Online Algorithm for Unsupervised Sensor Selection

Arun Verma
Dept. of IEOR
IIT Bombay, India
v.arun@iitb.ac.in

Manjesh K. Hanawal
Dept. of IEOR
IIT Bombay, India
mhanawal@iitb.ac.in

Csaba Szepesvári
DeepMind
London, UK
szepi@google.com

Venkatesh Saligrama
Dept. of ECE
Boston University, USA
srv@bu.edu

Abstract

In many security and healthcare systems, the detection and diagnosis systems use a sequence of sensors/tests. Each test outputs a prediction of the latent state and carries an inherent cost. However, the correctness of the predictions cannot be evaluated due to unavailability of the ground-truth annotations. Our objective is to learn strategies for selecting a test that gives the best trade-off between accuracy and costs in such unsupervised sensor selection (USS) problems. Clearly, learning is feasible only if ground truth can be inferred (explicitly or implicitly) from the problem structure. It is observed that this happens if the problem satisfies the 'Weak Dominance' (WD) property. We set up the USS problem as a stochastic partial monitoring problem and develop an algorithm with sub-linear regret under the WD property. We argue that our algorithm is optimal and evaluate its performance on problem instances generated from synthetic and real-world datasets.

1 Introduction

In many applications, one has to trade-off between accuracy and cost. For example, for detecting some event, it is not only the accuracy of a sensor that matters, but the associated sensing cost is important as well. Also, one may have to predict labels of instances for which ground-truth cannot be obtained. In such scenarios, feedback about the correctness of sensors' predictions remains unknown. Problems with this structure arise naturally in healthcare, security, and crowd-sourcing applications. In healthcare, the patients may not reveal the outcome of treatment due to privacy concerns; hence the effectiveness of the treatment is unknown. In crowd-sourcing systems, the expertise of self-listed-agents (workers) may not be known; therefore their quality cannot be identified. In a security application, specific threats may not have been seen before, and thus their in-situ ground-truth may not be available.

In this work, we focus on the study of sensor selection problems where we do not have the advantage of knowing the ground-truth and hence cannot measure the error rates of the sensors. Here sensors could correspond to medical tests (healthcare), detectors/scanners (security) or workers (crowd-sourcing). In these unsupervised sensor selection (USS) problems, the goal is to still find the 'best' sensor that gives the best trade-off between error and cost [1].

In USS setup, it is assumed that the sensors form a cascade, i.e., they are ordered by their prediction efficiency and costs– the average prediction error decreases hence, prediction efficiency increases with every stage of the cascade while the cost of acquiring it increases. Even though it is assumed that the sensor ordering is known and better sensors are associated with higher costs, the exact values of sensor errors are still unknown. The learner's goal is to find a sensor that has small value of total prediction cost for a given task, which includes both the cost of acquiring the sensor's outputs and the cost due to incorrect predictions.

Clearly, without the knowledge of the ground-truth, one cannot find the optimal sensor as the sensor accuracies cannot be computed. In the USS setup, the structure of the problem is exploited, and it is shown that under certain conditions, namely strong dominance (SD) and weak dominance (WD), learning is possible. The SD property requires the prediction accuracy of a sensor to stochastically dominate prediction accuracy of other sensors with lower costs in the cascade. Specifically, it

Proceedings of the 22[nd] International Conference on Artificial Intelligence and Statistics (AISTATS) 2019, Naha, Okinawa, Japan. PMLR: Volume 89.

assumes that if a sensor's prediction is correct, then all the sensors that follow this sensor in the cascade also have correct predictions.

Under the SD property, Hanawal et al. [1] established that USS problem is equivalent to a multi-armed bandit with side observations and exploit the equivalence to give an algorithm with sub-linear regret. SD property is quite strong and posits that disagreement probability of the predictions of two sensors is equal to the difference in error rates. This property implies that we can measure accuracy by measuring disagreement probabilities leading to a direct multi-armed bandit (MAB) reduction and analysis.

The WD property relaxes strict stochastic ordering on predictions and allows errors on some instances from better sensors. It is argued that the set of instances satisfying the WD property is maximally learnable, and any further relaxation of this property renders the problems unlearnable. The reduction techniques used under SD property does not apply/extend to WD property. For this case, a heuristic algorithm without any performance guarantee is given in [1]. Our work bridges this gap. Our contributions are summarized as follows:

- We develop an algorithm named USS-UCB that has sublinear regret under WD property. We characterize regret in terms of how 'well' the problem instances satisfy the WD property and then provide a bound that holds uniformly for all WD instances.
- We give problem independent bounds on the regret of USS-UCB. We show that it is of order $T^{2/3}$ under WD property and improves to $T^{1/2}$ under SD property. We establish that the bounds are optimal using results from partial monitoring in Section 3.
- Hanawal et al. assume that sensors are ordered, i.e., their accuracy improve with their index, and used this fact in their algorithms. We relax this assumption in Section 4 where the sensors can have an arbitrary order. For this setup, we show that the same WD property determines the learnability.
- We demonstrate performance of our algorithm on both synthetic and real datasets in Section 5. The experimental results show that regret of USS-UCB is always lower than the heuristic algorithm in [1] (See Fig. (3) in Section 5).

1.1 Related Work

Several works consider the problem of sensor selection in either batch, or online settings (e.g., Trapeznikov and Saligrama [2], Seldin et al. [3]). However, they all require that the label of each data point is available or the reward is obtained for each action. Zolghadr et al. [4] considers that the labels are available on payment. Greiner et al. [5], Póczos et al. [6] consider costs associated with tests. However, they assume that loss/reward associated with the players' action is revealed. In contrast, in our setting, the labels are not revealed at any point and are thus completely unsupervised, and the cost in our setup is related to sensing cost and not that of acquiring a label.

Platanios et al. [7, 8, 9] consider the problem of estimating accuracies of the multiple binary classifiers with unlabeled data. Most of these works make strong assumptions such as independence given the labels, knowledge of the true distribution of the labels. Platanios et al. [7] proposed logistic regression based methods using the classifiers' agreement rates over unlabeled data, [8] extend this work to use graphical models, and Platanios et al. [9] proposes method using probabilistic logic. Further, Platanios et al. [9] also uses weighted majority vote for label prediction. All this is in the batch setting and differs from our online setup.

In the crowd-sourcing problems, various methods have been proposed to estimate unknown skill-level of crowd-workers from the noisy labels they provide (Bonald and Combes [10], Kleindessner and Awasthi [11]). These methods assume that all workers are having the same cost and aggregate the predictions on a given dataset for estimating the accuracy of each worker. Unlike ours, these methods are not online.

Our work is closely related to the stochastic partial monitoring setting [12, 13, 14, 15], where the feedback from actions is indirectly tied to the rewards. In our setting, we exploit the problem structure to learn an optimal arm without explicitly knowing the loss associated with each action.

2 USS Problem

We cast the unsupervised, stochastic, cascaded sensor selection as an instance of stochastic partial monitoring problem (SPM). We use sensor and arm interchangeably in the following. Formally, a problem instance in our setting is specified by a pair $\theta = (P, c)$, where P is a distribution over the $K+1$ dimensional hypercube, and c is a K-dimensional, non-negative valued vector of costs. While c is known to the learner from the start, P is unknown. Henceforth, we identify problem instance by θ. The instance parameters specify the learner-environment interaction as follows: In each round $t = 1, 2, \ldots$, the environment generates a $K+1$-dimensional binary vector $(Y_t, Y_t^1, \ldots, Y_t^K) \in$

$\{0,1\}^{K+1}$ chosen at random from P. Here, Y_t^j is the output of sensor j, while Y_t is the (hidden) label to be guessed by the learner. Simultaneously, the learner chooses an index $I_t \in [K]$ where $[K] = \{1, 2, \ldots, K\}$, and observes the sensor outputs $Y_t^1, \ldots, Y_t^{I_t}$, i.e., the learner goes through the first I_t sensors and observes their outputs. Dropping the subindex t, write $S = (Y^1, \ldots, Y^K) \in \{0,1\}^K$. Then, P, the joint probability distribution of Y and S, can be expressed as $P = P_S \otimes P_{Y|S}$, where for any $s \in \{0,1\}^K$ and $y \in \{0,1\}$, $P_S(s) = \mathbb{P}\{S = s\}$ is (essentially) observable while $P_{Y|S}(y|s) = \mathbb{P}\{Y = y | S = s\}$ is not.

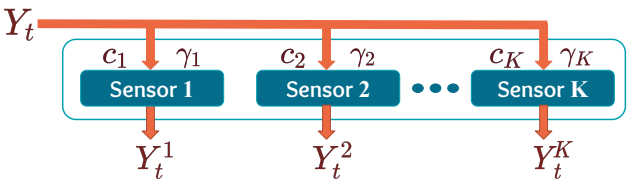


Figure 1: Cascaded Unsupervised Sensor Selection setup. Y_t is the hidden state of the instance and $Y_t^1, Y_t^2 \ldots Y_t^K$ are sensor outputs. c_j denotes the cost of using the sensor j and γ_j denotes error rate of the sensor j.

Hanawal et al. in addition assumes that the sensors are known to be ordered from least accurate to most accurate, i.e., $\gamma_j \doteq \gamma_j(\theta) \doteq \mathbb{P}\{Y \neq Y^j\}$ is decreasing in j. We relax this assumption later in the Section 4. The cost associated with sensor $j \in [K]$ is denoted by $c_j \geq 0$ and the cost of choosing action I_t is $C_{I_t} \doteq c_1 + \cdots + c_{I_t}$, as the selection has to be done sequentially. The *total cost* incurred by the learner in round t is thus $\lambda C_{I_t} + \mathbb{I}\{Y_t \neq Y_t^{I_t}\}$ where λ is a trade-off parameter between error rate and cost of using sensors[1]. Without loss of generality, we set $\lambda = 1$. The goal of the learner is to compete with the best choice knowing the θ. Let $c(j,\theta) \doteq \mathbb{E}\left[C_j + \mathbb{I}\{Y_t \neq Y_t^j\}\right]$ $(= C_j + \gamma_j)$ and $i^\star \doteq i^\star(\theta) \doteq \max\{l : l = \arg\min_{j \in [K]} c(j,\theta)\}$ be the optimal sensor. The cumulative (pseudo-)regret of the learner running an algorithm $\mathcal{A}$ up to the end of round T is

$$\mathcal{R}_T(\mathcal{A}, \theta) = \sum_{t=1}^{T} c(I_t, \theta) - T c(i^\star, \theta). \tag{1}$$

We say that the (expected) regret is sublinear if $\mathbb{E}[\mathcal{R}_T]/T \to 0$ as $T \to \infty$, where the expectation is over I_t, which is random as it depends on past random data. When the regret is sublinear, the learner collects almost as much reward in expectation in the long run as an oracle that knew the optimal action from the beginning. Let Θ_{SA} be the set of all stochastic, cascaded sensor selection problems. Thus, $\theta \in \Theta_{\text{SA}}$ such that $Y \sim \theta$ and $\gamma_j(\theta) := \mathbb{P}\{Y \neq Y^k\}$ is decreasing in k. Given a subset $\Theta \subset \Theta_{\text{SA}}$, we say that Θ is *learnable* if there exists a learning algorithm $\mathcal{A}$ such that for any $\theta \in \Theta$, the expected regret $\mathbb{E}[\mathcal{R}_T(\mathcal{A}, \theta)]$ of algorithm $\mathcal{A}$ on instance θ is sub-linear. A subset Θ is said to be a *maximal learnable problem class* if it is learnable and for any subset $\Theta' \subset \Theta_{\text{SA}}$ that contains Θ is not learnable.

[1] λ is a parameter that makes associated cost unit-less. For example, assume cost is in \$ and associated λ is p. If cost is increased by multiple of s ($s = 100$ for cost in cents) then the corresponding λ will be p/s and vice-versa.

2.1 Strong and Weak Dominance

The purpose of this section is to introduce the notions of strong and weak dominance from the work of Hanawal et al. [1]. While Hanawal et al. studied learning under strong dominance, here we will focus on weak dominance. We also modify the definition of weak dominance of Hanawal et al. to correct an oversight of them.

