k}^a(x) := \sum_{s \in \Gamma_{t,k}(x)} \mathbb{1}\left\{\pi_s = a\right\},$$

$$S_{t,k}^a(x) := \sum_{s \in \Gamma_{t,k}(x)} \mathbb{1}\left\{\pi_s = a\right\} \cdot Y_s^a,$$

$$\hat{f}_{t,k}^a(x) := S_{t,k}^a(x) / N_{t,k}^a(x).$$

In addition, given a constant $\theta > 0$ and a non-decreasing function $\varphi : \mathbb{N} \to [1, \infty)$ we define a corresponding uncertainty value $U_{t,k}^a(x)$ by

$$U_{t,k}^a(x) := \sqrt{\left(\theta \log t\right) / N_{t,k}^a(x)} + \varphi(t) \cdot r_{t,k}(x).$$

We shall combine $\hat{f}_{t,k}^a(x)$, $U_{t,k}^a(x)$, $N_{t,k}^a(x)$ and $r_{t,k}(x)$ to construct an index $\mathcal{I}_{t,k}^a(x)$ corresponding to an upper-confidence bound on the reward function $f^a(x)$. Our algorithm then proceeds as follows. At each time step t, a feature vector X_t is received. For each arm $a \in [A]$, the algorithm selects a number of neighbours $k_t(a)$ by minimising the uncertainty $U_{t,k}^a(X_t)$. The algorithm then selects the arm which maximises the index $\mathcal{I}_{t,k_t(a)}^a(X_t)$. The psuedo-code for this generalised k-NN index strategy is presented in Algorithm 1.

Algorithm 1: The k-NN index strategy

1. For $t = 1, \cdots, A$, do $\pi_t = t$;
2. For $t = A+1, \cdots, n$,
   (a) Observe X_t;
   (b) For $a = 1, \cdots, A$,
   Choose $k_t(a) \leftarrow \text{argmin}_{k \in [t-1]} \left\{ U^a_{t,k}(X_t) \right\}$;
   (c) Choose $\pi_t \in \text{argmax}_{a \in [A]} \left\{ \mathcal{I}^a_{t,k_t(a)}(X_t) \right\}$;
   (d) Receive reward $Y^{\pi_t}_t$;

By selecting $k_t(a)$ so as to minimise the $U^a_{t,k}(X_t)$ we avoid giving an explicit formula for k. This is fortuitous, since in order to obtain optimal regret bounds, any such formula would necessarily depend upon both the time horizon n and the intrinsic dimensionality of the data d, and in general, neither n nor d will be known a priori by the learner. Selecting $k_t(a)$ in this way is inspired by Kpotufe's procedure for selecting k in the regression setting, so as to minimise an upper bound on the squared error (Kpotufe (2011)).

3.2. k-Nearest Neighbour UCB

The k-Nearest Neighbour UCB algorithm (k-NN UCB) is a special case of Algorithm 1 with the following index function,

$$\mathcal{I}^a_{t,k}(x) = \hat{f}^a_{t,k}(x) + U^a_{t,k}(x). \tag{1}$$

The k-NN UCB algorithm satisfies the following regret bound whenever the noise is subgaussian (Assumption 4). First we let $\varphi^{-1}(\lambda) := \inf\{t \in \mathbb{N} : \varphi(t) \geq \lambda\}$ and define $M := \max_{a \in [A]} \{\sup_{x \in \mathcal{X}} \{\Delta^a(x)\}\}$. For all $n \in \mathbb{N}$ let $\overline{\log}(n) := \max\{1, \log(n)\}$.

Theorem 1 *Suppose that Assumption 1 holds with constants $R_{\mathcal{X}}, C_d, d$, Assumption 2 holds with Lipschitz constant λ, Assumption 3 holds with constants $\delta_\alpha, C_\alpha, \alpha > 0$ and Assumption 4 holds. Let π be the k-NN UCB algorithm (Algorithm 1 with $\mathcal{I}^a_{t,k}$ as in equation (1)). Then for all $\theta > 4$ there exists a constant C, depending solely upon $R_{\mathcal{X}}, C_d, d, \delta_\alpha, C_\alpha, \alpha$ and θ such that for all $n \in \mathbb{N}$ we have*

$$\mathbb{E}\left[R_n(\pi)\right] \leq M \cdot \varphi^{-1}(\lambda) + C \cdot A \cdot \left(M \cdot \varphi(n)^d + n \cdot \left(\frac{\varphi(n)^d \cdot \overline{\log}(n)}{n} \right)^{\min\left\{\frac{\alpha+1}{d+2}, 1\right\}} \right).$$

Theorem 1 follows from the more general Theorem 3 in Section 4. The full proof is given in Appendix C. Note that by taking $\varphi(n) = O(\log n)$ we obtain a regret bound which is minimax optimal up to logarithmic factors for any smooth compact embedded sub-manifold (See Theorem 7, Appendix I for details).

3.3. k-Nearest Neighbour KL-UCB

The k-Nearest Neighbour KL-UCB algorithm is another special case of Algorithm 1, customized for the setting of bounded rewards. The k-Nearest Neighbour KL-UCB algorithm is an adaptation of the KL-UCB algorithm of Garivier and Cappé (2011), which has shown strong empirical performance combined with tight regret bounds. Given $p, q \in [0,1]$ we define the Kullback-Leibler divergence $d(p,q) = p \log(p/q) + (1-p) \cdot \log((1-p)/(1-q))$.

$$\mathcal{I}^a_{t,k}(x) = \sup\left\{\omega \in [0,1] : N^a_{t,k}(x) \cdot d\left(\hat{f}^a_{t,k}(x), \omega\right) \leq \theta \cdot \log t\right\} + \varphi(t) \cdot r_{t,k}(x). \tag{2}$$

Theorem 2 *Suppose that Assumption 1 holds with constants $R_{\mathcal{X}}, C_d, d$, Assumption 2 holds with Lipschitz constant λ, Assumption 3 holds with constants $\delta_\alpha, C_\alpha, \alpha > 0$ and Assumption 5 holds. Let π be the k-NN KL-UCB algorithm (Algorithm 1 with $\mathcal{I}^a_{t,k}$ as in equation (2)). Then for all $\theta > 2$ there exists a constant C, depending solely upon $R_{\mathcal{X}}, C_d, d, \delta_\alpha, C_\alpha, \alpha$ and θ such that for all $n \in \mathbb{N}$ we have*

$$\mathbb{E}\left[R_n(\pi)\right] \leq \varphi^{-1}(\lambda) + C \cdot A \cdot \left(\varphi(n)^d + n \cdot \left(\frac{\varphi(n)^d \cdot \overline{\log}(n)}{n}\right)^{\min\left\{\frac{\alpha+1}{d+2}, 1\right\}}\right).$$

Theorem 2 follows from the more general Theorem 3 in Section 4. The full proof is given in Appendix D. As with Theorem 1 we may select $\varphi(n) = O(\log n)$ to obtain a regret bound which is minimax optimal up to logarithmic factors. Experiments on synthetic data indicate that the k-NN KL-UCB algorithm typically outperforms the k-NN UCB algorithm, just as the KL-UCB (Garivier and Cappé (2011)) algorithm typically outperforms the standard UCB algorithm (see Section 5).

4. Regret analysis

In order to prove Theorems 1 and 2 we first prove the more general Theorem 3. Suppose we have a k-NN index strategy (Algorithm 1) with index $\mathcal{I}^a_{t,k}$. We shall define for the index strategy a set of good events $\{\mathcal{G}_t\}_{t\in[n]}$ as follows. For each $a \in [A]$, $t \in [n]$ and $k \in [t-1]$ we define the event

$$\mathcal{G}^a_{t,k} := \{\varphi(t) \geq \lambda\} \cap \left\{\mathcal{I}^a_{t,k}(X_t) - 2 \cdot U^a_{t,k}(X_t) \leq f^a(X_t) \leq \mathcal{I}^a_{t,k}(X_t)\right\}.$$

Let $\mathcal{G}_t := \bigcap_{a\in[A]} \bigcap_{k\in[t-1]} \mathcal{G}^a_{t,k}$.

Theorem 3 *Suppose that Assumption 1 holds with constants $R_{\mathcal{X}}, C_d, d$, Assumption 2 holds with Lipschitz constant λ and Assumption 3 holds with constants $\delta_\alpha, C_\alpha, \alpha > 0$. Suppose π is a k-NN index strategy (Algorithm 1) with index $\mathcal{I}^a_{t,k}$. Then there exists a constant C, depending solely upon $R_{\mathcal{X}}, C_d, d, \delta_\alpha, C_\alpha, \alpha$ such that for all $n \in \mathbb{N}$ we have*

$$\mathbb{E}\left[R_n(\pi)\right] \leq C \cdot A \cdot \left(M \cdot \varphi(n)^d + n \cdot \left(\frac{\theta \cdot \varphi(n)^d \cdot \overline{\log}(n)}{n}\right)^{\min\left\{\frac{\alpha+1}{d+2}, 1\right\}}\right) + M \cdot \sum_{t\in[n]} \left(1 - \mathbb{P}\left[\mathcal{G}_t\right]\right).$$

Theorems 1 and 2 are deduced from Theorem 3 in Appendices C and D, respectively. In both cases, the deduction amounts to using concentration inequalities to show that the good events $\mathcal{G}_t$ hold with high probability. The proof of Theorem 3 consists of two primary components. Firstly, we prove an upper bound on the number of times an arm is pulled with covariates in a given region of the metric space with a sufficiently high local margin (see Lemma 3). A key difference with the regret bounds of (Rigollet and Zeevi (2010), Perchet et al. (2013)) is that these local bounds hold for arbitrary subsets, rather than just the members of the partition constructed by the algorithm. Secondly, we construct a partition of the covariate space based on local values of the margin, with regions of low margin partitioned into smaller pieces (see the proof of Proposition 1). The local upper bound is then applied to members of the partition to derive the regret bound.

Given a subset $B \subseteq \mathcal{X}$ and $a \in [A]$ we define $\Delta^a(B) := \sup_{x \in B} \{\Delta^a(x)\}$ and let

$$T_n^a(\pi, B) := \sum_{t \in [n]} \mathbb{1}\{\mathcal{G}_t\} \cdot \mathbb{1}\{X_t \in B\} \cdot \mathbb{1}\{\pi_t = a\}$$

$$\tilde{R}_n^a(\pi, B) := \sum_{t \in [n]} \mathbb{1}\{\mathcal{G}_t\} \cdot \mathbb{1}\{X_t \in B\} \cdot \mathbb{1}\{\pi_t = a\} \cdot \left(Y_t^{\pi_t^*} - Y_t^{\pi_t}\right).$$

Lemma 1: $\mathbb{E}[R_n(\pi)] \leq \sum_{a \in [A]} \mathbb{E}\left[\tilde{R}_n^a(\pi, \mathcal{X})\right] + M \cdot \sum_{t \in [n]} (1 - \mathbb{P}[\mathcal{G}_t])$.

Proof See Appendix F. ■

In light of Lemma 1, in order to prove Theorem 3 it suffices to prove the following proposition (Proposition 1).

Proposition 1: There exists a constant C, depending solely upon $R_{\mathcal{X}}, C_d, d, \delta_\alpha, C_\alpha, \alpha > 0$ such that for all $n \in \mathbb{N}$ we have

$$\mathbb{E}\left[\tilde{R}_n^a(\pi, \mathcal{X})\right] \leq C \cdot \left(M \cdot \varphi(n)^d + n \cdot \left(\frac{\theta \cdot \varphi(n)^d \cdot \overline{\log}(n)}{n}\right)^{\min\left\{\frac{\alpha+1}{d+2}, 1\right\}}\right).$$

Before proving Proposition 1 we require three lemmas (2, 3 and 4 below).

Lemma 2: For any $B \subseteq \mathcal{X}$ and $a \in [A]$ we have $\mathbb{E}\left[\tilde{R}_n^a(\pi, B)\right] \leq \Delta^a(B) \cdot \mathbb{E}[T_n^a(\pi, B)]$.

Proof See Appendix F. ■

The following key lemma bounds the number of times an arm is pulled in a given region of the covariate space.

Lemma 3: Given a subset $B \subseteq \mathcal{X}$ and an arm $a \in [A]$ with $4 \cdot \varphi(n) \cdot \mathrm{diam}(B) < \Delta^a(B)$, the following holds almost surely

$$T_n^a(\pi, B) \leq \frac{4\theta \cdot \overline{\log n}}{\left(\Delta^a(B) - 4 \cdot \varphi(n) \cdot \mathrm{diam}(B)\right)^2} + 1.$$

Proof Clearly we can assume that $T_n^a(\pi, B) > 1$. We define

$$t := \max\{s \in [n] : X_s \in B,\ \pi_s = a,\ \mathcal{G}_s \text{ holds}\}$$
$$k(B) := \max\left\{q \in [t-1] : X_{\tau_{t,q}(X_t)} \in B\right\}.$$

Note that as $\mathcal{G}_t$ holds we must have $\lambda \leq \varphi(t) \leq \varphi(n)$. Since $X_t \in B$ and $X_{\tau_{t,k(B)}(X_t)} \in B$ we must have $r_{t,k(B)}(X_t) \leq \text{diam}(B)$. Moreover, given any $s \in [t-1]$ with $X_s \in B$ we must have $\tau_{t,q}(X_t) = s$ for some $q \leq k(B)$. Thus, $T_n^a(\pi, B) \leq N_{t,k(B)}^a(X_t) + 1$.

Note that $T_n^a(\pi, B) > 1$ implies $t > A$. Choose $z_* \in [A]$ so that $f^{z_*}(X_t) = f^*(X_t)$. Since $\pi_t = a$ and $t > A$ we have $\mathcal{I}_{t,k_t(z_*)}^{z_*}(X_t) \leq \mathcal{I}_{t,k_t(a)}^a(X_t)$. On the other hand, since $\mathcal{G}_t$ holds we have, $f^{z_*}(X_t) \leq \mathcal{I}_{t,k_t(z_*)}^{z_*}(X_t)$ and $f^a(X_t) \geq \mathcal{I}_{t,k_t(a)}^a(X_t) - 2 \cdot U_{t,k_t(a)}^a(X_t)$. Thus, given above and the definitions of $k_t(a)$ and $U_{t,k}^a$ we have

$$\begin{aligned}(f^{z_*}(X_t) - f^a(X_t))/2 &\leq U_{t,k_t(a)}^a(X_t) \leq U_{t,k(B)}^a(X_t) \\ &= \sqrt{(\theta \log t)/N_{t,k(B)}^a(X_t)} + \varphi(t) \cdot r_{t,k(B)}(X_t) \\ &\leq \sqrt{(\theta \log t)/(T_n^a(\pi, B) - 1)} + \varphi(n) \cdot \text{diam}(B).\end{aligned}$$

By the Lipschitz assumption (Assumption 2) together and given $X_t \in B$ we must have

$$f^{z_*}(X_t) - f^a(X_t) \geq \Delta^a(B) - 2\lambda \cdot \text{diam}(B) \geq \Delta^a(B) - 2\varphi(n) \cdot \text{diam}(B).$$

Combining with the above proves the lemma. ■

Lemma 4 applies Assumption 1 to obtain an analogue of nested hyper-cubes within $[0,1]^d$. The proof adapts ideas from geometric measure theory (Käenmäki et al. (2012)).

Lemma 4: Suppose that Assumption 1 holds. Given $q \in \mathbb{N}\backslash\{0\}$, $\delta \in (0, R_{\mathcal{X}}]$ and $r \in (0, 1/3)$ there exists a finite collection of subsets $\{Z_{l,i} : l \in [q], i \in [m_l]\}$ which satisfies:

1. For each $l \in [q]$, $\{Z_{l,i}\}_{i \in [m_l]}$ is a partition of $\mathcal{X}$.
2. Given $l_1, l_2 \in [q]$ with $l_1 \leq l_2$, $i_1 \in [m_{l_1}]$ and $i_2 \in [m_{l_2}]$, either $Z_{l_1,i_1} \cap Z_{i_2,l_2} = \emptyset$ or $Z_{l_2,i_2} \subseteq Z_{l_1,i_1}$.
3. For all $l \in [q]$, $i \in [m_l]$ we have $\text{diam}(Z_{l,i}) \leq \delta \cdot r^l$ and $\mu(Z_{l,i}) \geq C_d \cdot \left((\delta/4) \cdot (1-3r) \cdot r^l\right)^d$.

Proof See Appendix G. ■

We are now ready to complete the proof of Proposition 1, which entails Theorem 3.

Proof [Proof of Proposition 1] Throughout the proof $c_1, \cdots, c_7$ will denote constants depending solely upon $R_{\mathcal{X}}, C_d, d, \delta_\alpha, C_\alpha, \alpha$. We shall apply Lemma 4 to construct a cover of $\mathcal{X}$ based upon the local value of Δ^a. First let $\delta(n) := \min\{R_{\mathcal{X}}, \delta_\alpha/(10 \cdot \varphi(n))\}$. Take some $q \in \mathbb{N}$ (to be specified later), let $\delta = \delta(n)$ and $r = 1/4$ and let $\{Z_{l,i} : l \in [q], i \in [m_l]\}$ be a collection of subsets satisfying properties (1),(2),(3) from Lemma 4. In particular, for all $l \in [q]$ and $i \in [m_l]$ we have $\text{diam}(Z_{l,i}) \leq \delta(n) \cdot 4^{-l}$ and $\mu(Z_{l,i}) \geq C_d \cdot (\delta(n)/16)^d \cdot 4^{-ld}$. First let $\mathcal{Z}_{\text{big}}^a := \{Z_{1,i} : i \in [m_1],\ \ \Delta^a(Z_{1,i}) \geq 5 \cdot \varphi(n) \cdot \delta(n)\}$. For each $l \in [q]$ we define

$$\mathcal{Z}_l^a := \left\{Z_{l,i} : i \in [m_l],\ \ 5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-l} \leq \Delta^a(Z_{l,i}) < 5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-l+1}\right\}.$$

Finally, define

$$Z^a_{\text{small}} := \left\{x \in \mathcal{X} : 0 < \Delta^a(x) < 5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-q}\right\}$$
$$Z^a_0 := \left\{x \in \mathcal{X} : \Delta^a(x) = 0\right\}.$$

We claim that for all $r \in [q]$ we have

$$\mathcal{X} \subseteq \bigcup \left(\mathcal{Z}^a_{\text{big}} \cup \left(\bigcup_{l \in [r]} \mathcal{Z}^a_l \right) \cup \left\{ Z_{r,i} : i \in [m_r],\ \ \Delta^a(Z_{r,i}) < 5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-r} \right\} \right).$$

For $r = 1$ the claim follows straightforwardly from the fact that $\{Z_{1,i}\}_{i \in [m_1]}$ is a partition of $\mathcal{X}$. Now suppose the claim holds for some $r \in [q-1]$. By properties (1) and (2) in Lemma 4 for any $i \in [m_r]$,

$$Z_{r,i} = \bigcup \left\{ Z_{r+1,j} : j \in [m_{r+1}],\ Z_{r+1,j} \subseteq Z_{r,i} \right\}.$$

Moreover, if $Z_{r+1,j} \subseteq Z_{r,i}$ then $\Delta^a\left(Z_{r+1,j}\right) \leq \Delta^a\left(Z_{r,i}\right)$. Thus, we have

$$\begin{aligned}
&\bigcup \left\{ Z_{r,i} : i \in [m_r],\ \ \Delta^a(Z_{r,i}) < 5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-r} \right\} \\
&\quad \subseteq \bigcup \left\{ Z_{r+1,i} : i \in [m_{r+1}],\ \ \Delta^a(Z_{r+1,i}) < 5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-r} \right\} \\
&\quad = \bigcup \left(\mathcal{Z}^a_{r+1} \cup \left\{ Z_{r+1,i} : i \in [m_{r+1}],\ \ \Delta^a(Z_{r+1,i}) < 5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-r-1} \right\} \right).
\end{aligned}$$

Hence, given that the claim holds for r it must also hold for $r+1$. From the special case where $r = q$ we deduce that,

$$\mathcal{X} \subseteq \bigcup \left(\mathcal{Z}^a_{\text{big}} \cup \left(\bigcup_{l \in [q]} \mathcal{Z}^a_l \right) \cup \left\{ Z^a_{\text{small}}, Z^a_0 \right\} \right).$$

Thus, given that $\mathbb{E}\left[\tilde{R}^a_n\left(\pi, Z^a_0\right)\right] = 0$ we have

$$\mathbb{E}\left[\tilde{R}^a_n\left(\pi, \mathcal{X}\right)\right] \leq \sum_{Z \in \mathcal{Z}^a_{\text{big}}} \mathbb{E}\left[\tilde{R}^a_n\left(\pi, Z\right)\right] + \sum_{l=1}^{q} \sum_{Z \in \mathcal{Z}^a_l} \mathbb{E}\left[\tilde{R}^a_n\left(\pi, Z\right)\right] + \mathbb{E}\left[\tilde{R}^a_n\left(\pi, Z^a_{\text{small}}\right)\right].$$

We begin by considering $\sum_{Z \in \mathcal{Z}^a_{\text{big}}} \mathbb{E}\left[\tilde{R}^a_n\left(\pi, Z\right)\right]$. Given $Z \in \mathcal{Z}^a_{\text{big}}$ we have $\operatorname{diam}(Z) \leq \delta(n)/4$, $5 \cdot \varphi(n) \cdot \delta(n) \leq \Delta^a(Z) \leq M$ and $\mu(Z) \geq C_d \cdot (\delta(n)/64)^d$. By Lemmas 2 and 3 we have

$$\begin{aligned}
\mathbb{E}\left[\tilde{R}^a_n\left(\pi, Z\right)\right] &\leq \Delta^a(Z) \cdot \left(\frac{4\theta \cdot \overline{\log}(n)}{\left(\Delta^a(Z) - 4 \cdot \varphi(n) \cdot \operatorname{diam}(Z)\right)^2} + 1 \right) \\
&\leq \frac{5\theta \cdot \overline{\log}(n)}{4 \cdot \varphi(n) \cdot \delta(n)} + M.
\end{aligned}$$

Moreover, since $\mu(Z) \geq C_d \cdot (\delta(n)/64)^d$ for $Z \in \mathcal{Z}^a_{\text{big}}$, we have $\#\mathcal{Z}^a_{\text{big}} \leq C_d^{-1} \cdot (\delta(n)/64)^{-d}$. Hence,

$$\sum_{Z \in \mathcal{Z}^a_{\text{big}}} \mathbb{E}\left[\tilde{R}^a_n(\pi, Z)\right] \leq c_1 \cdot \varphi(n)^d \cdot \left(\theta \cdot \overline{\log}(n) + M\right). \tag{3}$$

Now take $l \in [q]$ and consider $Z \in \mathcal{Z}^a_l$. We have $\operatorname{diam}(Z) \leq \delta(n) \cdot 4^{-l}$,

$$5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-l} \leq \Delta^a(Z) < 5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-l+1}$$

and $\mu(Z) \geq C_d \cdot (\delta(n)/16)^d \cdot 4^{-ld}$. Hence, by Lemma 3 we have

$$T^a_n(\pi, Z) \leq \frac{\theta \cdot \overline{\log}(n)}{(\varphi(n) \cdot \delta(n))^2} \cdot 4^{2l+1} + 1.$$