The strong dominance (SD) property is defined as follows:

Definition 1 (Strong Dominance (SD)). *An instance $\theta \in \Theta_{SA}$ is said to satisfy the strong dominance property if for $(Y, Y^1, \cdots, Y^K) \sim \theta$, it holds almost surely (a.s.) that*

$$Y^i = Y \text{ for some } i \in [K] \Rightarrow Y^j = Y \ \ \forall j > i. \tag{2}$$

The SD property implies that if a sensor predicts correctly then, a.s., all the sensors in the subsequent stages of the cascade also predict correctly. The set of all instances satisfying SD property, i.e., $\Theta_{\text{SD}} = \{\theta \in \Theta_{\text{SA}} : \theta \text{ satisfies SD condition}\}$ is learnable [1, Theorem 2]. The weaker version of the SD property is defined as follows:

Definition 2 (Weak Dominance (WD)). *An instance $\theta \in \Theta_{\text{SA}}$ is said to satisfy* weak dominance property *if*

$$\rho(\theta) := \min_{j > i^\star} \frac{C_j - C_{i^\star}}{\mathbb{P}\{Y^j \neq Y^{i^\star}\}} > 1. \tag{3}$$

Let $\Theta_{\text{WD}} = \{\theta \in \Theta_{\text{SA}} : \theta \text{ satisfies WD condition}\}$ denote the set of instances satisfying the WD property. The WD property holds for all problem instances where sensor K is an optimal sensor.

Hanawal et al. [1] claimed that Θ_{WD} is learnable. However, their definition allowed $\rho(\theta) \geq 1$. As it turns out, permitting $\rho(\theta) = 1$ can prevent Θ_{WD} from being learnable:

Proposition 1. *The set $\Theta'_{\text{WD}} = \{\theta \in \Theta_{\text{SA}} : \rho(\theta) \geq 1\}$ is not learnable.*

Proof. Let $C_2 - C_1 = 1/4$. Theorem 19 of Hanawal et al. [1] constructs instances $\theta, \theta' \in \Theta'_{\text{WD}}$ such that the optimal decision for θ is sensor 1, for θ' is sensor

2. The suboptimality gap on instance θ is $1/4$, while on instance θ' is ϵ, where $\epsilon \in [0,1]$ is a tunable parameter. At the same time $\mathbb{P}\left\{Y^1 \neq Y^2\right\} = 1/4$ in θ and $\mathbb{P}\left\{Y^1 \neq Y^2\right\} = 1/4 + \epsilon$ in θ'. Theorem 17 of Hanawal et al. [1] implies that a sound algorithm must check $1/4 = C_2 - C_1 \geq \mathbb{P}\left\{Y^1 \neq Y^2\right\}$. However, no finite amount of data is sufficient to decide this: In particular, one can show that if an algorithm on θ achieves sublinear regret, then it must suffer linear regret on θ' for $\epsilon > 0$ small enough. Hence, all algorithms will suffer linear regret on some instance in Θ'_{WD}. □

The following theorem is obtained directly from Theorem 14 and Theorem 19 in [1] after excluding the case $\rho = 1$ in their proofs.

Theorem 1. *The set Θ_{WD} is a maximal learnable set.*

In the following, we use an alternate characterization of the Θ_{WD} property given as

$$\xi := \min_{j > i^\star} \left\{ C_j - C_{i^\star} - \mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\} \right\} > 0 \qquad (4)$$

Notice that $\rho > 1$ if and only if $\xi > 0$. Larger the value of ξ 'stronger' is the Θ_{WD} property and easier it is to identify an optimal action. We later characterize the regret bounds in terms of ξ.

3 Algorithm Under WD Property

In the following, we let $i^\star$ denote the optimal arm with largest index, i.e., $i^\star = \max\{l : l = \arg\min_{j \in [K]} c(j,\theta)\}$. The optimal sensor $i^\star$ satisfies the following inequalities:

$$\forall j < i^\star \,:\, C_{i^\star} - C_j \leq \gamma_j - \gamma_{i^\star}\,, \qquad (5a)$$
$$\forall j > i^\star \,:\, C_j - C_{i^\star} > \gamma_{i^\star} - \gamma_j\,. \qquad (5b)$$

Note that the above decision criteria is risk-averse, i.e., if two sensors have the same optimal cost, the sensor with smaller error-rate will be chosen.

A natural candidate for a decision criteria is to replace error rates (γ_j) by their estimates and look for an index that satisfies (5a) and (5b). However, error rates (γ_j) cannot be estimated, implying that (5a) and (5b) can not lead to a sound algorithm. Recall the following result from [1]:

Proposition 2 ([1, Proposition 3])**.** *Let $\gamma_i = \gamma_i(\theta)$ for any θ, not necessarily in Θ_{SA}. Then, for any $i, j \in [K]$, $\gamma_i - \gamma_j = \mathbb{P}\left\{Y^i \neq Y^j\right\} - 2\mathbb{P}\left\{Y^i = Y, Y^j \neq Y\right\}$, and hence $\gamma_i - \gamma_j \leq \mathbb{P}\left\{Y^i \neq Y^j\right\}$.*

Using Proposition 2, criteria (5a) implies

$$\forall j < i^\star \,:\, C_{i^\star} - C_j \leq \mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\} \qquad (6)$$

where $\mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\}$ forms a proxy for $\gamma_j - \gamma_{i^\star}$. For the case $j > i^\star$, we can appeal to the WD property and can replace (5b) by

$$\forall j > i^\star \,:\, C_j - C_{i^\star} > \mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\} \qquad (7)$$

$\mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\}$ can be estimated as the distribution P_S is observable. Motivated by (6) and (7), we define the selection criteria based on the following sets:

$$\mathcal{B}^l = \left\{ i : \forall j < i, C_i - C_j \leq \mathbb{P}\left\{Y^i \neq Y^j\right\} \right\} \cup \{1\} \qquad (8)$$
$$\mathcal{B}^h = \left\{ i : \forall j > i, C_j - C_i > \mathbb{P}\left\{Y^i \neq Y^j\right\} \right\} \cup \{K\}. \qquad (9)$$

Lemma 1. *Let $\theta \in \Theta_{\text{WD}}$. Let $\mathcal{B} := \mathcal{B}(\theta) = \mathcal{B}^l \cap \mathcal{B}^h$. Then $\mathcal{B}$ contains the optimal sensor.*

The proof is in Appendix A.

3.1 USS-UCB

In bandit problems, the upper confidence bound (UCB) [16, 17] is highly effective for dealing with the trade-off between exploration and exploitation. Using UCB idea, we develop an algorithm, named USS-UCB, that utilizes the sets (8) and (9) and looks for an index that belongs to both. Since disagreement probabilities, $\left(p_{ij} \doteq \mathbb{P}\left\{Y^i \neq Y^j\right\}\right)$'s, are unknown (but fixed), they are replaced by their optimistic empirical estimates at round t, denoted by $\hat{p}_{ij}(t) + \Psi_{ij}(t)$ where $\hat{p}_{ij}(t)$ is empirical estimate of p_{ij} and $\Psi_{ij}(t)$ is the confidence term associated with $\hat{p}_{ij}(t)$ as in UCB algorithm. The new sets for selection criteria are defined as follows:

$$\hat{\mathcal{B}}^l_t = \{i : \forall j < i, C_i - C_j \leq \hat{p}_{ji}(t) + \Psi_{ji}(t)\} \cup \{1\}, \qquad (10a)$$
$$\hat{\mathcal{B}}^h_t = \{i : \forall j > i; C_j - C_i > \hat{p}_{ij}(t) + \Psi_{ij}(t)\} \cup \{K\}. \qquad (10b)$$

From the definition, it is easy to verify that $\hat{p}_{ij}(t) = \hat{p}_{ji}(t)$ and $\Psi_{ij}(t) = \Psi_{ji}(t)$ for any (i,j) pair. Therefore, it is enough for algorithm to only keep track of $\hat{p}_{ij}(t)$ and $\Psi_{ji}(t)$ for $i < j$.

Remark 1. *It might be tempting to use lower confidence, i.e., $\hat{p}_{ij}(t) - \Psi_{ij}(t)$ term instead of the upper confidence term in (10b). However, such a change can make the algorithm converge to a sub-optimal sensor. A detailed discussion is given in the supplementary material.*

The pseudo code of USS-UCB is given in Algorithm USS-UCB and it works as follows. It takes α as an input that trades-off between exploration and exploitation. In the first round, it selects sensor K and initializes the value of number of comparisons and counter of disagreements for each pair $(i,j), i < j$, denoted $\mathcal{N}_{ij}(1)$ and $\mathcal{D}_{ij}(1)$ (Line 2), respectively. In each subsequent round, the algorithm computes estimate for the disagreement probability $\hat{p}_{ij}(t)$ (Line 4)and the

USS-UCB Algorithm for **USS** under **WD** property

Input: $\alpha > 0.5$

1: Select sensor $I_1 = K$ and observe $Y_1^1, \ldots, Y_1^{I_1}$
2: Set $\mathcal{D}_{ij}(1) \leftarrow \mathbb{1}_{\{Y_1^i \neq Y_1^j\}}$, $\mathcal{N}_{ij}(1) \leftarrow 1 \ \ \forall i < j \leq I_1$
3: **for** $t = 2, 3, \ldots$ **do**
4: $\quad \hat{p}_{ij}(t) \leftarrow \frac{\mathcal{D}_{ij}(t-1)}{\mathcal{N}_{ij}(t-1)} \ \ \forall i < j \leq K$
5: $\quad \Psi_{ij}(t) \leftarrow \sqrt{\frac{\alpha \log f(t)}{\mathcal{N}_{ij}(t-1)}} \ \ \forall i < j \leq K$
6: $\quad$ Compute $\hat{\mathcal{B}}_t^l$ and $\hat{\mathcal{B}}_t^h$ as given in (10a) and (10b)
7: $\quad \hat{\mathcal{B}}_t := \hat{\mathcal{B}}_t^l \cap \hat{\mathcal{B}}_t^h$
8: $\quad I_t \leftarrow \min\left\{\hat{\mathcal{B}}_t \cup \{K\}\right\}$
9: $\quad$ Select sensor I_t and observe $Y_t^1, \ldots, Y_t^{I_t}$
10: $\quad \mathcal{D}_{ij}(t) \leftarrow \mathcal{D}_{ij}(t-1) + \mathbb{1}_{\{Y_t^i \neq Y_t^j\}} \ \ \forall i < j \leq I_t$
11: $\quad \mathcal{N}_{ij}(t) \leftarrow \mathcal{N}_{ij}(t-1) + 1 \ \ \forall i < j \leq I_t$
12: **end for**

associated confidence $\Psi_{ij}(t)$ (Line 5)). Then $\hat{p}_{ij}(t)$ and $\Psi_{ij}(t)$ are used for computing sets $\mathcal{B}_t^l$ and $\mathcal{B}_t^h$ (Line 6) which are then used to select the sensor. Specifically, the algorithm selects a sensor I_t that satisfies (10a) and (10b) (Line 8).

Since initial estimates for p_{ij} are not good enough, $\hat{\mathcal{B}}_t$ can be empty. In such a case, the algorithm selects the sensor K. After selection of sensor I_t, $Y_t^j, j \in [I_t]$ (Line 9) are observed which are then used to update the $\mathcal{D}_{ij}(t)$ (Line 10) and $\mathcal{N}_{ij}(t)$ (Line 11) in the algorithm.

3.2 Regret Analysis

Following notations and definition are useful in subsequent proofs. For the optimal sensor $i^\star$ and each $j \in [K]$, let

$$\Delta_j := C_j + \gamma_j - (C_{i^\star} + \gamma_{i^\star}), \tag{11}$$

$$\kappa_j := \begin{cases} p_{i^\star j} - (\gamma_j - \gamma_{i^\star}), & \text{if } j < i^\star \quad (12a) \\ p_{i^\star j} - (\gamma_{i^\star} - \gamma_j), & \text{if } j > i^\star \quad (12b) \end{cases}$$

$$\xi_j := \begin{cases} \Delta_j + \kappa_j, & \text{if } j < i^\star \quad (13a) \\ \Delta_j - \kappa_j, & \text{if } j > i^\star \quad (13b) \end{cases}$$

Notice that the values of κ_j and ξ_j for all $j \in [K]$ are positive under the WD property. Let $N_j(T)$ denote the number of times sensor j is selected until round T. The following proposition gives the mean number of times a sub-optimal sensor is selected.

Proposition 3. *Let $f(t)$ be a positive valued increasing function such that $C = \lim_{T\to\infty} \sum_{t=1}^{T} \frac{1}{f(t)^{2\alpha}} < \infty$ in USS-UCB. For any $\theta \in \Theta_{\text{WD}}$, the mean number of times a sensor $j \neq i^\star$ is selected, is bounded as follows:*

- *for any $j < i^\star$*

$$\mathbb{E}\left[N_j(T)\right] \leq \frac{C}{2\xi_j^2},$$

- *and for any $j > i^\star$*

$$\mathbb{E}\left[N_j(T)\right] \leq 1 + \frac{1}{\xi_j^2}\left(\alpha \log f(T) + \sqrt{\frac{\pi\alpha \log f(T)}{2}} + \frac{1}{2}\right).$$

Notice that the mean number of times a sensor $j < i^\star$ is selected, is finite. The regret bounds follows by noting that $\mathbb{E}\left[\mathcal{R}_T\right] = \sum_{j<i^\star} \mathbb{E}\left[N_j(T)\right] \Delta_j + \sum_{j>i^\star} \mathbb{E}\left[N_j(T)\right] \Delta_j$. Formally, we have the following regret bound.