Combining with Lemma 2 and $\Delta^a(Z) < 5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-l+1}$ we have

$$\mathbb{E}\left[\tilde{R}^a_n(\pi, Z)\right] \leq c_2 \cdot \theta \cdot \overline{\log}(n) \cdot 4^l.$$

Moreover, it follows from the definition of $\delta(n)$ that for all $Z \in \mathcal{Z}^a_l$ we have $\Delta^a(Z) < \delta_\alpha$. Hence, by Assumption 3 we have

$$\#\mathcal{Z}^a_l \cdot C_d \cdot (\delta(n)/16)^d \cdot 4^{-ld} \leq \sum_{Z \in \mathcal{Z}^a_l} \mu(Z) \leq C_\alpha \cdot \left(5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-l+1}\right)^\alpha.$$

Thus, we have

$$\sum_{Z \in \mathcal{Z}^a_l} \mathbb{E}\left[\tilde{R}^a_n(\pi, Z)\right] \leq c_3 \cdot \varphi(n)^d \cdot \theta \cdot \overline{\log}(n) \cdot 4^{l(d+1-\alpha)}. \tag{4}$$

Finally, $\Delta(Z^a_{\text{small}}) \leq 5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-q}$. Hence, by Assumption 3 we have $\mu(Z^a_{\text{small}}) \leq C_\alpha \cdot (5 \cdot \varphi(n) \cdot \delta(n))^\alpha \cdot 4^{-q\alpha}$. Hence, by Lemma 2 we have

$$\begin{aligned}\mathbb{E}\left[\tilde{R}^a_n(\pi, Z^a_{\text{small}})\right] &\leq \left(5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-q}\right) \cdot \mathbb{E}\left[T^a_n(\pi, Z^a_{\text{small}})\right] \\ &\leq \left(5 \cdot \varphi(n) \cdot \delta(n) \cdot 4^{-q}\right) \cdot n \cdot \mu(Z^a_{\text{small}}) \leq c_4 \cdot n \cdot 4^{-q(\alpha+1)}. \end{aligned} \tag{5}$$

Combining equations (3), (4) and (5) we have

$$\begin{aligned}\mathbb{E}\left[\tilde{R}^a_n(\pi, \mathcal{X})\right] &\leq c_5 \cdot \left(\varphi(n)^d \left(M + \theta \cdot \overline{\log}(n) \cdot \sum_{l=0}^{q} 4^{l(d+1-\alpha)}\right) + n \cdot 4^{-q(\alpha+1)}\right) \\ &\leq c_6 \cdot \left(\varphi(n)^d \left(M + \theta \cdot \overline{\log}(n) \cdot (1 + 4^{q(d+1-\alpha)})\right) + n \cdot 4^{-q(\alpha+1)}\right).\end{aligned}$$

Thus, if we take $q = \lceil \log\left(n / \left(\theta \cdot \varphi(n)^d \cdot \overline{\log}(n)\right)\right) / ((d+2)\log 4) \rceil$ we have

$$\begin{aligned}\mathbb{E}\left[\tilde{R}^a_n(\pi, \mathcal{X})\right] &\leq c_7 \cdot \left((M + \theta \cdot \overline{\log}(n)) \cdot \varphi(n)^d + n \cdot \left(\frac{\theta \cdot \varphi(n)^d \cdot \overline{\log}(n)}{n}\right)^{\frac{\alpha+1}{d+2}}\right) \\ &\leq c_8 \cdot \left(M \cdot \varphi(n)^d + n \cdot \left(\frac{\theta \cdot \varphi(n)^d \cdot \overline{\log}(n)}{n}\right)^{\min\left\{\frac{\alpha+1}{d+2}, 1\right\}}\right)\end{aligned}$$

■

5. Experimental results

We conduct an empirical illustration of the ability of both the K-NN UCB algorithm and the K-NN KL-UCB to adapt to the intrinsic dimensionality of the data. We consider four bandit scenarios. In each bandit scenario the marginal distribution over the covariates is supported on a $d = 2$ dimensional affine sub-manifold within $\mathbb{R}^D$. We vary the dimension of the ambient feature space D over the four scenarios so $D \in \{2, 5, 10, 15\}$. We compare four algorithms: The UCBogram (Rigollet and Zeevi (2010)), the ABSE algorithm (Perchet et al. (2013)), the K-NN UCB algorithm and the K-NN KL-UCB algorithm. For further details on experimental procedure and the generation of the synthetic data we refer to Appendix B.

The results are displayed in Figure 1 (Appendix A). When $d = D = 2$ the UCBogram performs comparably with the KNN based algorithms. However, the performance of the UCBogram deteriorates as we increase the dimension of the ambient feature space D. However, both the KNN UCB and the KNN KL-UCB algorithm are robust to increases in the ambient dimension D, significantly outperforming both the ABSE algorithm and the UCBogram when $D = 15$. This gives an empirical illustration of the fact that bounds in Theorem 1 and 2 do not depend upon the dimensionality of ambient feature space.

6. Discussion

We have presented the k-NN UCB algorithm for multi-armed bandits with covariates. The method is conceptually simple, and unlike previous methods, such as the UCBogram and Adaptively Binned Successive Elimination, the k-NN UCB algorithm does not require prior knowledge of either the time horizon or the intrinsic dimension of the marginal distribution over the covariates. We have proven two regret bounds. Theorem 1 demonstrates that the k-NN UCB algorithm is minimax optimal, up to logarithmic factors whenever the noise is subgaussian. Theorem 2 demonstrates that the k-NN KL-UCB algorithm is minimax optimal, up to logarithmic factors, in the bounded rewards setting. Overall, we see that both k-NN based algorithms automatically take advantage of both low intrinsic dimensionality of the marginal distribution over the covariates and low noise in the data, expressed as a margin condition. In addition we have illustrated the robustness of the k-NN based algorithms to the dimension of the ambient feature space with experimental results on synthetic data.

A challenging open question remains. Is it possible to obtain algorithms which are adaptive to an unknown Hölder exponent? Both the k-NN UCB and the k-NN KL-UCB algorithm may be straightforwardly adapted to fixed Hölder exponents $\beta < 1$ (in place of the Lipschitz assumption), provided the exponent is known a priori. However, it remains to be seen whether or not it is possible to construct an algorithm which automatically adapts to an unknown Hölder exponent.

Acknowledgements

The authors gratefully acknowledge the support of the EPSRC for the LAMBDA project (EP/N035127/1) and the Manchester Centre for Doctoral Training in Computer Science (EP/1038099/1). We would also like to thank Ata Kabán, Peter Tino, Jeremy Wyatt, Konstantinos Sechidis, Nikos Nikolaou and Sarah Nogueira for useful discussions. We would also like to thank the anonymous reviewers for their careful feedback.

References

Rajeev Agrawal. Sample mean based index policies by o (log n) regret for the multi-armed bandit problem. *Advances in Applied Probability*, 27(4):1054–1078, 1995.

Jean-Yves Audibert, Alexandre B Tsybakov, et al. Fast learning rates for plug-in classifiers. *The Annals of statistics*, 35(2):608–633, 2007.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

Alina Beygelzimer, John Langford, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandit algorithms with supervised learning guarantees. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 19–26, 2011.

Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1): 1–122, 2012.

Kamalika Chaudhuri and Sanjoy Dasgupta. Rates of convergence for nearest neighbor classification. In *Advances in Neural Information Processing Systems*, pages 3437–3445, 2014.

Armin Eftekhari and Michael B Wakin. New analysis of manifold embeddings and signal recovery from compressive measurements. *Applied and Computational Harmonic Analysis*, 39(1):67–109, 2015.

Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491, 1959.

Charlie Frogner, Chiyuan Zhang, Hossein Mobahi, Mauricio Araya, and Tomaso A Poggio. Learning with a wasserstein loss. In *Advances in Neural Information Processing Systems*, pages 2053–2061, 2015.

Aurélien Garivier and Olivier Cappé. The kl-ucb algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th annual Conference On Learning Theory*, pages 359–376, 2011.

Antti Käenmäki, Tapio Rajala, and Ville Suomala. Existence of doubling measures via generalised nested cubes. *Proceedings of the American Mathematical Society*, 140(9): 3275–3281, 2012.

Sham M Kakade, Shai Shalev-Shwartz, and Ambuj Tewari. Efficient bandit algorithms for online multiclass prediction. In *Proceedings of the 25th international conference on Machine learning*, pages 440–447. ACM, 2008.

Samory Kpotufe. k-nn regression adapts to local intrinsic dimension. In *Advances in Neural Information Processing Systems*, pages 729–737, 2011.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

John Langford and Tong Zhang. The epoch-greedy algorithm for multi-armed bandits with side information. In *Advances in neural information processing systems*, pages 817–824, 2008.

John M Lee. *Riemannian manifolds: an introduction to curvature*, volume 176. Springer Science & Business Media, 2006.

Ulrike von Luxburg and Olivier Bousquet. Distance-based classification with lipschitz functions. *Journal of Machine Learning Research*, 5(Jun):669–695, 2004.

Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.

Vianney Perchet, Philippe Rigollet, et al. The multi-armed bandit problem with covariates. *The Annals of Statistics*, 41(2):693–721, 2013.

Wei Qian and Yuhong Yang. Kernel estimation and model combination in a bandit problem with covariates. *Journal of Machine Learning Research*, 2016.

Henry W J Reeve and Gavin Brown. Minimax rates for cost-sensitive learning on manifolds with approximate nearest neighbours. In *Proceedings of the 28th International Conference on Algorithmic Learning Theory*, volume 76, pages 11–56, 2017.

Philippe Rigollet and Assaf Zeevi. Nonparametric bandits with covariates. *COLT 2010*, page 54, 2010.

Yevgeny Seldin, Peter Auer, John S Shawe-Taylor, Ronald Ortner, and François Laviolette. Pac-bayesian analysis of contextual bandits. In *Advances in Neural Information Processing Systems*, pages 1683–1691, 2011.

Aleksandrs Slivkins. Contextual bandits with similarity information. In *Proceedings of the 24th annual Conference On Learning Theory*, pages 679–702, 2011.

Alexandre B Tsybakov. Optimal aggregation of classifiers in statistical learning. *Annals of Statistics*, pages 135–166, 2004.