Theorem 2. *Let $f(t)$ be set as in Proposition 3. Then, for any $\theta \in \Theta_{\text{WD}}$, the expected regret of USS-UCB in T rounds is bounded as below:*

$$\mathbb{E}\left[\mathcal{R}_T\right] \leq \sum_{j<i^\star} \frac{\Delta_j C}{2\xi_j^2} + \sum_{j>i^\star} \Delta_j \left[1 + \frac{1}{\xi_j^2}\left(\alpha \log f(T) + \sqrt{\frac{\pi\alpha \log f(T)}{2}} + \frac{1}{2}\right)\right].$$

Corollary 1. *Let $\alpha = 1$ and $f(t) = t$ in Theorem 2. Then, expected regret of USS-UCB for any $\theta \in \Theta_{\text{WD}}$ in T rounds is of $O\left(\sum_{j>i^\star} \frac{\Delta_j \log T}{\xi^2}\right)$.*

Corollary 2. *Let technical conditions stated in Corollary 1 hold. Then expected regret of USS-UCB for any $\theta \in \Theta_{\text{SD}}$ in T rounds is of $O\left(\sum_{j>i^\star} \frac{\log T}{\xi}\right)$.*

Proof. Since $|\gamma_j - \gamma_{i^\star}| = p_{i^\star j}$ for $\theta \in \Theta_{\text{SD}}$, $\kappa_j = 0, \forall j \in [K] \Rightarrow \xi_j = \Delta_j$. Rest follows from Corollary 1. □

We next present problem independent bounds on the expected regret of USS-UCB.

Theorem 3. *Let $f(t)$ be set as in Proposition 3. The expected regret of USS-UCB in T rounds*

- *for any instance in Θ_{WD} is bounded as*
$$\mathbb{E}\left[\mathcal{R}_T\right] \leq 3\left(3\alpha K \log f(T)\right)^{1/3} T^{2/3}.$$
- *for any instance in Θ_{SD} is bounded as*
$$\mathbb{E}\left[\mathcal{R}_T\right] \leq 4\left(\alpha K T \log f(T)\right)^{1/2}.$$

Corollary 3. *The expected regret of USS-UCB on Θ_{SD} is $\tilde{O}(T^{1/2})$ and on Θ_{WD} it is $\tilde{O}(T^{2/3})$, where $\tilde{O}$ hides logarithmic terms.*

The proof of Theorem 3 can be found in the supplementary material. We note that the above uniform bounds do not contradict Theorem 19 in [1] which claimed non-existence of uniform bounds. The Θ_{WD} condition considered in [1] incorrectly includes the class of instances satisfying $\rho = 1$ which renders Θ_{WD} not learnable, whereas in our definition of Θ_{WD} these instances are excluded and Θ_{WD} is learnable.

Discussion on optimality of USS-UCB: Any partial monitoring problem can be classified as an

'easy', 'hard' or 'hopeless' problem if it has expected regret bounds of the order $\Theta(T^{1/2}), \Theta(T^{2/3})$ or $\Theta(T)$, respectively, and there exists no other class in between [14]. The class Θ_{SD} is regret equivalent to a stochastic multi-armed bandit with side observations [1], for which regret scales as $\Theta(T^{1/2})$, hence Θ_{SD} resides in the easy class and our bound on it is optimal. Since $\Theta_{\text{WD}} \supsetneq \Theta_{\text{SD}}$, Θ_{WD} is not easy, and also Θ_{WD} is learnable, it cannot be hopeless. Therefore, the class Θ_{WD} is hard. We thus conclude that the regret bound of USS-UCB is optimal in T. However, optimality concerning other leading constants (in terms of K) is to be explored further.

4 Unknown Ordering of Sensors

The sensor error rates are unknown in our setup and cannot be estimated due to unavailability of ground-truth. Thus, it may happen that we do not know whether error rate of the sensors in the cascade is decreasing or not. In this section, we remove the requirement that sensors are arranged in the decreasing order of their error rates and allow them to be arranged in an arbitrary order that is unknown. We denote the set of USS instances with unknown ordering of sensors by their error-rates as $\Theta'_{\text{SA}} \supset \Theta_{SA}$. The rest of the setup is same as in Section 2. We show that even with this relaxation, WD property defined earlier continues to characterize the learnability of the problem.

We begin with the following observation.

Lemma 2. *Let $i^\star$ be an optimal sensor. Then, error rate of any sensor $j < i^\star$ is higher than that of $i^\star$.*

Proof. We have $\gamma_j - \gamma_{i^\star} \geq C_{i^\star} - C_i$ for all $j \in [K]$. For $j < i^\star$, $C_{i^\star} - C_j \geq 0$ as costs are increasing with sensors. Hence $\gamma_j \geq \gamma_{i^\star}$. □

The following corollary directly follows from Prop. 2.

Corollary 4. *For any $i, j \in [K]$, $\max\{0, \gamma_j - \gamma_i\} \leq \mathbb{P}\left\{Y^i \neq Y^j\right\}$.*

The following two propositions provide the conditions on sensor costs that allows comparison of their total costs based on disagreement probabilities.

Proposition 4. *Let $i < j$. Assume*

$$C_j - C_i \notin \left(\max\{0, \gamma_i - \gamma_j\}, \mathbb{P}\left\{Y^i \neq Y^j\right\}\right]. \quad (14)$$

Then, $C_j - C_i > \max\{0, \gamma_i - \gamma_j\}$ iff $C_j - C_i > \mathbb{P}\left\{Y^j \neq Y^i\right\}$.

Proposition 5. *Let $i > j$. Assume*

$$C_i - C_j \notin \left(\max\{0, \gamma_j - \gamma_i\}, \mathbb{P}\left\{Y^i \neq Y^j\right\}\right]. \quad (15)$$

Then, $C_i - C_j \leq \max\{0, \gamma_j - \gamma_i\}$ iff $C_j - C_i \leq \mathbb{P}\left\{Y^i \neq Y^j\right\}$.

From Lemma (2), for any $j < i^*$ we have $\max\{0, \gamma_j - \gamma_{i^\star}\} = \gamma_j - \gamma_{i^\star}$. Propositions (4) and (5) then suggests that the value of $\mathbb{P}\left\{Y^i \neq Y^j\right\}$ are sufficient to select the optimal sensor if the sensors costs satisfy (Eq. (14)) for all $j > i^\star$ and Eqn. (15) for all $j < i^\star$. Since the values of $\mathbb{P}\left\{Y^i \neq Y^j\right\}$ can be estimated for all $i, j \in [K]$, we can establish the following result.

Proposition 6. *Let $i^\star$ be an optimal sensor. Any problem instance $\theta \in \Theta'_{\text{SA}}$ is learnable if*

$$\forall\, j > i^\star \;\; C_j - C_{i^\star} \notin \left(\max\{0, \gamma_{i^\star} - \gamma_j\}, \mathbb{P}\left\{Y^i \neq Y^j\right\}\right].$$

Notice that for $j > i^\star$, $C_j - C^\star_i \geq 0$ and $C_j - C_{i^\star} \geq \gamma_{i^\star} - \gamma_j$. Hence, the learnability condition reduces to $\forall\, j > i^\star, C_j - C_{i^\star} > \Pr\{Y^{i^\star} \neq Y^j\}$, i.e., same as the WD condition. Hence, we have the following result.

Theorem 4. *The set $\{\theta \in \Theta'_{\text{SA}} : \rho(\theta) > 1\}$ is learnable.*

5 Experiments

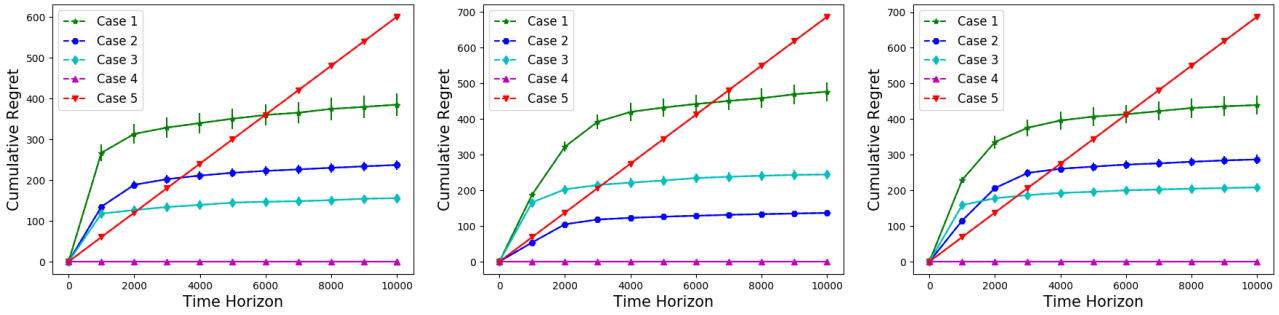
In this section, we evaluate the performance of USS-UCB on different problem instances derived from synthetic and two 'real' datasets: PIMA Indians Diabetes [18] and Heart Disease (Cleveland) [19, 20]. In our experiments, each sensor is represented by a classifier that is arranged in order of their decreasing misclassification error, i.e., error-rate for each dataset. The cost of using a classifier is assigned based on its error-rate – smaller the error-rate higher the cost. The case where sensors' error-rate need not to decrease in the cascade is also considered.

Synthetic Dataset: We generate synthetic Bernoulli Symmetric Channel (BSC) dataset [1] as follows: The input, Y_t, is generated from i.i.d. Bernoulli(0.7) random variable. The problem instance used in experiment has three sensors with error rates $\gamma_1 = 0.4, \gamma_2 = 0.1, \gamma_3 = 0.05$. To ensure strong dominance, we impose the condition given in Eq. (2) during data generation. When sensor 1 predicts correctly, we introduce error up to 10% to the outputs of sensor 2 and 3. We use five problem instances by varying the associated cost of each sensor as given in Table 1.

Table 1: BSC Dataset. WD doesn't hold for Case 5. Optimal classifier's cost is in red bold font.

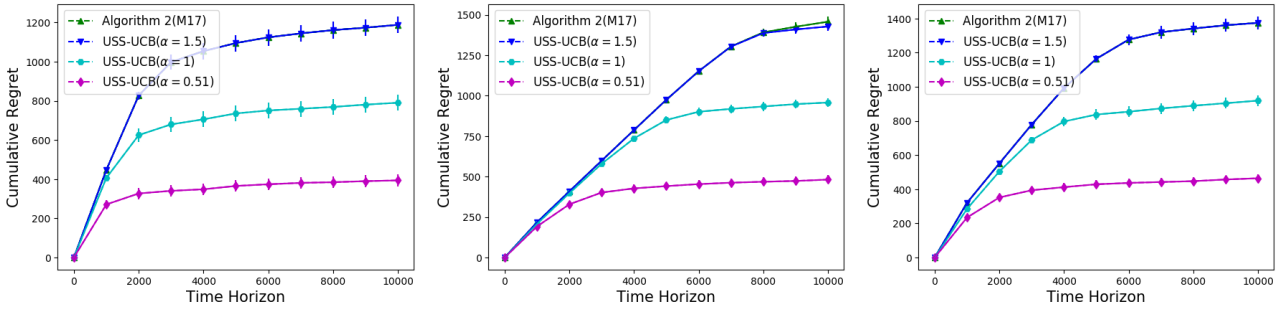
Values/Classifiers	Clf. 1	Clf. 2	Clf. 3	WD Prop.
Case 1 Costs	**0**	0.6	0.8	✓
Case 2 Costs	0	**0.15**	0.35	✓
Case 3 Costs	**0**	0.65	0.9	✓
Case 4 Costs	0.2	0.36	**0.4**	✓
Case 5 Costs	0	**0.11**	0.22	×

Real Datasets: Both real datasets specify the costs of acquiring individual features. We split these



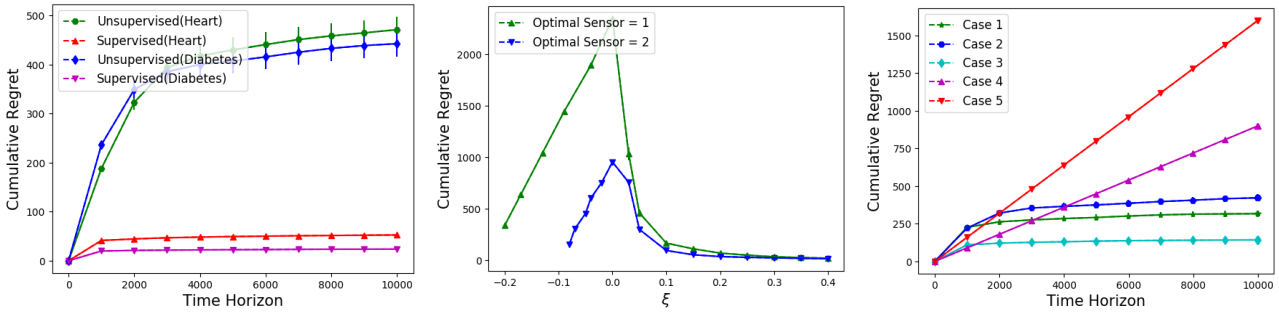
(a) Synthetic BSC Dataset (b) Pima Indian Diabetes (c) Heart Disease

Figure 2: Cumulative regret for different problem instances of USS-UCB with parameter $\alpha = 0.51$.