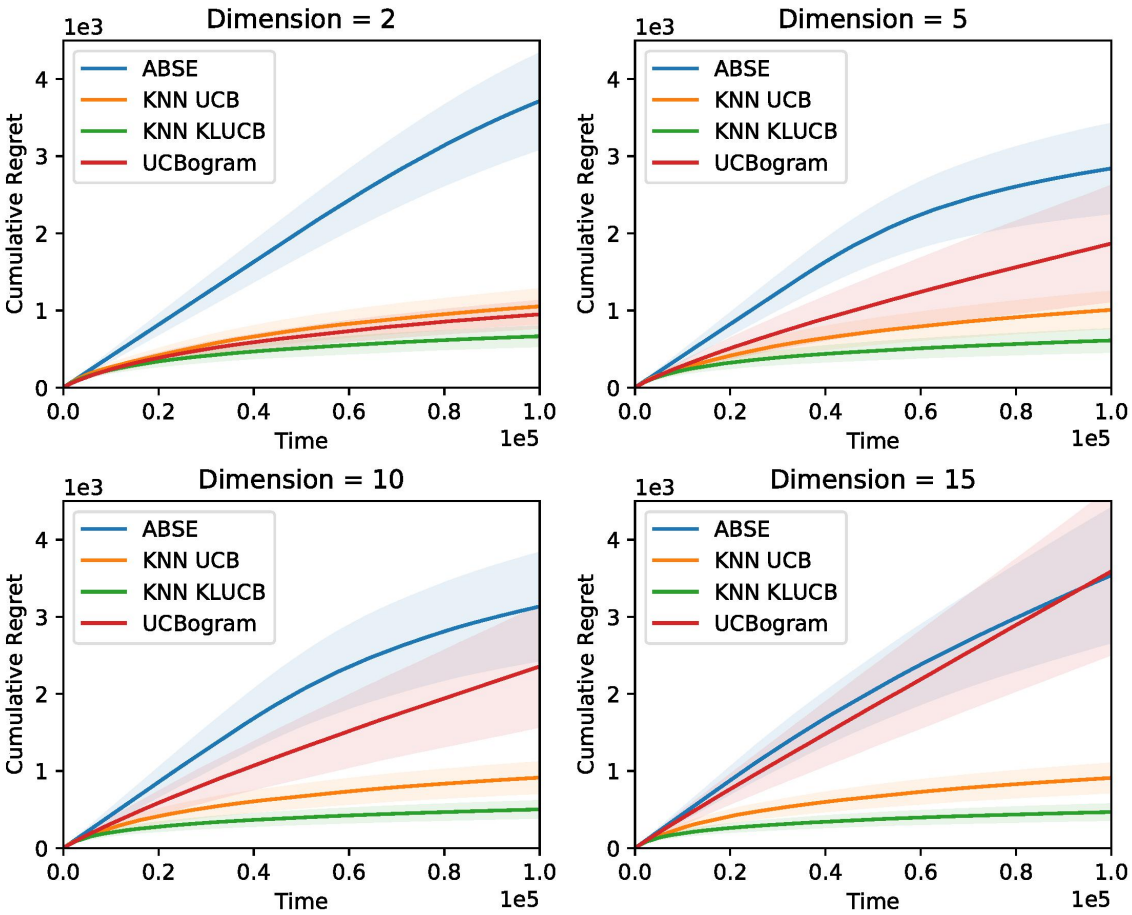
Chih-Chun Wang, Sanjeev R Kulkarni, and H Vincent Poor. Arbitrary side observations in bandit problems. *Advances in Applied Mathematics*, 34(4):903–938, 2005a.

Chih-Chun Wang, Sanjeev R Kulkarni, and H Vincent Poor. Bandit problems with side observations. *IEEE Transactions on Automatic Control*, 50(3):338–355, 2005b.

Yuhong Yang, Dan Zhu, et al. Randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates. *The Annals of Statistics*, 30(1):100–121, 2002.

Appendix A. Empirical results

Figure 1: A comparison of four algorithms for multi-armed bandits with covariates: the UCBogram, the ABSE algorithm, the KNN UCB algorithm and the KNN KL-UCB algorithm. In each experiment the covariates are supported on a $d = 2$-dimensional sub-manifold $\mathcal{M} \subset \mathbb{R}^D$ with the dimension of the ambient space varied $D \in \{2, 5, 10, 15\}$. For each algorithm in each scenario we plot the mean and standard deviation over fifty runs. For further discussion see Section 5 .



Appendix B. Experimental Procedure

In this section we give a detailed account of the experimental procedure for the empirical results in Section 5. We constructed a synthetic A-armed bandit problem with covariates on a d-dimensional sub-manifold within a D-dimensional feature space $\mathbb{R}^D$ as follows:

To construct a marginal distributions μ on $\mathbb{R}^D$, supported on a d-dimensional sub-manifold $\mathcal{M} \subset \mathbb{R}^D$, we first construct an affine map $\phi : \mathbb{R}^d \to \mathbb{R}^D$. We do this by randomly generating d orthonormal D-dimensional vectors $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_d\} \subset \mathbb{R}^D$, letting $\tilde{\phi}(z) = \sum_{l=1}^{d} z_l \cdot \boldsymbol{u}_l$ for $z = (z_l)_{l=1}^{d} \in \mathbb{R}^d$, and letting $\phi(z) = \tau \circ \tilde{\phi}$, where $\tau : \mathbb{R}^D \to \mathbb{R}^D$ is a similarity mapping such that $\phi\left([0,1]^d\right) \subseteq [0,1]^D$. It follows that $\mathcal{M} := \phi\left([0,1]^d\right) \subseteq [0,1]^D$ is a compact subset of the d-dimensional manifold $\phi(\mathbb{R}^d)$ and the inverse $\phi^{-1} : \mathcal{M} \to [0,1]^d$ is well-defined. We construct a measure $\tilde{\mu}$ on $[0,1]^d$ by taking

$$S := \bigcup_{\boldsymbol{\omega} \in [5]^d} \left\{ z \in \mathbb{R}^d : \|10 \cdot z - 2\boldsymbol{\omega} + \mathbf{1}_d\|_\infty \leq 1/5 \right\},$$

and letting $\tilde{\mu}$ be the uniform measure on S. We then obtain μ supported on $\mathcal{M}$ by $\mu := \tilde{\mu} \circ \phi^{-1}$. The use of S in the construction of μ ensures that that we have large margins $\Delta(x)$ with high probability. We construct reward functions f^a for each arm $a \in [A]$ as follows: First we define a kernel function $h : \mathbb{R}^d \to \mathbb{R}$ by $h(x) := \max\{1 - \|x\|_\infty, 0\}$. For each arm $a \in [A]$, and each vector $\boldsymbol{\omega} \in [5]^d$ we select $\zeta^a(\boldsymbol{\omega}) \in \{-1, +1\}$ randomly (i.i.d with probability 0.5). We construct a Lipschitz function $\tilde{f}^a : [0,1]^d \to [0,1]$ by

$$\tilde{f}^a(z) := \frac{1}{2} + \frac{1}{10} \cdot \sum_{\omega \in [5]^d} \zeta^a(\boldsymbol{\omega}) \cdot h\left(10z - 2\boldsymbol{\omega} + \mathbf{1}_d\right),$$

where $\mathbf{1}_d$ is a d-dimensional vector consisting entirely of ones. Finally, we define $f^a : \mathcal{M} \to [0,1]$ by $f^a(x) = \tilde{f}^a\left(\phi^{-1}(x)\right)$. We generate $(X, Y^1, \cdots, Y^A) \in \mathcal{M} \times \{0,1\}^A$ by $X \sim \mu$ and $\mathbb{E}\left[Y^a | X = x\right] = f^a(X)$.

In our experiments we consider four bandit scenarios and four algorithms. In each bandit scenario the dimension of the manifold $d = 2$ and the number of arms $A = 2$. We vary the dimension of the ambient feature space D over the scenarios by taking $D \in \{2, 5, 10, 15\}$. We compare four algorithms: The UCBogram Rigollet and Zeevi (2010), the ABSE algorithm Perchet et al. (2013), the K-NN UCB algorithm with $\theta = 2$, $\varphi \equiv 1$ and the K-NN KL-UCB algorithm with $\theta = 1$, $\varphi \equiv 1$. For each of the sixteen combinations of bandit scenario and algorithm we conduct fifty runs, with fifty different random seeds, each time with a horizon of one hundred thousand.

Appendix C. Proof of Theorem 1

In this section we will deduce Theorem 1 from Theorem 3.

Theorem 1 *Suppose that Assumption 1 holds with constants $R_{\mathcal{X}}, C_d, d$, Assumption 2 holds with Lipschitz constant λ, Assumption 3 holds with constants $\delta_\alpha, C_\alpha, \alpha > 0$ and Assumption 4 holds. Let π be the k-NN UCB algorithm (Algorithm 1 with $\mathcal{I}^a_{t,k}$ as in equation (1)). Then*

for all $\theta > 4$ there exists a constant C, depending solely upon $R_{\mathcal{X}}, C_d, d, \delta_\alpha, C_\alpha, \alpha$ and θ such that for all $n \in \mathbb{N}$ we have

$$\mathbb{E}\left[R_n(\pi)\right] \leq M \cdot \varphi^{-1}(\lambda) + C \cdot A \cdot \left(M \cdot \varphi(n)^d + n \cdot \left(\frac{\varphi(n)^d \cdot \overline{\log}(n)}{n} \right)^{\min\left\{\frac{\alpha+1}{d+2}, 1\right\}} \right).$$

Lemma 5: Suppose that Assumption 2 holds with Lipschitz constant λ and Assumption 4 holds. Let π be the k-NN UCB algorithm (Algorithm 1 with $\mathcal{I}^a_{t,k}$ as in equation (1)). Then for all $\theta > 0$, all $a \in [A]$, $t \in \{\varphi^{-1}(\lambda), \cdots, n\}$ and $k \in [t-1]$ we have

$$1 - \mathbb{P}\left[\mathcal{G}^a_{t,k}\right] \leq 2 \cdot e \cdot \lceil (\theta/2) \cdot (\log t)^2 \rceil \cdot t^{-\theta/2}.$$

Proof Recall that

$$\mathcal{G}^a_{t,k} := \{\varphi(t) \geq \lambda\} \cap \left\{\mathcal{I}^a_{t,k}(X_t) - 2 \cdot U^a_{t,k}(X_t) \leq f^a(X_t) \leq \mathcal{I}^a_{t,k}(X_t)\right\}.$$

Hence, for $t \geq \varphi^{-1}(\lambda)$, if $\mathcal{G}^a_{t,k}$ does not hold then

$$\begin{aligned} \left|\hat{f}^a_{t,k}(X_t) - f^a(X_t)\right| > U^a_{t,k}(X_t) &= \sqrt{(\theta \log t) / N^a_{t,k}(X_t)} + \varphi(t) \cdot r_{t,k}(X_t) \\ &\geq \sqrt{(\theta \log t) / N^a_{t,k}(X_t)} + \lambda \cdot r_{t,k}(X_t). \end{aligned}$$

For $s \in [t-1]$ we define

$$\begin{aligned} \epsilon_s &= \mathbb{1}\left\{s \in \Gamma_{t,k}(X_t)\right\} \cdot \mathbb{1}\left\{\pi_s = a\right\} \\ Z_s &= \mathbb{1}\left\{s \in \Gamma_{t,k}(X_t)\right\} \cdot \left(Y^a_s - f^a(X_s)\right). \end{aligned}$$

Hence, $N^a_{t,k}(X_t) = \sum_{s \in [t-1]} \epsilon_s$. By the Lipschitz property (Assumption 2) for all $s \in \Gamma_{t,k}(X_t)$ we have

$$|f^a(X_s) - f^a(X_t)| \leq \lambda \cdot \rho\left(X_s, X_t\right) \leq \lambda \cdot r_{t,k}(X_t).$$

Thus, if $\mathcal{G}^a_{t,k}$ does not hold then

$$\left| \sum_{s \in [t-1]} \epsilon_s \cdot Z_s \right| > \sqrt{(\theta \log t) \cdot N^a_{t,k}(X_t)}.$$

By Corollary 5, Appendix E, for any given $\{X_s\}_{s \in [n]}$ we have

$$\mathbb{P}\left[\sum_{s \in [t-1]} \epsilon_s \cdot Z_s > \sqrt{(\theta \log t) \cdot N^a_{t,k}(X_t)} \;\middle|\; \{X_s\}_{s \in [n]} \right] \leq e \cdot \lceil (\theta/2) \cdot (\log t)^2 \rceil \cdot t^{-\theta/2}.$$

By the law of total expectation this implies

$$\mathbb{P}\left[\sum_{s \in [t-1]} \epsilon_s \cdot Z_s > \sqrt{(\theta \log t) \cdot N^a_{t,k}(X_t)} \right] \leq e \cdot \lceil (\theta/2) \cdot (\log t)^2 \rceil \cdot t^{-\theta/2}.$$

By applying Corollary 5 to $\{-Z_s\}_{s\in[t-1]}$ we also have the lower tail inequality. Hence the lemma holds. ■

Proof [Proof of Theorem 1] By Lemma 5 we have

$$\begin{aligned}\sum_{t\in[n]}\left(1-\mathbb{P}\left[\mathcal{G}_t\right]\right) &\leq \varphi^{-1}(\lambda)+\sum_{a\in[A]}\sum_{t=\varphi^{-1}(\lambda)}^{n}\sum_{k\in[t-1]}\left(1-\mathbb{P}\left[\mathcal{G}_{t,k}^a\right]\right)\\ &\leq \varphi^{-1}(\lambda)+2e\cdot\sum_{a\in[A]}\sum_{t=\varphi^{-1}(\lambda)}^{n}\sum_{k\in[t-1]}\lceil(\theta/2)\cdot(\log t)^2\rceil\cdot t^{-\theta/2}\\ &\leq \varphi^{-1}(\lambda)+2e\cdot A\cdot\sum_{t=1}^{\infty}\lceil(\theta/2)\cdot(\log t)^2\rceil\cdot t^{1-\theta/2}.\end{aligned}$$

Given $\theta>4$ we have $\sum_{t=1}^{\infty}\lceil(\theta/2)\cdot(\log t)^2\rceil\cdot t^{1-\theta/2}<\infty$. Hence, by applying Theorem 3, the regret bound in Theorem 1 holds. ■

Appendix D. Proof of Theorem 2

In this section we will deduce Theorem 2 from Theorem 3.

Theorem 2 *Suppose that Assumption 1 holds with constants $R_{\mathcal{X}}, C_d, d$, Assumption 2 holds with Lipschitz constant λ, Assumption 3 holds with constants $\delta_\alpha, C_\alpha, \alpha>0$ and Assumption 5 holds. Let π be the k-NN KL-UCB algorithm (Algorithm 1 with $\mathcal{I}_{t,k}^a$ as in equation (2)). Then for all $\theta>2$ there exists a constant C, depending solely upon $R_{\mathcal{X}}, C_d, d, \delta_\alpha, C_\alpha, \alpha$ and θ such that for all $n\in\mathbb{N}$ we have*

$$\mathbb{E}\left[R_n(\pi)\right]\leq\varphi^{-1}(\lambda)+C\cdot A\cdot\left(\varphi(n)^d+n\cdot\left(\frac{\varphi(n)^d\cdot\overline{\log}(n)}{n}\right)^{\min\left\{\frac{\alpha+1}{d+2},1\right\}}\right).$$

Lemma 6: Suppose that Assumption 2 holds with Lipschitz constant λ and Assumption 5 holds. Let π be the k-NN KL-UCB algorithm (Algorithm 1 with $\mathcal{I}_{t,k}^a$ as in equation (2)). Then for all $\theta>0$, all $a\in[A]$, $t\in\{\varphi^{-1}(\lambda),\cdots,n\}$ and $k\in[t-1]$ we have

$$1-\mathbb{P}\left[\mathcal{G}_{t,k}^a\right]<2\cdot e\cdot\lceil\theta\cdot(\log t)^2\rceil\cdot t^{-\theta}.$$

Proof Recall that in the k-NN KL-UCB algorithm we have

$$\mathcal{I}_{t,k}^a(x)=\sup\left\{\omega\in[0,1]:N_{t,k}^a(x)\cdot d\left(\hat{f}_{t,k}^a(x),\omega\right)\leq\theta\cdot\log t\right\}+\varphi(t)\cdot r_{t,k}(x).$$

For the purposes of the proof we also define a lower confidence bound,

$$\mathcal{H}_{t,k}^a(x)=\inf\left\{\omega\in[0,1]:N_{t,k}^a(x)\cdot d\left(\hat{f}_{t,k}^a(x),\omega\right)\leq\theta\cdot\log t\right\}-\varphi(t)\cdot r_{t,k}(x).$$

Suppose that $\mathcal{H}^a_{t,k}(X_t) > f^a(X_t)$. Then since $t \geq \varphi^{-1}(\lambda)$ we have

$$N^a_{t,k}(x) \cdot d\left(\hat{f}^a_{t,k}(X_t), f^a(X_t) + \lambda \cdot r_{t,k}(X_t)\right) > \theta \cdot \log t.$$

For $s \in [t-1]$ we define

$$\begin{aligned} \epsilon_s &= \mathbb{1}\left\{s \in \Gamma_{t,k}(X_t)\right\} \cdot \mathbb{1}\left\{\pi_s = a\right\} \\ Z^{\mathcal{H}}_s &= \mathbb{1}\left\{s \in \Gamma_{t,k}(X_t)\right\} \cdot Y^a_s. \end{aligned}$$

Fix $\{X_l\}_{l\in[n]}$. By the Lipschitz property (Assumption 2) for all $s \in \Gamma_{t,k}(X_t)$ we have

$$\mathbb{E}\left[Z^{\mathcal{H}}_s \,\middle|\, \{X_l\}_{l\in[n]}\right] = f^a(X_s) \leq f^a(X_t) + \lambda \cdot \rho\left(X_s, X_t\right).$$

Hence, by Corollary 6, Appendix E we have

$$\begin{aligned} &\mathbb{P}\left[\mathcal{H}^a_{t,k}(X_t) > f^a(X_t) \,\middle|\, \{X_l\}_{l\in[n]}\right] \\ &\leq \mathbb{P}\left[N^a_{t,k}(x) \cdot d\left(\hat{f}^a_{t,k}(X_t), f^a(X_t) + \lambda \cdot r_{t,k}(X_t)\right) > \theta \cdot \log t \,\middle|\, \{X_l\}_{l\in[n]}\right] \\ &\leq e\lceil \theta \cdot \log(t)^2 \rceil t^{-\theta}. \end{aligned}$$

Similarly, by applying Corollary 6 with $\{\epsilon_s\}_{s\in[t-1]}$ and $\left\{Z^{\mathcal{I}}_s\right\}_{s\in[t-1]}$ where $Z^{\mathcal{I}}_s = \mathbb{1}\left\{s \in \Gamma_{t,k}(X_t)\right\} \cdot (1 - Y^a_s)$. We have,

$$\begin{aligned} &\mathbb{P}\left[\mathcal{I}^a_{t,k}(X_t) < f^a(X_t) \,\middle|\, \{X_l\}_{l\in[n]}\right] \\ &\leq \mathbb{P}\left[N^a_{t,k}(x) \cdot d\left(\hat{f}^a_{t,k}(X_t), f^a(X_t) - \lambda \cdot r_{t,k}(X_t)\right) > \theta \cdot \log t \,\middle|\, \{X_l\}_{l\in[n]}\right] \\ &\leq e\lceil \theta \cdot \log(t)^2 \rceil t^{-\theta}. \end{aligned}$$

Thus, by the total law of expectations we have

$$1 - \mathbb{P}\left[\mathcal{H}^a_{t,k}(X_t) \leq f^a(X_t) \leq \mathcal{I}^a_{t,k}(X_t)\right] \leq 2e\lceil \theta \cdot \log(t)^2 \rceil t^{-\theta}.$$

By Pinsker's inequality, if $N^a_{t,k}(x) \cdot d\left(\hat{f}^a_{t,k}(x), \omega\right) \leq \theta \cdot \log t$ then

$$\left|\omega - \hat{f}^a_{t,k}(x)\right| \leq \sqrt{\frac{\theta \cdot \log t}{2 \cdot N^a_{t,k}(x)}}.$$

Hence, we have

$$\mathcal{I}^a_{t,k}(X_t) \leq \hat{f}^a_{t,k}(X_t) + \sqrt{\frac{\theta \cdot \log t}{2 \cdot N^a_{t,k}(x)}} + \varphi(t) \cdot r_{t,k}(X_t) \leq \hat{f}^a_{t,k}(X_t) + U^a_{t,k}(X_t).$$

Similarly, $\mathcal{H}^a_{t,k}(X_t) \geq \hat{f}^a_{t,k}(X_t) - U^a_{t,k}(X_t)$, so $\mathcal{H}^a_{t,k}(X_t) \geq \mathcal{I}^a_{t,k}(X_t) - 2 \cdot U^a_{t,k}(X_t)$. Thus, for $t \geq \varphi^{-1}(\lambda)$, $\mathcal{H}^a_{t,k}(X_t) \leq f^a(X_t) \leq \mathcal{I}^a_{t,k}(X_t)$ implies $\mathcal{G}^a_{t,k}$, so

$$1 - \mathbb{P}\left[\mathcal{G}^a_{t,k}\right] \leq 1 - \mathbb{P}\left[\mathcal{H}^a_{t,k}(X_t) \leq f^a(X_t) \leq \mathcal{I}^a_{t,k}(X_t)\right] \leq 2e\lceil \theta \cdot \log(t)^2 \rceil t^{-\theta}.$$

■

Proof [Proof of Theorem 2] By Lemma 6 we have

$$\begin{aligned}
\sum_{t\in[n]} \left(1 - \mathbb{P}\left[\mathcal{G}_t\right]\right) &\leq \varphi^{-1}(\lambda) + \sum_{a\in[A]} \sum_{t=\varphi^{-1}(\lambda)}^{n} \sum_{k\in[t-1]} \left(1 - \mathbb{P}\left[\mathcal{G}^a_{t,k}\right]\right) \\
&\leq \varphi^{-1}(\lambda) + 2e \cdot \sum_{a\in[A]} \sum_{t=\varphi^{-1}(\lambda)}^{n} \sum_{k\in[t-1]} \lceil \theta \cdot (\log t)^2 \rceil \cdot t^{-\theta} \\
&\leq \varphi^{-1}(\lambda) + 2e \cdot A \cdot \sum_{t=1}^{\infty} \lceil \theta \cdot (\log t)^2 \rceil \cdot t^{1-\theta}.
\end{aligned}$$

Given $\theta > 2$ we have $\sum_{t=1}^{\infty} \lceil \theta \cdot (\log t)^2 \rceil \cdot t^{1-\theta} < \infty$. Hence, by applying Theorem 3, the regret bound in Theorem 2 holds. ■

Appendix E. Concentration Inequalities

The following theorem is closely related to (Garivier and Cappé, 2011, Theorem 11).