(a) Synthetic BSC Dataset (b) Pima Indian Diabetes (c) Heart Disease

Figure 3: Comparison between Heuristic Algorithm 2 proposed in [1] and USS-UCB with parameter $\alpha = \{1.5, 1, 0.51\}$ for Case 1 of the synthetic BSC dataset and real datasets.



(a) Real Datasets (b) Synthetic BSC Dataset (c) Synthetic BSC Dataset

Figure 4: Comparison between unsupervised and supervised setting is shown for Case 1 of real datasets (4a). Cumulative regret v/s WD property for BSC Dataset using different costs. Right figure: Cumulative regret v/s Time Horizon for synthetic BSC Dataset when sensor are not ordered by their error rates (4b). Sensor 2 and 3 are interchanged in the sequence while keeping the cost same as given in the Table 1 for synthetic BSC dataset (4c). Note that, $i^\star = K$ for Case 4 and WD automatically holds but after interchanging last two classifiers, WD does not hold for Case 4.

features into three subsets based on their costs and train three linear classifiers on these subsets using logistic regression. For PIMA-Diabetes dataset (# of samples=768) the first classifier is associated with patient history/profile at the cost of \$6, the 2nd classifier, in addition, utilizes glucose tolerance test (cost \$ 29) and the 3rd classifier uses all attributes including insulin test (cost \$46). For the Heart dataset (# of samples=297) we associate 1st classifier with the first 7 attributes that include cholesterol readings, blood-sugar, and rest-ECG (cost \$32), the 2nd classifier utilizes, in addition, the thalach, exang and oldpeak attributes that cost \$397 and the 3rd classifier utilizes more extensive tests at a total cost of \$601. We scale costs using a tuning parameter λ (since the costs of features are all greater than one) and consider minimizing a combined objective ($\lambda Cost + Error$) as stated in Section 2. In our setup, high (low)-values

Table 2: Real Datasets. WD doesn't hold for Case 5. Optimal classifier's cost is in red bold font.

Values/ Classifiers	PIMA-Diabetes			Heart Disease			WD Pro.
	Clf. 1	Clf. 2	Clf. 3	Clf. 1	Clf. 2	Clf. 3	
Error-rate	0.3125	0.2331	0.2279	0.29292	0.20202	0.14815	
Cost (in $)	4	29	46	32	397	601	
λ in Case 1	**0.01**	0.0106	0.015	**0.0001**	0.0008	0.001	✓
λ in Case 2	0.01	**0.004**	0.0038	0.0001	**0.0001**	0.00035	✓
λ in Case 3	**0.01**	0.0113	0.015	**0.0001**	0.0009	0.001	✓
λ in Case 4	0.0001	0.0001	**0.0001**	0.00001	0.00004	**0.0001**	✓
λ in Case 5	0.01	**0.002**	0.0055	0.0042	**0.0001**	0.00027	×

for λ correspond to low (high)-budget constraint. For example, if we set a fixed budget of \$50, this corresponds to high-budget (small λ) and low budget (large λ) for PIMA Diabetes (3rd classifier optimal) and Heart Disease (1st classifier optimal) respectively. For performance evaluation, different values of λ are used in five problem instances for both real datasets as given in Table 2.

Verifying WD property: As we know the error-rate associated with each sensor, we can find an optimal sensor for a given problem instance. Once the optimal sensor is known, WD property is verified by using estimates of disagreement probability after T rounds.

Expected Cumulative Regret v/s Time Horizon: The *Expected Cumulative Regret* of USS-UCB with $\alpha = 0.51$ versus *Time Horizon* plots for the Synthetic BSC Dataset and two real datasets are shown in Figure 2. These plots verify that any instance that satisfies WD property has sub-linear regret. The online USS-UCB selects an instance randomly from the dataset (with replacement) in each round for fixed time horizon. Further, we make a comparison of Algorithm 2 of [1] and USS-UCB for different values of α. With same value of $\alpha = 1.5$, Algorithm 2 of [1] and USS-UCB gives same regret whereas USS-UCB with $\alpha = 0.51$ gives best result. as shown in the Figure 3. We verify that if WD holds in any problem instance with the arbitrary ordering of sensors by error rates, then the problem is learnable as shown in Figure 4c. We fix the time horizon to 10000 for our experiments. We repeat each experiment 100 times, and average regret with 95% confidence bound is presented.

Supervised v/s Unsupervised Learning: We compare USS-UCB against an algorithm where the learner receives feedback. In particular, for each action in each round, in the bandit setting, the learner knows whether or not the corresponding sensor output is correct. We implement the "supervised bandit" setting by replacing Step 4 in USS-UCB with estimated marginal error rates. We notice that for both high as well as low-cost scenarios, while supervised algorithm does have lower regret, the USS-UCB cumulative regret is also sublinear as shown in Figure 4a. It is qualitatively interesting because these plots demonstrate that, in typical cases, our unsupervised algorithm learn as good as the supervised setting.

Learnability v/s WD Property: To verify the relationship between learnability and WD property, we experiment with different problem instances of synthetic BSC dataset that are parameterized by varying costs. We test the hypothesis that set of problem instances satisfying the WD property is a maximal learnable set. We fixed an optimal sensor and vary the costs in such a way that we continuously pass from the situation where WD holds ($\xi := \min_{j>i^\star} \xi_j$ and $\xi > 0$) to the case where WD does not hold ($\xi \leq 0$ or $C_j - C_{i^\star} \in (\gamma_{i^\star} - \gamma_j, p_{i^\star j}]$ for any $j > i^\star$). If WD does not hold for any problem instance then USS-UCB converges to sub-optimal sensor j instead of optimal sensor $i^\star$. In such problem instances, as $C_j - C_{i^\star}$ increase, the cumulative regret (1) will also increase due to selection of sub-optimal sensor j by USS-UCB until WD does not hold for that problem instance i.e., $\xi > 0$. The difference $C_j - C_{i^\star}$ is lower bounded by $\gamma_{i^\star} - \gamma_j$ in such cases, therefore, ξ cannot be less than $-\Delta_j$. We start experiments with the minimum possible value of ξ for which problem instance does satisfy WD property and then increase the value of ξ. Figure 4b depicts cumulative regret USS-UCB v/s ξ plots for Synthetic BSC Dataset. It can be seen clearly that there is indeed a transition at $\xi = 0$.

6 Conclusion

We studied the problem of selecting the best sensor in a cascade of sensors where they are ordered according to their prediction accuracies. The best sensor optimally trades-off between sensor costs and their prediction accuracy. The challenge in this setup is that the ground truth is not revealed at any time and hence setup is completely unsupervised. We modeled it as stochastic partial monitoring problem and proposed an algorithm that gives sub-linear regret under the Weak Dominance (WD) property. We showed that our algorithm enjoys regret of order $\tilde{O}(T^{2/3})$ (hiding logarithmic terms) and when the problem instance satisfies the more stringent Strong Dominance property, the regret bound improves to $\tilde{O}(T^{1/2})$. We showed that our algorithm enjoys the same performance under WD property even if the sensor ordering is not necessarily according to the decreasing value of their prediction accuracies.

In the current work, we did not exploit any side information (contexts) available with the tasks. It would be interesting to study the contextual version of this problem where the optimal sensor could be job dependent.

Acknowledgment

Arun Verma is partially supported by MHRD Fellowship, Govt. of India. M.K. Hanawal is supported by IIT Bombay IRCC SEED grant (16IRCCSG010) and INSPIRE faculty fellowship (IFA-14/ENG-73) from DST, Govt. of India. V. Saligrama acknowledges the support of the NSF through grant 1527618. AV and MKH would like to thank Prof. N. Hemachandra, IEOR, IIT Bombay for many useful discussions. This work was done when Csaba Szepesvári was at leave from the University of Alberta.

References

[1] Manjesh Hanawal, Csaba Szepesvari, and Venkatesh Saligrama. Unsupervised sequential sensor acquisition. In *Artificial Intelligence and Statistics*, pages 803–811, 2017.

[2] Kirill Trapeznikov and Venkatesh Saligrama. Supervised sequential classification under budget constraints. In *Artificial Intelligence and Statistics*, pages 581–589, 2013.

[3] Yevgeny Seldin, Peter L Bartlett, Koby Crammer, and Yasin Abbasi-Yadkori. Prediction with limited advice and multiarmed bandits with paid observations. In *ICML*, pages 280–287, 2014.

[4] Navid Zolghadr, Gábor Bartók, Russell Greiner, András György, and Csaba Szepesvári. Online learning with costly features and labels. In *Advances in Neural Information Processing Systems*, pages 1241–1249, 2013.

[5] Russell Greiner, Adam J Grove, and Dan Roth. Learning cost-sensitive active classifiers. *Artificial Intelligence*, 139(2):137–174, 2002.

[6] Barnabás Póczos, Yasin Abbasi-Yadkori, Csaba Szepesvári, Russell Greiner, and Nathan Sturtevant. Learning when to stop thinking and do something! In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 825–832. ACM, 2009.

[7] Emmanouil Antonios Platanios, Avrim Blum, and Tom M Mitchell. Estimating accuracy from unlabeled data. In *UAI*, pages 682–691, 2014.

[8] Emmanouil Antonios Platanios, Avinava Dubey, and Tom Mitchell. Estimating accuracy from unlabeled data: A bayesian approach. In *International Conference on Machine Learning*, pages 1416–1425, 2016.

[9] Emmanouil Platanios, Hoifung Poon, Tom M Mitchell, and Eric J Horvitz. Estimating accuracy from unlabeled data: A probabilistic logic approach. In *Advances in Neural Information Processing Systems*, pages 4361–4370, 2017.

[10] Thomas Bonald and Richard Combes. A minimax optimal algorithm for crowdsourcing. In *Advances in Neural Information Processing Systems*, pages 4352–4360, 2017.

[11] Matthäus Kleindessner and Pranjal Awasthi. Crowdsourcing with arbitrary adversaries. In *International Conference on Machine Learning*, pages 2713–2722, 2018.

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

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

[14] Gábor Bartók, Dean P Foster, Dávid Pál, Alexander Rakhlin, and Csaba Szepesvári. Partial monitoring—classification, regret bounds, and algorithms. *Mathematics of Operations Research*, 39(4):967–997, 2014.

[15] Yifan Wu, András György, and Csaba Szepesvári. Online learning with gaussian payoffs and side observations. In *Advances in Neural Information Processing Systems*, pages 1360–1368, 2015.

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

[17] 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.

[18] UCI Machine Learning, Kaggle. Pima Indians Diabetes Database. 2016. URL https://www.kaggle.com/uciml/pima-indians-diabetes-database.

[19] Robert Detrano. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, MD, Ph.D., Donor: David W. Aha. 1998. URL https://archive.ics.uci.edu/ml/datasets/Heart+Disease.

[20] Dua Dheeru and Efi Karra Taniskidou. UCI machine learning repository, 2017. URL http://archive.ics.uci.edu/ml.