Theorem 4 *Let $(Z_t)_{t\in[n]}$ be a sequence of real-valued random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\{\mathcal{F}_t\}_{t\in\{0\}\cup[n]}$ be an increasing sequence of sigma fields such that for each t, $\sigma(Z_1, \cdots, Z_t) \subset \mathcal{F}_t$ and for $s > t$, Z_s is independent from $\mathcal{F}_t$. Let $\{\epsilon_t\}_{t\in[n]}$ be a sequence of Bernoulli random variables such that $\sigma(\epsilon_t) \subset \mathcal{F}_{t-1}$. For each $t \in [n]$ we let*

$$S(t) = \sum_{s\in[t]} \epsilon_s \cdot Z_s, \qquad N(t) = \sum_{s\in[t]} \epsilon_s, \qquad \hat{\xi}_t = S(t)/N(t).$$

Suppose we have a function $\phi : [0, \infty) \to \mathbb{R}$ with the following properties

- *ϕ is twice differentiable with $\phi''(\rho) > 0$ for all $\rho > 0$,*
- *For all $t \in [n]$ and $\rho \geq 0$ we have $\log\left(\mathbb{E}\left[\exp\left(\rho \cdot Z_t\right)\right]\right) \leq \phi(\rho)$,*
- *For all $t \in [n]$ we have $\mathbb{P}\left[Z_t > \lim_{\rho\to\infty} \phi'(\rho)\right] = 0$,*
- *$\phi(0) = 0$.*

We define the Legendre transform $\phi^* : \mathbb{R} \to \mathbb{R}$ by $\phi^*(x) := \sup_{\rho\geq 0}\{\rho \cdot x - \phi(\rho)\}$. For all $\delta > 0$ we have

$$\mathbb{P}\left[N(n) \cdot \phi^*\left(\hat{\xi}_n\right) > \delta\right] \leq e\lceil \delta \log(n) \rceil \exp(-\delta).$$

Proof Since Z_{t+1} is independent from $\mathcal{F}_t$ and $\epsilon_{t+1} \in \{0,1\}$ is $\mathcal{F}_t$ measureable we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(\rho \cdot \epsilon_{t+1} \cdot Z_{t+1}\right)|\mathcal{F}_t\right] &= \mathbb{E}\left[(1-\epsilon_{t+1}) + \epsilon_{t+1} \cdot \exp\left(\rho \cdot Z_{t+1}\right)|\mathcal{F}_t\right] \\ &= (1-\epsilon_{t+1}) + \epsilon_{t+1} \cdot \mathbb{E}\left[\exp\left(\rho \cdot Z_{t+1}\right)\right] \\ &\leq (1-\epsilon_{t+1}) + \epsilon_{t+1} \cdot \exp(\phi(\rho)) = \exp\left(\epsilon_{t+1} \cdot \phi(\rho)\right).\end{aligned}$$

For each $\rho \in \mathbb{R}$ we define $\{W_t^\rho\}_{t\in\{0\}\cup[n]}$ by $W_0^\rho = 1$ and $W_t^\rho := \exp\left(\rho \cdot S(t) - N(t) \cdot \phi(\rho)\right)$. Thus, W_t^ρ is $\mathcal{F}_t$-measureable. Moreover, by the above we have

$$\begin{aligned}\mathbb{E}\left[W_{t+1}^\rho|\mathcal{F}_t\right] &= \mathbb{E}\left[\exp\left(\rho \cdot \epsilon_{t+1} \cdot Z_{t+1} - \epsilon_{t+1} \cdot \phi(\rho)\right) \cdot W_t^\rho|\mathcal{F}_t\right] \\ &= W_t^\rho \cdot \mathbb{E}\left[\exp\left(\rho \cdot \epsilon_{t+1} \cdot Z_{t+1}\right)|\mathcal{F}_t\right] \cdot \exp\left(-\epsilon_{t+1} \cdot \phi(\rho)\right) \leq W_t^\rho.\end{aligned}$$

Hence, $\{W_t^\rho\}_{t\in\{0\}\cup[n]}$ is a super Martingale with respect to $\{\mathcal{F}_t\}_{t\in\{0\}\cup[n]}$.

By considering the derivative $\frac{\partial}{\partial\rho}\left(\rho \cdot x - \phi(\rho)\right) = x - \phi'(\rho)$ and noting that $\phi'' > 0$ on $(0,\infty)$ we see that for all $\rho \geq 0$ we have $\phi^*\left(\phi'(\rho)\right) = \rho \cdot \phi'(\rho) - \phi(\rho)$. In particular, $\phi^*\left(\phi'(0)\right) = 0$ since $\phi(0) = 0$, and for all $\rho > 0$, we have

$$\frac{\partial}{\partial\rho}\left(\phi^*\left(\phi'(\rho)\right)\right) = \rho \cdot \phi''(\rho) > 0.$$

Thus, $\lim_{\rho\to\infty}\phi^*\left(\phi'(\rho)\right) > 0$. Moreover, $(\phi^*) \circ (\phi') : [0,\infty) \to [0, \lim_{\rho\to\infty}\phi^*\left(\phi'(\rho)\right))$ is an increasing bijection.

Now fix $\gamma > 1$, to be determined later, and let

$$\Delta := \min\left\{\lim_{\rho\to\infty}\phi^*\left(\phi'(\rho)\right), \delta\right\} > 0.$$

For each $q \in \{0\} \cup \mathbb{N}$ we let $t_q := \lfloor(\delta/\Delta) \cdot \gamma^q\rfloor$ and let $Q := \lceil\log(n)/\log\gamma\rceil$, so $t_Q \geq n$. Note also that if $N(n) \leq t_0 \leq \delta/\Delta \leq \delta \cdot \left(\lim_{\rho\to\infty}\phi^*\left(\phi'(\rho)\right)\right)^{-1}$ we have $\phi^*(\hat{\xi}_n) \leq \delta/N(n)$ with probability one, since $\mathbb{P}\left[Z_t > \lim_{\rho\to\infty}\phi'(\rho)\right] = 0$ for each $t \in [n]$ and ϕ^* is everywhere non-decreasing.

Hence, if we let

$$A_q := \left\{N(n) \cdot \phi^*\left(\hat{\xi}_n\right) > \delta\right\} \cap \left\{t_{q-1} < N(n) \leq t_q\right\},$$

then we have

$$\mathbb{P}\left[N(n) \cdot \phi^*\left(\hat{\xi}_n\right) > \delta\right] \leq \sum_{q=1}^{Q}\mathbb{P}\left[A_q\right].$$

Now since $0 < \Delta \leq \lim_{\rho\to\infty}\phi^*\left(\phi'(\rho)\right)$ and $\gamma > 1$, for each $q = 1, \cdots, Q$ we may choose $\rho_q \in (0,\infty)$ so that $\phi^*(\phi'(\rho_q)) = \Delta \cdot \gamma^{-q}$.

Hence, if A_q holds then since $t_{q-1} < N(n) \leq t_q$ we must have

$$\phi^*(\phi'(\rho_q)) = \frac{\Delta}{\gamma^q} \leq \frac{\delta}{N(n)} < \frac{\Delta}{\gamma^{q-1}} = \gamma \cdot \phi^*(\phi'(\rho_q)).$$

Thus, as $N(n) \cdot \phi^*(\hat{\xi}_n) > \delta$ and ϕ^* is non-decreasing we have $\hat{\xi}_n > \phi'(\rho_q)$. Thus,

$$\begin{aligned}\rho_q \cdot S(n) - N(n) \cdot \phi(\rho_q) &= N(n) \cdot \Big(\rho_q \cdot \hat{\xi}_n - \phi(\rho_q)\Big) \\ &\geq N(n) \cdot \left(\rho_q \cdot \phi'(\rho_q) - \phi(\rho_q)\right) \\ &= N(n) \cdot \phi^*(\phi'(\rho_q)) > \frac{\delta}{\gamma}.\end{aligned}$$

Hence,

$$\mathbb{P}\left[A_q\right] \leq \mathbb{P}\left[W_n^{\rho_q} > \exp\left(\frac{\delta}{\gamma}\right)\right] \leq \mathbb{E}\left[W_n^{\rho_q}\right] \cdot \exp\left(-\frac{\delta}{\gamma}\right) \leq \exp\left(-\frac{\delta}{\gamma}\right),$$

by the super-Martingale property. Hence, for any $\gamma > 1$ we have

$$\mathbb{P}\left[N(n) \cdot \phi^*\left(\hat{\xi}_n\right) > \delta\right] \leq \left\lceil \frac{\log(n)}{\log \gamma} \right\rceil \cdot \exp\left(-\frac{\delta}{\gamma}\right).$$

Taking $\gamma = \delta / (\delta - 1)$ completes the proof. ■

Corollary 5 *Let $(Z_t)_{t\in[n]}$ be a sequence of sub-Gaussian random variables, with $\mathbb{E}\left[\exp\left(\rho \cdot Z_t\right)\right] \leq \exp(\rho^2/2)$ for all $t \in [n]$ and $\rho \in \mathbb{R}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\{\mathcal{F}_t\}_{t\in\{0\}\cup[n]}$ be an increasing sequence of sigma fields such that for each t, $\sigma\left(Z_1, \cdots, Z_t\right) \subset \mathcal{F}_t$ and for $s > t$, Z_s is independent from $\mathcal{F}_t$. Let $\{\epsilon_t\}_{t\in[n]}$ be a sequence of Bernoulli random variables such that $\sigma(\epsilon_t) \subset \mathcal{F}_{t-1}$. For all $\delta > 0$, with $N(n)$ and $\hat{\xi}_n$ as in the statement of Theorem 4, we have*

$$\mathbb{P}\left[S(n) > \sqrt{2\delta \cdot N(n)}\right] \leq e\lceil \delta \log(n) \rceil \exp(-\delta).$$

Proof Apply Theorem 4 with $\phi(\rho) = \rho^2/2$. ■

Corollary 6 *Let $(Z_t)_{t\in[n]}$ be a sequence of random variables in bounded in $[0,1]$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\max_{t\in[n]}\left\{\mathbb{E}\left[Z_t\right]\right\} \leq \xi_{\max}$. Let $\{\mathcal{F}_t\}_{t\in\{0\}\cup[n]}$ be an increasing sequence of sigma fields such that for each t, $\sigma\left(Z_1, \cdots, Z_t\right) \subset \mathcal{F}_t$ and for $s > t$, Z_s is independent from $\mathcal{F}_t$. Let $\{\epsilon_t\}_{t\in[n]}$ be a sequence of Bernoulli random variables such that $\sigma(\epsilon_t) \subset \mathcal{F}_{t-1}$. Let $N(n)$ and $\hat{\xi}_n$ be as in the statement of Theorem 4. For all $\delta > 0$ we have*

$$\mathbb{P}\left[\hat{\xi}_n > \xi_{\max} \ \& \ N(n) \cdot d\left(\hat{\xi}_n, \xi_{\max}\right) > \delta\right] \leq e\lceil \delta \log(n) \rceil \exp(-\delta).$$

Proof Let $\phi(\rho) := \log\left(1 + \xi_{\max} \cdot (\exp(\rho) - 1)\right)$. Note that for any $z \in [\xi_{\max}, 1]$ we have $d(z, \xi_{\max}) = \sup_{\rho\geq 0}\left\{\rho \cdot z - \phi(\rho)\right\}$. Since $\mathbb{E}\left[Z_t\right] \leq \xi_{\max}$, for all $\rho \geq 0$ we have $\mathbb{E}\left[\exp\left(\rho \cdot Z_t\right)\right] \leq 1 + \mathbb{E}\left[Z_t\right] \cdot (\exp(\rho) - 1) \leq \exp(\phi(\rho))$. Hence, the corollary follows from Theorem 4. ■

Appendix F. Local Regret Lemmas

In this section we will prove lemmas 1 and 2.

Lemma 1: $\mathbb{E}\left[R_n(\pi)\right] \leq \sum_{a\in[A]} \mathbb{E}\left[\tilde{R}_n^a(\pi,\mathcal{X})\right] + M \cdot \sum_{t\in[n]} \left(1-\mathbb{P}\left[\mathcal{G}_t\right]\right)$.

Proof We decompose the expected regret as follows

$$\begin{aligned}
R_n(\pi) &= \sum_{t\in[n]} \left(Y_t^{\pi_t^*} - Y_t^{\pi_t}\right) \\
&\leq \sum_{a\in[A]} \left(\sum_{t\in[n]} \mathbb{1}\left\{\mathcal{G}_t\right\} \cdot \mathbb{1}\left\{X_t \in \mathcal{X}\right\} \cdot \mathbb{1}\left\{\pi_t = a\right\} \cdot \left(Y_t^{\pi_t^*} - Y_t^{\pi_t}\right)\right) \\
&\qquad + \sum_{t\in[n]} \left(1 - \mathbb{1}\left\{\mathcal{G}_t\right\}\right) \cdot \left(Y_t^{\pi_t^*} - Y_t^{\pi_t}\right).
\end{aligned}$$

Recall that $M := \sup\left\{\Delta^a(x) : a \in [A],\ \ x \in \mathcal{X}\right\}$. Moreover, given any history $\mathcal{D}_{t-1}$, reward vector X_t and arm π_t we have,

$$\mathbb{E}\left[Y_t^{\pi_t^*} - Y_t^{\pi_t} | \mathcal{D}_{t-1}, X_t, \pi_t\right] = f^*(X_t) - f^{\pi_t}(X_t) \leq M.$$

Hence, $\mathbb{E}\left[Y_t^{\pi_t^*} - Y_t^{\pi_t} | \neg\mathcal{G}_t\right] \leq M$, since $\mathcal{G}_t$ is determined by $\mathcal{D}_{t-1}$, X_t, π_t, by the tower property. Thus, taking expectations in the above decomposition completes the proof of the lemma. ■

Lemma 2: For any $B \subseteq \mathcal{X}$ and $a \in [A]$ we have $\mathbb{E}\left[\tilde{R}_n^a(\pi,B)\right] \leq \Delta^a(B) \cdot \mathbb{E}\left[T_n^a(\pi,B)\right]$.

Proof By the definitions of $\tilde{R}_n^a(\pi,B)$ and $T_n^a(\pi,B)$ we have

$$\begin{aligned}
&\mathbb{E}\left[\tilde{R}_n^a(\pi,B)\right] \\
&= \sum_{t\in[n]} \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}\left\{\mathcal{G}_t\right\} \cdot \mathbb{1}\left\{X_t \in B\right\} \cdot \mathbb{1}\left\{\pi_t = a\right\} \cdot \left(Y_t^{\pi_t^*} - Y_t^{\pi_t}\right) | \mathcal{D}_{t-1}, X_t, \pi_t\right]\right] \\
&= \sum_{t\in[n]} \mathbb{E}\left[\mathbb{1}\left\{\mathcal{G}_t\right\} \cdot \mathbb{1}\left\{X_t \in B\right\} \cdot \mathbb{1}\left\{\pi_t = a\right\} \cdot \mathbb{E}\left[Y_t^{\pi_t^*} - Y_t^{\pi_t} | \mathcal{D}_{t-1}, X_t, \pi_t\right]\right] \\
&= \mathbb{E}\left[\sum_{t\in[n]} \mathbb{1}\left\{\mathcal{G}_t\right\} \cdot \mathbb{1}\left\{X_t \in B\right\} \cdot \mathbb{1}\left\{\pi_t = a\right\} \cdot \left(f^*(X_t) - f^a(X_t)\right)\right] \\
&\leq \Delta^a(B) \cdot \mathbb{E}\left[\sum_{t\in[n]} \mathbb{1}\left\{\mathcal{G}_t\right\} \cdot \mathbb{1}\left\{X_t \in B\right\} \cdot \mathbb{1}\left\{\pi_t = a\right\}\right] \\
&= \Delta^a(B) \cdot \mathbb{E}\left[T_n^a(\pi,B)\right].
\end{aligned}$$

■

Appendix G. Nested Partitions Lemma

In this section we prove Lemma 4. The proof utilises ideas from (Käenmäki et al. (2012)).

Lemma 4: Suppose that Assumption 1 holds. Given $q \in \mathbb{N}\backslash\{0\}$, $\delta \in (0, R_{\mathcal{X}}]$ and $r \in (0, 1/3)$ there exists a finite collection of subsets $\{Z_{l,i} : l \in [q], i \in [m_l]\}$ which satisfies:

1. For each $l \in [q]$, $\{Z_{l,i}\}_{i\in[m_l]}$ is a partition of $\mathcal{X}$.
2. Given $l_1, l_2 \in [q]$ with $l_1 \leq l_2$, $i_1 \in [m_{l_1}]$ and $i_2 \in [m_{l_2}]$, either $Z_{l_1,i_1} \cap Z_{i_2,l_2} = \emptyset$ or $Z_{l_2,i_2} \subseteq Z_{l_1,i_1}$.
3. For all $l \in [q]$, $i \in [m_l]$ we have $\operatorname{diam}(Z_{l,i}) \leq \delta \cdot r^l$ and $\mu\left(Z_{l,i}\right) \geq C_d \cdot \left((\delta/4) \cdot (1-3r) \cdot r^l\right)^d$.

Lemma 7: Suppose that $\epsilon > 0$, $U \subseteq V \subseteq \mathcal{X}$ and U is a maximal ϵ-separated subset of V. Suppose further that there exists a function $g : V \to U$ such that for all $v \in V$, $\rho(v, g(v)) = \min_{u\in U} \{\rho(u, v)\}$. Then for all $u \in U$, $\rho(u, g(u)) < \epsilon$ and if $v \in V\backslash \{g(u)\}$ then $\rho(u, v) \geq \epsilon/2$.

Proof Follows from the definition of a maximal ϵ-separated set. ■

Proof [Proof of Lemma 4] Let $\{x_{q,i}\}_{i\in[m_q]}$ be a finite maximal $(\delta/2) \cdot (1-r) \cdot r^q$ separated subset of $\mathcal{X}$. This is possible by Assumption 1. For $l \in [q-1]$ we let $\{x_{l,i}\}_{i\in[m_l]}$ be a maximal $(\delta/2) \cdot (1-r) \cdot r^l$ separated subset of $\{x_{l+1,i}\}_{i\in[m_{l+1}]}$, and define a function $g_l : \{x_{l+1,i}\}_{i\in[m_{l+1}]} \to \{x_{l,i}\}_{i\in[m_l]}$ by $g_l(x_{l+1,j}) = x_{l,i_j}$ where $i_j = \min\left\{\operatorname{argmin}_{i\in[m_l]} \left\{\rho(z, x_{l,i_j})\right\}\right\}$.

The collection of sets $\{Z_{l,i} : l \in [q], i \in [m_l]\}$ is defined as follows. First define a partition $\{Z_{q,i}\}_{i\in[m_q]}$ by

$$Z_{q,i} := B\left(x_{q,i}, (\delta/2) \cdot (1-r) \cdot r^q\right) \backslash \bigcup_{j<i} Z_{q,j}.$$

Then for $l = q-1, \cdots, 1$ we define partitions $\{Z_{l,i}\}_{i\in[m_l]}$ by

$$Z_{l,i} = \bigcup \left\{Z_{l+1,j} : g_l(x_{l+1,j}) = x_{l,i}\right\}.$$

Properties (1) and (2) in Proposition 4 are immediate.

We claim that for all $l \in [q]$ and $i \in [m_l]$, $Z_{l,i} \subseteq B\left(x_{l.i}, (\delta/2) \cdot r^l\right)$. For $l = q$ this follows from the construction of $\{Z_{q,i}\}_{i\in[m_q]}$. For $l \in [q-1]$, we assume that the claim holds for $l+1$. Given $z \in Z_{q,i}$ for some $i \in [m_q]$, by construction we must have $z \in Z_{l+1,j}$ for some $j \in [m_{l+1}]$ with $g(x_{l+1,j}) = x_{l,i}$. Hence, by assumption $\rho(z, x_{l+1,j}) < (\delta/2) \cdot r^{l+1}$. Also, by Lemma 7 we have $\rho(x_{l+1,j}, x_{l,i}) \leq (\delta/2) \cdot (1-r) \cdot r^l$. Hence, $\rho(z, x_{l,i}) < (\delta/2) \cdot r^l$, which proves the claim.

In addition we claim for all $l \in [q]$ and $i \in [m_l]$, $B\left(x_{l,i}, (\delta/4) \cdot (1-3r) \cdot r^l\right) \subseteq Z_{l,i}$. Indeed, for $l = q$, it follows from the fact that $\{x_{q,i}\}_{i\in[m_q]}$ is $(\delta/2) \cdot (1-r) \cdot r^q$ separated that

$$B\left(x_{q,i}, (\delta/4) \cdot (1-3r) \cdot r^q\right) \subseteq B\left(x_{q,i}, (\delta/4) \cdot r^q\right) \subseteq Z_{q,i}.$$

For $l \in [q-1]$, $i \in [m_l]$ we consider $z \notin Z_{l,i}$. Take $j \in [m_{l+1}]$ so that $z \in Z_{l+1,j}$. Given the construction of $Z_{l,i}$ we have $g(x_{l+1,j}) \neq x_{l,i}$. Hence, by Lemma 7 we have $\rho(x_{l+1,j}), x_{l,i}) \geq (\delta/4) \cdot (1-r) \cdot r^l$. Moreover, by the previous claim we have $\rho(z, x_{l+1,j}) < (\delta/2) \cdot r^{l+1}$. Hence, we have $\rho(z, x_{l,i}) \geq (\delta/4) \cdot (1-3r) \cdot r^l$. which completes the proof of the second claim.

Now take $l \in [q]$, $i \in [m_l]$. Since $Z_{l,i} \subseteq B\left(x_{l.i}, (\delta/2) \cdot r^l\right)$ we must have $\text{diam}(Z_{l,i}) \leq \delta \cdot r^l$. By Assumption 1, combined with the fact that $B\left(x_{l,i}, (\delta/4) \cdot (1-3r) \cdot r^l\right) \subseteq Z_{l,i}$ gives $\mu\left(Z_{l,i}\right) \geq C_d \cdot \left((\delta/4) \cdot (1-3r) \cdot r^l\right)^d$. Hence, property (3) in Proposition 4 also holds.

■

Appendix H. Dyadic sub-intervals

The Adaptively Binned Successive Elimination requires the following assumption.

Assumption 6 (Dyadic cubes assumption): There exists $C_d, d > 0$ such that for every $B \subset [0,1]^D$ of the form $B = 2^{-q} \cdot \prod_{i=1}^{D} [z_i, z_i + 1]$ with $z_1, \cdots, z_D, q \in \mathbb{N} \cup \{0\}$, we have either $\mu(B) \geq C_d \cdot \text{diam}(B)^d$ or $\mu(B) = 0$.

In the following example Assumption 1 holds yet Assumption 6 does not.

Example 1 *We define $\theta = \sum_{n=1}^{\infty} 2^{-n!}$, take $I_\theta = [0, \theta]$ and let μ_θ denote the normalised Lebesgue measure on I_θ. For any $x \in supp(\mu_\theta) = I_\theta$, and $r \in (0, \theta)$, $B(x,r) \cap I_\theta$ is an interval of diameter at least r, so we have $\mu_\theta(B(x,r)) \geq r/\theta$. Hence, Assumption 1 holds with $C_d = \theta^{-1}$ and $d = 1$. On the other hand, Assumption 6 does not hold. Indeed given $q \in \mathbb{N}$, we consider the dyadic interval $B_q := [\sum_{n=1}^{q} 2^{-n!}, \sum_{n=1}^{q} 2^{-n!} + 2^{q!}]$. Then $diam(B_q) = 2^{-q!}$. However, $B_q \cap I_\theta = [\sum_{n=1}^{q} 2^{-n!}, \sum_{n=1}^{\infty} 2^{-n!}]$, so*

$$\mu(B_q) = \theta^{-1} \cdot diam(B_q) = \theta^{-1} \cdot \sum_{n=q+1} 2^{-n!} \leq (2/\theta) \cdot 2^{-(q+1)!} = (2/\theta) \cdot diam(B_q)^{q+1}.$$

Consequently, given any $C_d, d > 0$ we can take $q \in \mathbb{N}$ sufficiently large that $q > d$ and $(2/\theta) \cdot 2^{-q!} < C_d$. It follows that whilst $\mu(B_q) \neq \emptyset$ we do have $\mu(B_q) \leq (2/\theta) \cdot 2^{-q!} \cdot diam(B_q)^{q+1} < C_d \cdot diam(B_q)^d$.

Appendix I. Manifolds and the minimax lower bound

In this section we shall recall some results regarding manifolds. This will serve two proposes. Firstly, we will make precise the sense in which Assumption 1 holds for all well-behaved measures μ supported on a d-dimensional submanifold of Euclidean space. Secondly, we will demonstrate that the regret bounds in Theorems 1 and 2 are minimax optimal up to logarithmic factors.