[21] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

Supplementary Material: Online Algorithm for Unsupervised Sensor Selection

A Proof of Lemma 1

Lemma 1. *Let $\theta \in \Theta_{\text{WD}}$. Let $\mathcal{B} := \mathcal{B}(\theta) = \mathcal{B}^l \cap \mathcal{B}^h$. Then $\mathcal{B}$ contains the optimal sensor.*

Proof. Let $i^\star$ be an optimal sensor. Define

$$\mathcal{B}^l := \left\{i : i \in [2, K]; \forall j < i \ \ni \ C_i - C_j \leq \mathbb{P}\{Y^i \neq Y^j\}\right\} \cup \{1\} \tag{16}$$

$$\mathcal{B}^l_{-1} := \mathcal{B}^l \setminus \{1\} \tag{17}$$

$$\mathcal{B}^h := \left\{i : i \in [1, K-1]; \forall j > i \ \ni \ C_j - C_i > \mathbb{P}\{Y^i \neq Y^j\}\right\} \cup \{K\} \tag{18}$$

$$\mathcal{B}^h_{-K} := \mathcal{B}^h \setminus \{K\} \tag{19}$$

$$\mathcal{B} := \mathcal{B}^l \cap \mathcal{B}^h \tag{20}$$

Consider following three cases:

I. $1 < i^\star < K$

II. $i^\star = 1$

III. $i^\star = K$

Case I: $1 < i^\star < K$
As $i^\star$ is an optimal sensor therefore $\forall j > i^\star : C_j - C_{i^\star} > \mathbb{P}\{Y^{i^\star} \neq Y^j\} \Rightarrow C_j - C_{i^\star} \nleq \mathbb{P}\{Y^{i^\star} \neq Y^j\} \Rightarrow \forall j > i^\star \notin \mathcal{B}^l_{-1}$. If any sensor $l \in \mathcal{B}^l_{-1}$ then $l \leq i^\star$ i.e.,

$$\mathcal{B}^l_{-1} = \{l_1, l_2, \ldots, l_m, i^\star\} \qquad \text{where } 1 < l_1 < \cdots < l_m < i^\star \tag{21}$$

$$\mathcal{B}^l = \mathcal{B}^l_{-1} \cup \{1\} = \{1, l_1, l_2, \ldots, l_m, i^\star\} \tag{22}$$

Similarly, $\forall j < i^\star : C_{i^\star} - C_j \leq \mathbb{P}\{Y^{i^\star} \neq Y^j\} \Rightarrow C_{i^\star} - C_j \ngtr \mathbb{P}\{Y^{i^\star} \neq Y^j\} \Rightarrow \forall j < i^\star \notin \mathcal{B}^h_{-K}$. If any sensor $h \in \mathcal{B}^h_{-K}$ then $h \geq i^\star$ i.e.,

$$\mathcal{B}^h_{-K} = \{i^\star, h_1, \ldots, h_n\} \text{ where } i^\star < h_1 < \cdots < h_n < K \tag{23}$$

$$\mathcal{B}^h = \mathcal{B}^h_{-K} \cup \{K\} = \{i^\star, h_1, h_2, \ldots, h_n, K\} \tag{24}$$

From (20), (22) and (24), we get

$$\begin{aligned} \mathcal{B} &= \mathcal{B}^l \cap \mathcal{B}^h \\ &= \{1, l_1, l_2, \ldots, i^\star\} \cap \{i^\star, h_1, h_2, \ldots, h_k, K\} \\ \Rightarrow \mathcal{B} &= \{i^\star\} \end{aligned} \tag{25}$$

Case II: $i^\star = 1$
Using (21), we get $\mathcal{B}^l_{-1} = \phi$, hence $\mathcal{B}^l = \{1\}$. Similarly, using (24), we have $\mathcal{B}^h = \{1, h_1, h_2, \ldots, h_n, K\}$ that implies

$$\mathcal{B} = \{1\} \Rightarrow \mathcal{B} = \{i^\star\} \tag{26}$$

Case III: $i^\star = K$
Using (23), we get $\mathcal{B}^h_{-K} = \phi$, hence $\mathcal{B}^h = \{K\}$. Similarly, using (22), we have $\mathcal{B}^l = \{1, l_1, l_2, \ldots, l_m, K\}$ that implies

$$\mathcal{B} = \{K\} \Rightarrow \mathcal{B} = \{i^\star\} \tag{27}$$

(25),(26) and (27) $\Rightarrow \mathcal{B}$ is a singleton set and contains the optimal sensor. □

The following definition is convenient for the proof arguments.

Definition 3 (Action Preference ($\succ_t$)). *The sensor i is optimistically preferred over sensor j in round t if:*

$$i \succ_t j := \begin{cases} C_i - C_j \leq \hat{p}_{ji}(t) + \Psi_{ji}(t) \text{ if } j < i & \text{(28a)} \\ C_j - C_i > \hat{p}_{ij}(t) + \Psi_{ij}(t) \text{ if } j > i & \text{(28b)} \end{cases}$$

B Discussion of Remark 1

The algorithm can converge to a sub-optimal sensor when we replace the term $\hat{p}_{i^\star j}(t) + \Psi_{i^\star j}(t)$ in (10b) by $\hat{p}_{i^\star j}(t) - \Psi_{i^\star j}(t)$. To verify this claim, assume algorithm selects sub-optimal sensor j in around t and $j < i^\star$ then,

$$C_{i^\star} - C_j > \hat{p}_{i^\star j}(t) - \Psi_{i^\star j}(t)$$

Since sensor $i^\star$ is not used then there is no update in $\hat{p}_{i^\star j}(t+1)$ but by definition $\Psi_{i^\star j}(t+1) > \Psi_{i^\star j}(t)$ therefore,

$$C_{i^\star} - C_j > \hat{p}_{i^\star j}(t+1) - \Psi_{i^\star j}(t+1).$$

Hence sub-optimal sensor will always be preferred over the optimal sensor in the subsequent rounds. This can be avoided by using UC term in (10b) because,

$$\hat{p}_{i^\star j}(t+1) + \Psi_{i^\star j}(t+1) > \hat{p}_{i^\star j}(t) + \Psi_{i^\star j}(t)$$

The sub-optimal sensor j will not be preferred after sufficient n rounds,

$$\Rightarrow C_{i^\star} - C_j < \hat{p}_{i^\star j}(t+n) + \Psi_{i^\star j}(t+n). \tag{29}$$

As using LC term can make the decisions stuck to sub-optimal sensor, UC term is used in (10b).

C Proof of Proposition 3

We first recall the standard Hoeffding's inequality [21, Theorem 2] that we use in the proof.

Theorem 5. *Let $X_1, \ldots, X_n$ be independent random variables with common range $[0,1]$, $\mu = \mathbb{E}[X_i]$, and $\hat{\mu}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$. Then for all $\epsilon \geq 0$,*

$$\mathbb{P}\{\hat{\mu}_n - \mu \leq -\epsilon\} \leq e^{-2n\epsilon^2} \tag{30a}$$

$$\mathbb{P}\{\hat{\mu}_n - \mu \geq \epsilon\} \leq e^{-2n\epsilon^2} \tag{30b}$$

We need the following lemmas to prove the Proposition 3.

Lemma 3.

$$\operatorname{erf}(x) = \int_0^x \mathrm{e}^{-t^2} dt = \int \mathrm{e}^{-x^2} dx \tag{31}$$

Proof. Leibniz's rule for $0 < g(x) \leq h(x) < \infty$,

$$\frac{d}{dx}\int_{g(x)}^{h(x)} f(x,t)dt = f(x,h(x))\frac{dh(x)}{dx} - f(x,g(x))\frac{dg(x)}{dx} + \int_{g(x)}^{h(x)} \frac{\partial f(x,t)}{\partial x} dt$$

Leibniz's integral rule without any common variable,

$$\frac{d}{dx}\int_{g(x)}^{h(x)} f(t)dt = f(h(x))\frac{dh(x)}{dx} - f(g(x))\frac{dg(x)}{dx}$$

Using Leibniz's rule in (31), we get

$$\frac{d}{dx}\int_0^x \mathrm{e}^{-t^2} dt = \mathrm{e}^{-x^2}\frac{dx}{dx} - 1\frac{d0}{dx} = \mathrm{e}^{-x^2}$$

$$\Rightarrow d\int_0^x \mathrm{e}^{-t^2}dt = \mathrm{e}^{-x^2}dx$$

Integrating both side,

$$\int d\int_0^x \mathrm{e}^{-t^2}dt = \int \mathrm{e}^{-x^2}dx$$
$$\Rightarrow \int_0^x \mathrm{e}^{-t^2}dt = \int \mathrm{e}^{-x^2}dx$$

□

Lemma 4. *Let $a, b, c, d \in \mathbb{R}^+$ and $t_c = b\sqrt{at} \pm \sqrt{acd}$. Then*

$$\int \mathrm{e}^{-t_c^2}dt = \mp\frac{\sqrt{\pi cd}\,\mathrm{erf}(t_c)}{\sqrt{a}b^2} - \frac{\mathrm{e}^{-t_c^2}}{ab^2} + C$$
$$\text{where } \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int \mathrm{e}^{-x^2}dx \qquad \text{(using Lemma 3)}$$

Proof. Let $x = t_c \Rightarrow x = b\sqrt{at} \pm \sqrt{acd}$. Then,

$$t = \frac{(x \mp \sqrt{acd})^2}{ab^2}$$

Now differentiate x w.r.t. t,

$$\frac{dx}{dt} = \frac{b\sqrt{a}}{2\sqrt{t}} \Rightarrow dt = \frac{2\sqrt{t}}{b\sqrt{a}}dx = \frac{2(x \mp \sqrt{acd})}{ab^2}dx$$

By changing the variable from t to x in given integral,

$$\int \mathrm{e}^{-t_c^2}dt = \int \mathrm{e}^{-x^2}\frac{2(x \mp \sqrt{acd})}{ab^2}dx$$
$$\Rightarrow \int \mathrm{e}^{-t_c^2}dt = \frac{2}{ab^2}\int x\mathrm{e}^{-x^2}dx \mp \frac{2\sqrt{cd}}{\sqrt{a}b^2}\int \mathrm{e}^{-x^2}dx \tag{32}$$

As $\int \mathrm{e}^{-cx^2}dx = \sqrt{\frac{\pi}{4c}}\,\mathrm{erf}(\sqrt{c}x) + C$ and $\int x\mathrm{e}^{-cx^2}dx = -\frac{\mathrm{e}^{-cx^2}}{2c} + C$, then (32) with $c = 1$ is,

$$= -\frac{\mathrm{e}^{-cx^2}}{ab^2} \mp \frac{\sqrt{\pi cd}\,\mathrm{erf}(x)}{\sqrt{a}b^2} + C$$
$$\Rightarrow \int \mathrm{e}^{-t_c^2}dt = \mp\frac{\sqrt{\pi cd}\,\mathrm{erf}(t_c)}{\sqrt{a}b^2} - \frac{\mathrm{e}^{-t_c^2}}{ab^2} + C$$

□

Lemma 5. *Let $a, b, c, d \in \mathbb{R}^+$ and $t_0 = cdb^{-2}$. Then*

$$\int_{t_0}^{\infty} \mathrm{e}^{-(b\sqrt{at}-\sqrt{acd})^2}dt = \frac{\sqrt{\pi cd}}{\sqrt{a}b^2} + \frac{1}{ab^2}$$

Proof. Using Lemma 4 with $t_c = b\sqrt{at} - \sqrt{acd}$.

$$\int_{t_0}^{\infty} \mathrm{e}^{-t_c^2}dt = \left(\frac{\sqrt{\pi cd}\,\mathrm{erf}(t_c)}{\sqrt{a}b^2} - \frac{\mathrm{e}^{-t_c^2}}{ab^2}\right)\Bigg|_{t_0}^{\infty}$$

Since $t_0 = cdb^{-2}$, $t_c = 0$ for $t = t_0$, $\mathrm{erf}(0) = 0$ and $\mathrm{erf}(\infty) = 1$, we get

$$\Rightarrow \int_{t_0}^{\infty} \mathrm{e}^{-t_c^2}dt = \frac{\sqrt{\pi cd}}{\sqrt{a}b^2} + \frac{1}{ab^2} = \frac{1}{b^2}\left(\sqrt{\frac{\pi cd}{a}} + \frac{1}{a}\right)$$

□

Lemma 6. *Let $b, c, d \in \mathbb{R}^+$, $\{X_t\}_{t\geq 1}$ be a sequence of independent random variables, $\hat{\mu}_t = \frac{1}{t}\sum_{s=1}^{t} X_s$, and $\mu = \mathbb{E}[X_t]$ where $X_t \in [0,1]$, $\forall t$. Then*

$$\sum_{t=1}^{n} \mathbb{P}\left\{\hat{\mu}_t - \mu \geq b - \sqrt{\frac{cd}{t}}\right\} \leq 1 + \frac{1}{b^2}\left(cd + \sqrt{\frac{\pi cd}{2}} + \frac{1}{2}\right)$$