I.1. Manifolds, reach and regular measures

Suppose we have a C^∞-smooth sub-manifold of $\mathcal{M} \subset \mathbb{R}^D$ of dimension d (see Lee (2006)). We shall make use of the concept of reach τ introduced by Federer (1959) and investigated

by Niyogi et al. (2008). The reach τ of a manifold $\mathcal{M}$ is defined by

$$\tau := \sup\left\{r > 0 : \forall z \in \mathbb{R}^D \ \inf_{q\in\mathcal{M}}\{\|z-q\|_2\} < r \implies \exists!\ p \in \mathcal{M},\ \ \|z-p\|_2 = \inf_{q\in\mathcal{M}}\{\|z-q\|_2\}\right\}.$$

Note that Niyogi et al. (2008) refers to the condition number $1/\tau$, which is the reciprocal of the reach τ. We let $V_{\mathcal{M}}$ denote the Riemannian volume.

Definition 1 (Regular sets and measures): Suppose we have a measure υ on the metric space $(\mathcal{X}, \rho)$. A subset $A \subset \mathcal{X}$ is said to be a (c_0, r_0)-regular set with respect to the measure υ if for all $x \in A$ and all $r \in (0, r_0)$ we have $\upsilon(A \cap B_r(x)) \geq c_0 \cdot \upsilon(B_r(x))$, where $B_r(x)$ denotes the open metric ball of radius r, centred at x. A measure μ with support $\text{supp}(\mu) \subset \mathcal{X}$ is said to be $(c_0, r_0, \nu_{\min}, \nu_{\max})$-regular measure with respect to υ if $\text{supp}(\mu)$ is a (c_0, r_0)-regular set with respect to υ and μ is absolutely continuous with respect to υ with Radon-Nikodym derivative $\nu(x) = d\mu(x)/d\upsilon(x)$, such that for all $x \in \text{supp}(\mu)$ we have $\nu_{\min} \leq \nu(x) \leq \nu_{\max}$.

I.2. The dimension assumption on manifolds

In this section we justify Assumption 1 showing that it holds whenever the marginal μ is regular with respect to a d-dimensional manifold. The proof follows straightforwardly from Eftekhari and Wakin (2015).

Proposition 2: Let $\mathcal{M} \subseteq \mathbb{R}^D$ be a C^∞-smooth compact sub-manifold of dimension d and reach τ. Suppose that μ is a $(c_0, r_0, \nu_{\min}, \nu_{\max})$-regular measure with respect to $V_{\mathcal{M}}$. Then μ satisfies the dimension assumption (Assumption 1) with constants $R_{\mathcal{X}} = \min\{\tau/4, r_0\}$, d and $C_d = \nu_{\min} \cdot c_0 \cdot v_d \cdot 2^{-d}$, where v_d denotes the Lebesgue measure of the unit ball in $\mathbb{R}^d$.

Proof Take $x \in \text{supp}(\mu)$ & $r \in (0, R_{\mathcal{X}})$. By (Eftekhari and Wakin, 2015, Lemma 12) we have

$$V_{\mathcal{M}}(B_r(x)) \geq \left(1 - \frac{r^2}{4\tau^2}\right)^{\frac{d}{2}} \cdot v_d \cdot r^d \geq v_d \cdot 2^{-d} \cdot r^d. \tag{6}$$

Moreover, since μ is $(c_0, r_0, \nu_{\min}, \nu_{\max})$-regular we have

$$\begin{aligned}\mu(B_r(x)) &\geq \nu_{\min} \cdot V_{\mathcal{M}}(B_r(x) \cap \text{supp}(\mu)) \\ &\geq \nu_{\min} \cdot c_0 \cdot V_{\mathcal{M}}(B_r(x)).\end{aligned}$$

Combining with (6) proves the proposition. ■

I.3. A lower bound on regret for bandits on manifolds

The following result demonstrates that the regret bounds in Theorems 1 and 2 are minimax optimal up to logarithmic factors, for all sufficiently well-behaved manifolds. The theorem follows straightforwardly from the proof of (Reeve and Brown, 2017, Proposition A.1), which generalises (Audibert et al., 2007, Theorem 3.5) to embedded manifolds.

Theorem 7 *Let $\mathcal{M} \subset \mathbb{R}^D$ be a compact C^∞ sub-manifold of dimension d and reach τ, and take $A \geq 2$. There exists a universal positive constant $Z > 0$ and positive constants $C_0, R_0, V_-, V_+ > 0$ determined by γ, τ such that for all $c_0 \in (0, C_0)$, $r_0 \in (0, R_0)$, $\delta_\alpha \in (0, Z)$, $\alpha \in (0, d)$, $C_\alpha > 0$, $\lambda > 0$, there exists a constant $C > 0$, depending solely upon (d, τ), $(c_0, r_0, \nu_{\min}, \nu_{\max})$, $(\alpha, \delta_\alpha, C_\alpha)$ and λ such that the following holds: Given any policy π and $n \in \mathbb{N}$ there exists a distribution $\mathbb{P}$ on pairs (X, Y) with $X \in \mathbb{R}^D$ and $Y = (Y^a)_{a\in[A]} \in \{0,1\}^A$ such that the marginal over X, μ is $(c_0, r_0, \nu_{\min}, \nu_{\max})$-regular with respect to $V_{\mathcal{M}}$ (so $\mathbb{P}$ satisfies Assumption 1 by Lemma 2), the reward functions f^a are λ-Lipschitz (Assumption 2) and $\mathbb{P}$ satisfies the margin condition with constants $\delta_\alpha, C_\alpha, \alpha$ (Assumption 3) and*

$$\mathbb{E}\left[R_n(\pi)\right] \geq C \cdot n^{1-\frac{\alpha+1}{d+1}}.$$

Proof Let $\mathcal{P}$ denote the set of all distributions $\mathbb{P}$ on (X, Y) with $X \in \mathbb{R}^D$ and $Y = (Y^a)_{a\in[A]} \in \mathcal{Y} := \{0,1\}^A$ such that:

1. The marginal of $\mathbb{P}$ over X is $(c_0, r_0, \nu_{\min}, \nu_{\max})$-regular with respect to $V_{\mathcal{M}}$,
2. The reward functions $f^a : x \mapsto \mathbb{E}\left[Y^a | X = x\right]$ are λ-Lipschitz,
3. $\mathbb{P}$ satisfies the margin condition with constants $\delta_\alpha, C_\alpha, \alpha$.

Let $\mathcal{Y}_{\text{class}} := \left\{Y = (Y^a)_{a\in[A]} \in \mathcal{Y} : \sum_{a\in[A]} Y^a = 1\right\}$ and let $\mathcal{P}_{\text{class}} := \{\mathbb{P} \in \mathcal{P} : \mathbb{P}\left[Y \in \mathcal{Y}_{\text{class}}\right] = 1\}$. Let ϕ be a supervised classification procedure. We may view ϕ as a map from pairs $(\mathcal{F}_{n-1}, X_n)$, consisting of data set $\mathcal{F}_{n-1} = \{(X_t, Y_t)\}_{t\in[n-1]} \in (\mathcal{M} \times \mathcal{Y}_{\text{class}})^{n-1}$ and a covariate $X_n \in \mathcal{M}$, to an output $\phi\left(\mathcal{F}_{n-1}, X_n\right) \in [A]$. From the proof of (Reeve and Brown, 2017, Proposition A.1) we see that there exists a constant $C > 0$, depending solely upon (d, τ), $(c_0, r_0, \nu_{\min}, \nu_{\max})$, $(\alpha, \delta_\alpha, C_\alpha)$ and λ together with a finitely supported probability measure p on $\mathcal{P}_{\text{class}}$ such that for all classification procedures ϕ we have

$$\begin{aligned}
&\int_{\mathcal{P}_{\text{class}}} \left(\int_{(\mathcal{M}\times\mathcal{Y}_{\text{class}})^n} \left(Y_n^{\pi_*(X_n)} - Y_n^{\phi(\mathcal{F}_{n-1}, X_n)} \right) d\mathbb{P}^n\left(\mathcal{F}_n\right) \right) dp\left(\mathbb{P}\right) \\
&= \int_{\mathcal{P}_{\text{class}}} \left(\int_{(\mathcal{M}\times\mathcal{Y}_{\text{class}})^n} \mathbb{1}\left\{ Y_n^{\phi(\mathcal{F}_{n-1}, X_n)} \neq 1 \right\} - \mathbb{1}\left\{ Y_n^{\pi_*(X_n)} \neq 1 \right\} d\mathbb{P}^n\left(\mathcal{F}_n\right) \right) dp\left(\mathbb{P}\right) \\
&\geq C \cdot (n-1)^{-\frac{\alpha+1}{d+2}} \geq C \cdot n^{-\frac{1+\alpha}{2+d}}.
\end{aligned}$$

Now let π be any bandit policy. For each $t \in [n]$ we may convert π into a classification procedure $\phi_t : (\mathcal{F}_{n-1}, X_n) \mapsto \phi_t\left(\mathcal{F}_{n-1}, X_n\right) \in [A]$ by first applying π to $\{(X_s, Y_s)\}_{s\in[t-1]} \subset \mathcal{F}_{n-1}$ and letting $\phi_t\left(\mathcal{F}_{n-1}, X_n\right) = \pi_t\left(X_n\right)$. Hence, for each $t \in [n]$ we have,

$$\int_{\mathcal{P}_{\text{class}}} \left(\int_{(\mathcal{M}\times\mathcal{Y}_{\text{class}})^n} \left(Y_n^{\pi_*(X_n)} - Y_n^{\pi_t(X_n)} \right) d\mathbb{P}^n\left(\mathcal{F}_n\right) \right) dp\left(\mathbb{P}\right) \geq C \cdot n^{-\frac{1+\alpha}{2+d}}.$$

By symmetry, for each $t \in [n]$ we have,

$$\int_{\mathcal{P}_{\text{class}}} \left(\int_{(\mathcal{M}\times\mathcal{Y}_{\text{class}})^n} \left(Y_t^{\pi_*(X_t)} - Y_t^{\pi_t(X_t)} \right) d\mathbb{P}^n\left(\mathcal{F}_n\right) \right) dp\left(\mathbb{P}\right) \geq C \cdot n^{-\frac{1+\alpha}{2+d}}.$$

Hence, by Fubini's theorem

$$\begin{aligned}\int_{\mathcal{P}_{\text{class}}} \mathbb{E}\left[R_n(\pi)\right] dp\left(\mathbb{P}\right) &= \int_{\mathcal{P}_{\text{class}}} \left(\int_{(\mathcal{M}\times\mathcal{Y}_{\text{class}})^n} \sum_{t\in[n]} \left(Y_t^{\pi_*(X_t)} - Y_t^{\pi_t(X_t)} \right) d\mathbb{P}^n\left(\mathcal{F}_n\right) \right) dp\left(\mathbb{P}\right) \\ &= \sum_{t\in[n]} \int_{\mathcal{P}_{\text{class}}} \left(\int_{(\mathcal{M}\times\mathcal{Y}_{\text{class}})^n} \left(Y_t^{\pi_*(X_t)} - Y_t^{\pi_t(X_t)} \right) d\mathbb{P}^n\left(\mathcal{F}_n\right) \right) dp\left(\mathbb{P}\right) \\ &\geq C \cdot n^{1-\frac{\alpha+1}{d+2}}.\end{aligned}$$

In particular, there must exist some $\mathbb{P} \in \mathcal{P}_{\text{class}} \subset \mathcal{P}$ with $\mathbb{E}\left[R_n(\pi)\right] \geq C \cdot n^{1-\frac{\alpha+1}{d+2}}$. ■