Proof. Assume $t_0 = \lceil cdb^{-2} \rceil$ and $t_0 << n$. We divide sum of the interest into two parts as:

$$\sum_{t=1}^{n} \mathbb{P}\left\{\hat{\mu}_t - \mu \geq b - \sqrt{\frac{cd}{t}}\right\} = \sum_{t=1}^{t_0} \mathbb{P}\left\{\hat{\mu}_t - \mu \geq b - \sqrt{\frac{cd}{t}}\right\} + \sum_{t=t_0}^{n} \mathbb{P}\left\{\hat{\mu}_t - \mu \geq b - \sqrt{\frac{cd}{t}}\right\}$$

As $\mathbb{P}\{\text{any event}\} \leq 1$ and for $t > t_0$, $b - \sqrt{\frac{cd}{t}} > 0$. Using Hoeffding's inequality (30b), we get

$$\begin{aligned}\sum_{t=1}^{n} \mathbb{P}\left\{\hat{\mu}_t - \mu \geq b - \sqrt{\frac{cd}{t}}\right\} &\leq \lceil t_0 \rceil + \sum_{t=\lceil t_0 \rceil}^{n} \mathrm{e}^{-2\left(b\sqrt{t} - \sqrt{cd}\right)^2} \\ &\leq 1 + \frac{cd}{b^2} + \int_{t_0}^{\infty} \mathrm{e}^{-\left(b\sqrt{2t} - \sqrt{2cd}\right)^2} dt\end{aligned}$$

Now using Lemma 5 with $a = 2$,

$$\sum_{t=1}^{n} \mathbb{P}\left\{\hat{\mu}_t - \mu \geq b - \sqrt{\frac{cd}{t}}\right\} \leq 1 + \frac{1}{b^2}\left(cd + \sqrt{\frac{\pi cd}{2}} + \frac{1}{2}\right)$$ □

Proposition 3. *Let $f(t)$ be a positive valued increasing function such that $C = \lim_{T\to\infty} \sum_{t=1}^{T} \frac{1}{f(t)^{2\alpha}} < \infty$ in USS-UCB. For any $\theta \in \Theta_{\text{WD}}$, the mean number of times a sensor $j \neq i^\star$ is selected, is bounded as follows:*

- *for any $j < i^\star$*
$$\mathbb{E}[N_j(T)] \leq \frac{C}{2\xi_j^2},$$
- *and for any $j > i^\star$*
$$\mathbb{E}[N_j(T)] \leq 1 + \frac{1}{\xi_j^2}\left(\alpha \log f(T) + \sqrt{\frac{\pi\alpha \log f(T)}{2}} + \frac{1}{2}\right).$$

Proof. Assume $N_T(j)$ be the number of times sensor j is selected till T rounds and I_t be the sensor selected by algorithm at round t. Then mean number of pulls for any arm j is:

$$\mathbb{E}[N_T(j)] = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}_{\{I_t = j\}}\right] = \sum_{t=1}^{T} \mathbb{P}\{I_t = j\}$$

We prove the proposition by considering the case $j < i^\star$ and $j > i^\star$ separately.

- **Case I:** $j < i^\star$
  If sensor j is preferred over $i^\star$ at round t then,
  $$C_{i^\star} - C_j > \hat{p}_{ji^\star}(t) + \Psi_{ji^\star}(t) \qquad \text{(from 28b)}$$
  It is easy to verify that $C_{i^\star} - C_j = p_{i^\star j} - \xi_j$. By definition, $\hat{p}_{ji^\star}(t) = \hat{p}_{i^\star j}(t)$ and $\Psi_{ji^\star}(t) = \Psi_{i^\star j}(t)$,
  $$\hat{p}_{i^\star j}(t) + \Psi_{i^\star j}(t) < p_{i^\star j} - \xi_j \qquad \text{(from 11, 12a and 13a)}$$

If algorithm selects sensor j in round t then it is preferred over an optimal sensor in that round, i.e.,

$$\mathbb{P}\{I_t = j\} = \mathbb{P}\{I_t = j, j \succ_t i^\star\} \leq \mathbb{P}\{j \succ_t i^\star\} = \mathbb{P}\left\{\hat{p}_{i^\star j}(t) + \sqrt{\frac{\alpha \log f(t)}{\mathcal{N}_{i^\star j}(t-1)}} < p_{i^\star j} - \xi_j\right\}$$

As $\mathcal{N}_{i^\star j}(t-1)$ is a random variable, Hoeffding's inequality (30a) cannot be directly used here. Let $\hat{p}_{i^\star j,s}$ denote the value of $\hat{p}_{i^\star j}(t)$ when $\mathcal{N}_{i^\star j}(t-1) = s$. Then, we get

$$\begin{aligned}\mathbb{P}\{I_t = j\} &\leq \sum_{s=1}^{t} \mathbb{P}\left\{\hat{p}_{i^\star j,s} + \sqrt{\frac{\alpha \log f(t)}{s}} \leq p_{i^\star j} - \xi_j\right\} \\ &= \sum_{s=1}^{t} \mathbb{P}\left\{\hat{p}_{i^\star j,s} - p_{i^\star j} \leq -\left(\xi_j + \sqrt{\frac{\alpha \log f(t)}{s}}\right)\right\}\end{aligned}$$

Now using Hoeffding's inequality (30a),

$$\begin{aligned}\Rightarrow \mathbb{P}\{I_t = j\} &\leq \sum_{s=1}^{t} \mathrm{e}^{-2s\left(\xi_j + \sqrt{\frac{\alpha \log f(t)}{s}}\right)^2} \leq \sum_{s=1}^{t} \left(\mathrm{e}^{-2\xi_j^2 s}\, \mathrm{e}^{-2\alpha \log f(t)}\right) \\ &\leq \int_0^t \frac{\mathrm{e}^{-2\xi_j^2 s}}{f(t)^{2\alpha}} ds \leq \int_0^\infty \frac{\mathrm{e}^{-2\xi_j^2 s}}{f(t)^{2\alpha}} ds \\ &\leq \left(\frac{\mathrm{e}^{-2\xi_j^2 s}}{-2f(t)^{2\alpha}\xi_j^2}\right)\Bigg|_0^\infty = \frac{1}{2\xi_j^2 f(t)^{2\alpha}}\end{aligned}$$

The mean number of time a sub-optimal sensor selected in T rounds is:

$$\mathbb{E}[N_T(j)] = \sum_{t=1}^{T} \mathbb{P}\{I_t = j\} \leq \sum_{t=1}^{T} \frac{1}{2\xi_j^2 f(t)^{2\alpha}} \leq \frac{1}{2\xi_j^2} \sum_{t=1}^{\infty} \frac{1}{f(t)^{2\alpha}} = \frac{C}{2\xi_j^2}.$$

- **Case II:** $j > i^\star$
  If sensor j is preferred over $i^\star$ at round t then,

$$\begin{aligned} C_j - C_{i^\star} &\leq \hat{p}_{i^\star j}(t) + \Psi_{i^\star j}(t) && \text{(from 28a)} \\ \Rightarrow\ \hat{p}_{i^\star j}(t) + \Psi_{i^\star j}(t) &\geq p_{i^\star j} + \xi_j && \text{(from 11, 12b and 13b)}\end{aligned}$$

The mean number of time a sub-optimal sensor selected in T rounds is given by:

$$\begin{aligned}\mathbb{E}[N_T(j)] &= \sum_{t=1}^{T} \mathbb{P}\{I_t = j\} = \sum_{t=1}^{T} \mathbb{P}\{j \succ_t i^\star, I_t = j\} \\ &= \sum_{t=1}^{T} \mathbb{P}\left\{\hat{p}_{i^\star j}(t) + \sqrt{\frac{\alpha \log f(t)}{\mathcal{N}_{i^\star j}(t-1)}} \geq p_{i^\star j} + \xi_j,\ I_t = j\right\} \\ &\leq \sum_{t=1}^{T} \mathbb{P}\left\{\hat{p}_{i^\star j}(t) + \sqrt{\frac{\alpha \log f(T)}{\mathcal{N}_{i^\star j}(t-1)}} \geq p_{i^\star j} + \xi_j,\ I_t = j\right\}\end{aligned}$$

As $\mathbb{P}\{A \cap B\} \leq \min\{\mathbb{P}\{A\}, \mathbb{P}\{B\}\} \Rightarrow \mathbb{P}\{A \cap B\} \leq \mathbb{P}\{A\}$ or $\mathbb{P}\{A \cap B\} \leq \mathbb{P}\{B\}$, we get

$$\begin{aligned}&\leq \sum_{s=1}^{T} \mathbb{P}\left\{\hat{p}_{i^\star j,s} + \sqrt{\frac{\alpha \log f(T)}{s}} \geq p_{i^\star j} + \xi_j\right\} \\ &= \sum_{s=1}^{T} \mathbb{P}\left\{\hat{p}_{i^\star j,s} - p_{i^\star j} \geq \xi_j - \sqrt{\frac{\alpha \log f(T)}{s}}\right\}.\end{aligned}$$

Using Lemma 6 with $b = \xi_j$, $c = \alpha$, $d = \log f(T)$, we get

$$\mathbb{E}[N_T(j)] \leq 1 + \frac{1}{\xi_j^2}\left(\alpha \log f(T) + \sqrt{\frac{\pi\alpha \log f(T)}{2}} + \frac{1}{2}\right).$$

□

D Proof of Theorem 3

Theorem 3. *Let $f(t)$ be set as in Proposition 3. The expected regret of USS-UCB in T rounds*

- *for any instance in Θ_{WD} is bounded as*
$$\mathbb{E}\left[\mathcal{R}_T\right] \leq 3\left(3\alpha K \log f(T)\right)^{1/3} T^{2/3}.$$
- *for any instance in Θ_{SD} is bounded as*
$$\mathbb{E}\left[\mathcal{R}_T\right] \leq 4\left(\alpha K T \log f(T)\right)^{1/2}.$$

Proof. Let $N_j(T)$ is the number of times sensor j selected in T rounds. Then expected cumulative regret of USS-UCB for any instance $\theta \in \Theta_{\text{SA}}$ is:

$$\begin{aligned}\mathbb{E}\left[\mathcal{R}_T\right] &= \sum_{j\neq i^\star} \mathbb{E}\left[N_j(T)\right]\Delta_j \\ &= \sum_{j<i^\star} \mathbb{E}\left[N_j(T)\right]\Delta_j + \sum_{j>i^\star} \mathbb{E}\left[N_j(T)\right]\Delta_j\end{aligned}$$

Now, using the fact that for $j < i^\star$, $\Delta_j = \xi_j - \kappa_j$ and for $j > i^\star$, $\Delta_j = \xi_j + \kappa_j$, we get

$$\begin{aligned}\mathbb{E}\left[\mathcal{R}_T\right] &= \sum_{j<i^\star} \mathbb{E}\left[N_j(T)\right](\xi_j - \kappa_j) + \sum_{j>i^\star} \mathbb{E}\left[N_j(T)\right](\xi_j + \kappa_j) \\ &\leq \sum_{j<i^\star} \mathbb{E}\left[N_j(T)\right]\xi_j + \sum_{j>i^\star} \mathbb{E}\left[N_j(T)\right](\xi_j + \kappa_j) \\ \Rightarrow \mathbb{E}\left[\mathcal{R}_T\right] &\leq \underbrace{\sum_{j<i^\star} \mathbb{E}\left[N_j(T)\right]\xi_j}_{\mathbb{E}\left[\mathcal{R}_T^1\right]} + \underbrace{\sum_{j>i^\star} \mathbb{E}\left[N_j(T)\right](\xi_j + \beta)}_{\mathbb{E}\left[\mathcal{R}_T^2\right]},\end{aligned} \tag{33}$$

where

$$\forall j,\ \kappa_j \leq \beta \quad \begin{cases} = 0 & \text{if } \theta \in \Theta_{\text{SD}} \\ \leq 1 & \text{if } \theta \in \Theta_{\text{WD}} \text{ and sensors are ordered by their error-rate} \\ \leq 2 & \text{if } \theta \in \Theta_{\text{WD}} \text{ and sensors are arbitrarily ordered by their error-rates} \end{cases} \tag{34}$$

In the following, we set $\xi = \min\limits_{j>i^\star} \xi_j$. Now we consider the case $\xi \geq 1$ and $\xi < 1$ separately.

- **Case I:** $\xi \geq 1 \Leftrightarrow \forall j > i^\star,\ \xi_j \geq 1$
Using Proposition 3 to upper bound $\mathbb{E}\left[\mathcal{R}_T^2\right]$, we get

$$\begin{aligned}\mathbb{E}\left[\mathcal{R}_T^2\right] &\leq \sum_{j>i^\star} \left(1 + \frac{1}{\xi_j^2}\left(\alpha \log f(T) + \sqrt{\frac{\pi\alpha \log f(T)}{2}} + \frac{1}{2}\right)\right)(\xi_j + \beta) \\ &\leq K\left((\xi_j + 2) + \left(\alpha \log f(T) + \sqrt{\frac{\pi\alpha \log f(T)}{2}} + \frac{1}{2}\right)\left(\frac{1}{\xi_j} + \frac{\beta}{\xi_j^2}\right)\right) \\ \Rightarrow \mathbb{E}\left[\mathcal{R}_T^2\right] &\leq K(\xi_j + 2) + K\left(\alpha \log f(T) + \sqrt{\frac{\pi\alpha \log f(T)}{2}} + \frac{1}{2}\right)\left(\frac{1}{\xi} + \frac{\beta}{\xi^2}\right)\end{aligned} \tag{35}$$

For $\xi \geq 1$, $\left(\frac{1}{\xi} + \frac{\beta}{\xi^2}\right) \leq \beta + 1$. Further, by definition $\forall j > i^\star,\ \ \xi_j = C_j - C_{i^\star} - \mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\}$, one can easily verify that $\max\limits_{j>i^\star} \xi_j \leq C_K - C_1$ as $\mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\} \geq 0$. Assume $C_K - C_1 \leq C_1^K$, then (35) can be written as:

$$\mathbb{E}\left[\mathcal{R}_T^2\right] \leq K(C_1^K + 2) + (\beta + 1)K\left(\alpha \log f(T) + \sqrt{\frac{\pi\alpha \log f(T)}{2}} + \frac{1}{2}\right)$$

$$\leq \frac{(2C_1^K+\beta+5)K}{2} + (\beta+1)K\left(\alpha\log f(T) + \sqrt{\frac{\pi\alpha\log f(T)}{2}}\right) \tag{36}$$

For any ξ', $\mathbb{E}\left[\mathcal{R}_T^1\right]$ can written as:

$$\begin{aligned}\mathbb{E}\left[\mathcal{R}_T^1\right] &= \sum_{\substack{\xi'>\xi_j\\ j<i^\star}} \mathbb{E}\left[N_j(T)\right]\xi_j + \sum_{\substack{\xi'<\xi_j\\ j<i^\star}} \mathbb{E}\left[N_j(T)\right]\xi_j \\ &\leq \sum_{j<i^\star} \mathbb{E}\left[N_j(T)\right]\xi' + \sum_{\substack{\xi'<\xi_j\\ j<i^\star}} \frac{C}{2\xi_j^2}\xi_j \qquad \text{(using Proposition 3)} \\ \Rightarrow \mathbb{E}\left[\mathcal{R}_T^1\right] &\leq T\xi' + \frac{CK}{2\xi'} \qquad \left(\text{since } \sum_{j<i^\star}\mathbb{E}\left[N_j(T)\right] \leq T\right)\end{aligned} \tag{37}$$

From definition, $\forall j < i^\star$, $\xi_j = \mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\} - (C_{i^\star} - C_j)$. As sensors are ordered by increasing cost, one can verify that $\xi_j < 1$ for $\theta \in \Theta_{\text{WD}}$. With this fact, by combining (36) and (37), (33) can be written as:

$$\mathbb{E}\left[\mathcal{R}_T\right] \leq T\xi' + \frac{CK}{2\xi'} + \frac{(2C_1^K+\beta+5)K}{2} + 3K\left(\alpha\log f(T) + \sqrt{\frac{\pi\alpha\log f(T)}{2}}\right)$$

Choose $\xi' = \sqrt{\frac{CK}{T}}$ which maximize the upper bound and we get,

$$\mathbb{E}\left[\mathcal{R}_T\right] \leq \sqrt{2CKT} + \frac{(2C_1^K+\beta+5)K}{2} + 3K\left(\alpha\log f(T) + \sqrt{\frac{\pi\alpha\log f(T)}{2}}\right) \tag{38}$$

- **Case II:** $\xi < 1$
  Assume $T \geq T_0$ for $j > i^\star$ such that

$$1 + \frac{1}{\xi_j^2}\left(\alpha\log f(T) + \sqrt{\frac{\alpha\pi\log f(T)}{2}} + \frac{1}{2}\right) \leq \frac{2\alpha\log f(T)}{\xi_j^2} \tag{39}$$

For $\alpha = 1$ and $T_0 = 56$ (39) holds for all $T \geq T_0$. Let $0 < \xi' < \xi$. Then $\mathbb{E}\left[\mathcal{R}_T^2\right]$ can be written as:

$$\mathbb{E}\left[\mathcal{R}_T\right] \leq \sum_{\substack{\xi'>\xi_j\\ j<i^\star}} \mathbb{E}\left[N_j(T)\right]\xi_j + \sum_{\substack{\xi'<\xi_j\\ j<i^\star}} \mathbb{E}\left[N_j(T)\right]\xi_j + \sum_{\substack{\xi'>\xi_j\\ j>i^\star}} \mathbb{E}\left[N_j(T)\right](\xi_j+\beta) + \sum_{\substack{\xi'<\xi_j\\ j>i^\star}} \mathbb{E}\left[N_j(T)\right](\xi_j+\beta)$$

Since $\sum\limits_{\substack{\xi'>\xi_j\\ j<i^\star}} \mathbb{E}\left[N_j(T)\right] \leq T$ and for every $j > i^\star$, $\xi_j > \xi'$. Using Proposition 3 and (39), we get

$$\begin{aligned}\mathbb{E}\left[\mathcal{R}_T\right] &\leq T\xi' + \sum_{\xi'<\xi_j} \frac{C}{2\xi_j^2}\xi_j + \sum_{\xi'<\xi_j} \frac{2\alpha\log f(T)}{\xi_j^2}(\xi_j+\beta) \\ &\leq T\xi' + \frac{CK}{2\xi'} + 2\alpha K\log f(T)\left(\frac{1}{\xi'} + \frac{\beta}{\xi'^2}\right)\end{aligned}$$

As $C = \lim\limits_{T\to\infty}\sum\limits_{t=1}^{T}\frac{1}{t^{2\alpha}}$, one can verify that for $T_0 = 56$ and $\alpha = 1$, $C < 2\alpha\log f(T)$ holds. Using this fact,

$$\begin{aligned}&\leq T\xi' + 4\alpha K\log f(T)\left(\frac{1}{\xi'} + \frac{\beta}{\xi'^2}\right) \\ \mathbb{E}\left[\mathcal{R}_T\right] &\leq T\xi' + 4\alpha K\log f(T)\left(\frac{1}{\xi'} + \frac{\beta}{\xi'^2}\right)\end{aligned} \tag{40}$$

We first consider Θ_{WD} class of problems. For $\xi' < 1$ and $\beta \leq 2$, we have $\left(\frac{1}{\xi'} + \frac{\beta}{\xi'^2}\right) \leq \frac{\beta+1}{\xi'^2} \leq \frac{3}{\xi'^2}$. Then

$$\mathbb{E}\left[\mathcal{R}_T\right] \leq T\xi' + \frac{12\alpha K \log f(T)}{\xi'^2}$$

Choose $\xi' = \left(\frac{24\alpha K \log f(T)}{T}\right)^{1/3}$ which maximize above upper bound and we get,

$$\begin{aligned}
\Rightarrow \mathbb{E}\left[\mathcal{R}_T\right] &\leq \left(24\alpha K \log f(T)\right)^{1/3} T^{2/3} + \frac{\left(24\alpha K \log f(T)\right)^{1/3}}{2} T^{2/3} \\
&\leq 2\left(3\alpha K \log f(T)\right)^{1/3} T^{2/3} + \left(3\alpha K \log f(T)\right)^{1/3} T^{2/3} \\
\Rightarrow \mathbb{E}\left[\mathcal{R}_T\right] &\leq 3\left(3\alpha K \log f(T)\right)^{1/3} T^{2/3} \qquad (41)
\end{aligned}$$

As $C < 2\alpha \log f(T)$ and $K << T$, it is clear that upper bound in (41) is worse than (40). Hence it completes our proof for the case when any problem instance belongs to Θ_{WD}.

Now we consider any problem instance $\theta \in \Theta_{\text{SD}}$. For any $\theta \in \Theta_{\text{SD}} \Rightarrow \forall j \in [K],\ \kappa_j = 0 \Rightarrow \beta = 0$. Hence (40) can be written as

$$\mathbb{E}\left[\mathcal{R}_T\right] \leq T\xi' + \frac{4\alpha K \log f(T)}{\xi'}$$

Choose $\xi' = \left(\frac{4\alpha K \log f(T)}{T}\right)^{1/2}$ which maximize above upper bound and we get,

$$\Rightarrow \mathbb{E}\left[\mathcal{R}_T\right] \leq 2\left(4\alpha K T \log f(T)\right)^{1/2} = 4\left(\alpha K T \log f(T)\right)^{1/2} \qquad (42)$$

As $C < 2\alpha \log f(T)$ and $K << T$ then upper bound of expected regret in (40) is $3\left(\alpha K T \log f(T)\right)^{1/2}$ which is better than (42). It complete proof for second part of Theorem 3. □

E Proof of Proposition 4

Proposition 4. *Let $i < j$. Assume*

$$C_j - C_i \notin \left(\max\{0, \gamma_i - \gamma_j\}, \mathbb{P}\left\{Y^i \neq Y^j\right\}\right]. \qquad (14)$$

Then, $C_j - C_i > \max\{0, \gamma_i - \gamma_j\}$ iff $C_j - C_i > \mathbb{P}\left\{Y^j \neq Y^i\right\}$.

Proof. Assume that $C_j - C_i \geq \max\{0, \gamma_i - \gamma_j\}$. Since $C_j - C_i \notin \left[\max\{0, \gamma_i - \gamma_j\}, \mathbb{P}\left\{Y^i \neq Y^j\right\}\right]$, we get $C_j - C_i > \mathbb{P}\left\{Y^j \neq Y^i\right\}$.

The other direction follows by noting that $\mathbb{P}\left\{Y^j \neq Y^i\right\} \geq \max\{0, \gamma_i - \gamma_j\}$. □

F Proof of Proposition 5

Proposition 5. *Let $i > j$. Assume*

$$C_i - C_j \notin \left(\max\{0, \gamma_j - \gamma_i\}, \mathbb{P}\left\{Y^i \neq Y^j\right\}\right]. \qquad (15)$$

Then, $C_i - C_j \leq \max\{0, \gamma_j - \gamma_i\}$ iff $C_j - C_i \leq \mathbb{P}\left\{Y^i \neq Y^j\right\}$.

Proof. Assume that $C_i - C_j \leq \max\{0, \gamma_j - \gamma_i\}$. Since $\max\{0, \gamma_j - \gamma_i\} \leq \mathbb{P}\left\{Y^i \neq Y^j\right\}$, we get $C_j - C_i \leq \mathbb{P}\left\{Y^i \neq Y^j\right\}$.

The condition $C_j - C_i \leq \mathbb{P}\left\{Y^i \neq Y^j\right\}$ along with $C_i - C_j \notin \left(\max\{0, \gamma_j - \gamma_i\}, \mathbb{P}\left\{Y^i \neq Y^j\right\}\right]$ implies the other direction, i.e., $C_i - C_j \leq \max\{0, \gamma_j - \gamma_i\}$. □

G Proof of Proposition 6

Proposition 6. *Let $i^\star$ be an optimal sensor. Any problem instance $\theta \in \Theta'_{\text{SA}}$ is learnable if*

$$\forall\, j > i^\star \;\; C_j - C_{i^\star} \notin \left(\max\{0, \gamma_{i^\star} - \gamma_j\}, \mathbb{P}\left\{Y^i \neq Y^j\right\}\right].$$

Proof. From Proposition 4 and 5, if the optimal sensor satifies for $j > i^\star$

$$C_j - C_{i^\star} \notin \left(\max\{0, \gamma_i - \gamma_j\}, \mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\}\right]$$

and for $j < i^\star$

$$C_{i^*} - C_j \notin \left(\max\{0, \gamma_j - \gamma_i\}, \mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\}\right],$$

Then, for $j > i^\star, C_j - C_{i^\star} > \gamma_{i^\star} - \gamma_j$ iff $C_j - C_{i^\star} > \mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\}$
and for $j < i^\star, C_{i^\star} - C_j \leq \gamma_j - \gamma_{i^\star}$ iff $C_j - C_{i^\star} \leq \mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\}$. Hence we can use $\mathbb{P}\left\{Y^i \neq Y^j\right\}$ as a proxy for $\gamma_i - \gamma_j$ to make decision about the optimal arm.

Now notice that for $j < i^\star$, $C_{i^\star} - C_j \leq \gamma_j - \gamma_{i^\star} \leq \max\{0, \gamma_j - \gamma_{i^\star}\}$ (from Lemma 2). Hence for $j < i^\star$ the condition

$$C_{i^*} - C_j \notin \left(\max\{0, \gamma_j - \gamma_i\}, \mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\}\right]$$

is satisfied. Then, the condition

$$C_j - C_{i^\star} \notin \left(\max\{0, \gamma_i - \gamma_j\}, \mathbb{P}\left\{Y^{i^\star} \neq Y^j\right\}\right]$$

for $j > i^\star$ is sufficient for learnability. □

H Additional experiments for Section 5

Synthetic Datasets: The d-dimensional samples are randomly generated. Each sample is represented by $(x_1, \ldots, x_d)$ such that $\forall i,\ x_i$ is drawn from $(-1, 1)$ uniformly at random. We have generated two such datasets: Synthetic Dataset 1 with $d = 3$ and Synthetic Dataset 2 with $d = 5$. Both of these datasets have 10000 samples.

We train five linear classifiers on Synthetic Dataset 1 by varying the hyper-parameters in logistic regression and SVM. We use 80:20 train-test split of the dataset and then compute their error-rate for the whole dataset, i.e., the ratio of the total number of misclassification to total samples. The error-rate and cost of classifiers for the five problem instances are given in Table 3.

Table 3: Synthetic Dataset 1. For Case 5, WD property does not hold. Optimal classifier's cost is in red bold font.

Values/ Classifiers	Clf. 1	Clf. 2	Clf. 3	Clf. 4	Clf. 5	WD Prop.
Error-rate	0.2877	0.2448	0.2128	0.1714	0.1371	
Case 1 Costs	**0.05**	0.20	0.36	0.54	0.75	✓
Case 2 Costs	0.02	**0.045**	0.20	0.29	0.4	✓
Case 3 Costs	0.01	0.021	0.032	**0.043**	0.25	✓
Case 4 Costs	0.01	0.022	0.035	0.08	**0.1**	✓
Case 5 Costs	0.01	0.021	**0.032**	0.1	0.133	✗

Similar to Synthetic Dataset 1, we train four linear classifiers on Synthetic Dataset 2. Their error-rate and associated cost for the five problem instances are given in Table 4.

Table 4: Synthetic Dataset 2. For Case 5, WD property does not hold. Optimal classifier's cost is in red bold font.

Values/ Classifiers	Clf. 1	Clf. 2	Clf. 3	Clf. 4	WD Prop.
Error-rate	0.2340	0.1977	0.1673	0.1418	
Case 1 Costs	**0.05**	0.2	0.36	0.6	✓
Case 2 Costs	0.022	**0.045**	0.36	0.6	✓
Case 3 Costs	0.01	0.021	**0.032**	0.2	✓
Case 4 Costs	0.01	0.021	0.032	**0.045**	✓
Case 5 Costs	0.005	0.011	**0.018**	0.075	×

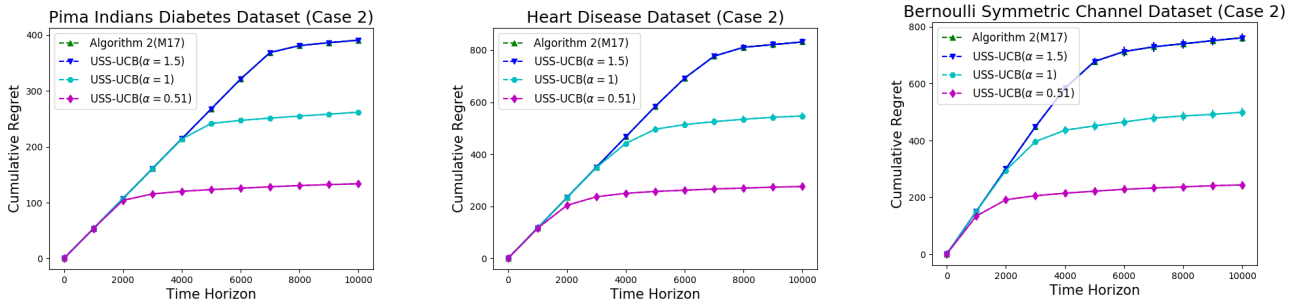


Figure 5: Comparison between Heuristic Algorithm 2 proposed in [1] and USS-UCB with parameter $\alpha = \{1.5, 1, 0.51\}$ for Case 2 of the real datasets and synthetic BSC dataset.

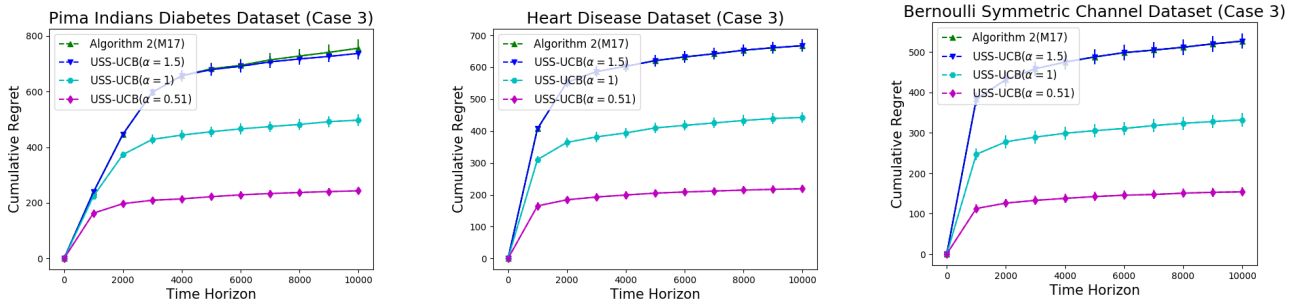


Figure 6: Comparison between Heuristic Algorithm 2 proposed in [1] and USS-UCB with parameter $\alpha = \{1.5, 1, 0.51\}$ for Case 3 of the real datasets and synthetic BSC dataset.

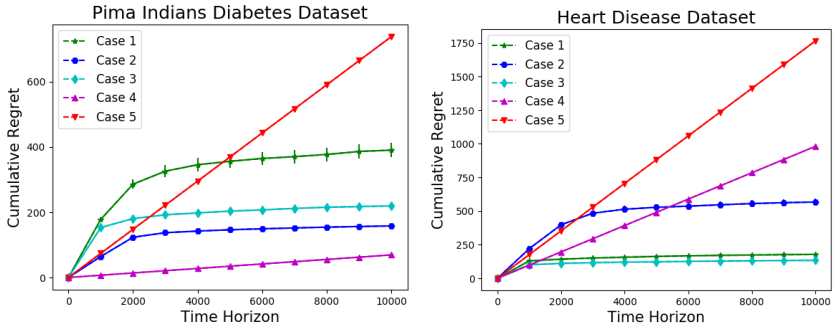


Figure 7: Cumulative regret of USS-UCB($\alpha = 0.51$) for different problem instances of the Real Datasets where last two classifier are interchanged in the sequence while keeping the cost same as given in the Table 2. Note that, $i^\star = K$ for Case 4 and WD automatically holds but after interchanging last two classifiers, WD does not hold for Case 4.

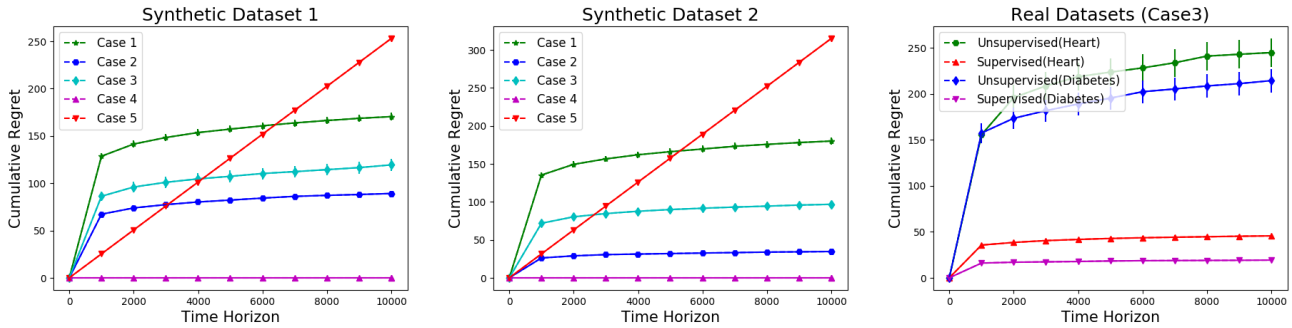


Figure 8: Cumulative regret of USS-UCB($\alpha = 0.51$) for different problem instances of the Synthetic Dataset 1 and 2. Rightmost figure: Comparison between unsupervised and supervised setting for Case 3 of real datasets.

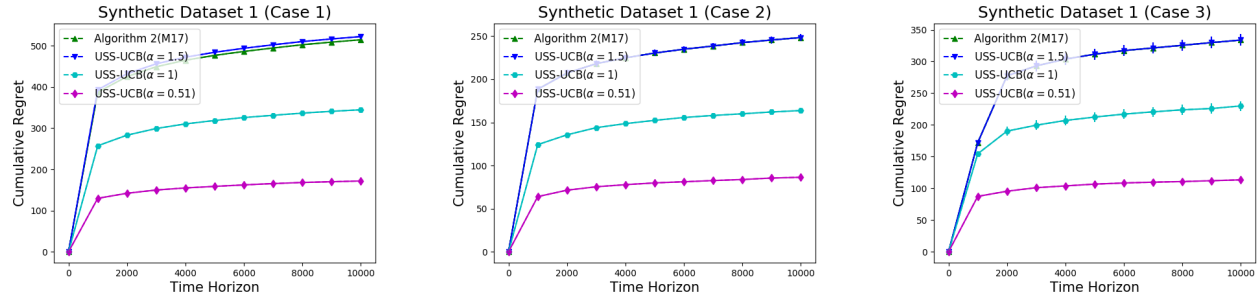


Figure 9: Comparison between Heuristic Algorithm 2 proposed in [1] and USS-UCB with parameter $\alpha = \{1.5, 1, 0.51\}$ for Synthetic Dataset 1.

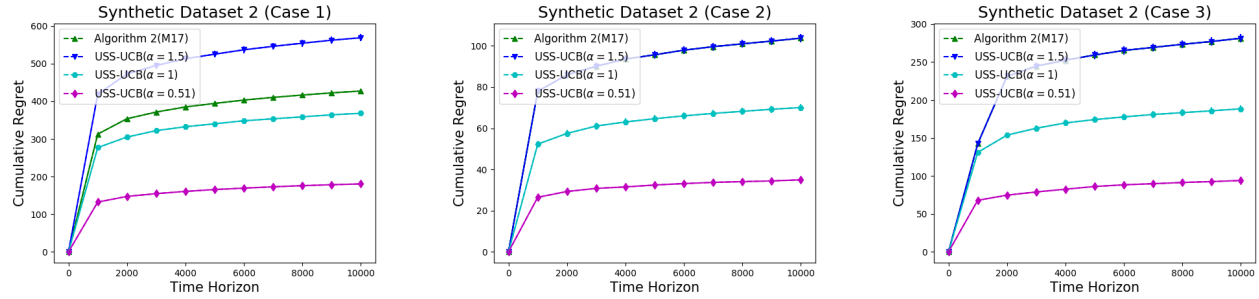


Figure 10: Comparison between Heuristic Algorithm 2 proposed in [1] and USS-UCB with parameter $\alpha = \{1.5, 1, 0.51\}$ for Synthetic Dataset 2.

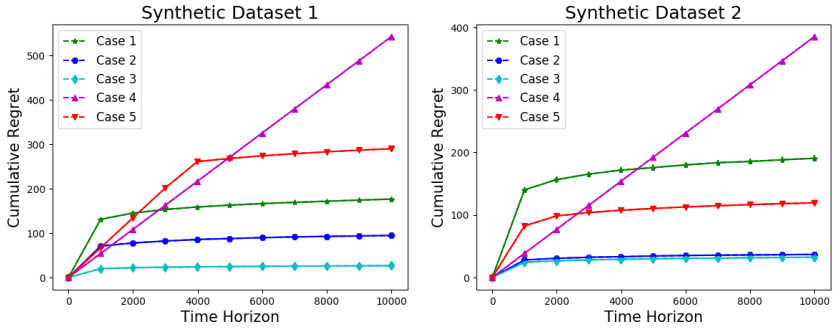


Figure 11: Cumulative regret of USS-UCB($\alpha = 0.51$) for different problem instances of the Synthetic Dataset 1 and 2 with last two classifier are interchanged in the sequence while keeping the cost same as given in the Table 3 and 4. Note that, $i^\star = K$ for Case 4 and WD automatically holds but after interchanging last two classifiers, WD does not hold for Case 4 whereas holds for Case 5.