Decision Making Problems with Funnel Structure: A Multi-Task Learning Approach with Application to Email Marketing Campaigns

Ziping Xu
University of Michigan

Amirhossein Meisami
Adobe Inc.

Ambuj Tewari
University of Michigan

Abstract

This paper studies the decision making problem with *Funnel Structure*. Funnel structure, a well-known concept in the marketing field, occurs in those systems where the decision maker interacts with the environment in a layered manner receiving far fewer observations from deep layers than shallow ones. For example, in the email marketing campaign application, the layers correspond to Open, Click and Purchase events. Conversions from Click to Purchase happen very infrequently because a purchase cannot be made unless the link in an email is clicked on.

We formulate this challenging decision making problem as a contextual bandit with funnel structure and develop a multi-task learning algorithm that mitigates the lack of sufficient observations from deeper layers. We analyze both the prediction error and the regret of our algorithms. We verify our theory on prediction errors through a simple simulation. Experiments on both a simulated environment and an environment based on real-world data from a major email marketing company show that our algorithms offer significant improvement over previous methods.

1 Introduction

We consider decision making problems arising in online recommendation systems or advertising systems (Pescher et al., 2014; Manikrao and Prabhakar, 2005). Traditional approaches to these problems only optimize a single reward signal (usually purchase or final conversion), whose positive rate can be extremely low

in some real recommendation systems. This reward sparsity can lead to a slow learning speed and unstable models. Nevertheless, some non-sparse signals are usually available in these applications albeit in a layered manner. These signals can be utilized to boost the performance of the final sparse signal. As a special case, funnel structure generates a sequence of binary signals by layers and the observations are cumulative products of the sequence. An example of email conversion funnel is shown in Figure 1.

Funnel structure characterizes a wide range of problems in advertising systems. In the email campaign problem, the learning agent decides the time to send emails to maximize purchases. Apart from the final reward on purchase, we also observe the opening and clicking status of an email. There are also papers studying the participation funnel in MOOCs (Clow, 2013; Borrella et al., 2019). Students go through the layers of Awareness, Registration, Activity, Progress and Completion until they drop or complete the course. For both of the funnels, the drop-off fraction at each layer is large. For example, in email campaigns, the conversion rates are typically 10% for Open, 4% for Click and 2.5% for Purchase.

Funnel structure studies in the marketing field. Conversion funnel has been at the center of the marketing literature for several decades (Howard and Sheth, 1969; Barry, 1987; Mulpuru, 2011). This line of work focuses on the attribution of advertising effects and is more interested in analyzing buyers' behavior at each layer. Schwartz et al. (2017) learns contextual bandit with a Funnel Structure. However, their model directly learns on the final purchase signal and signals on other stages are only used for performance evaluation. Hence, it lacks a comprehensive method that exploits the structural information of a funnel.

Multi-task learning. Learning on a funnel structure can be seen as a multi-task learning problem, where predicting the conversion on each stage is a single task. Previous literature analyzed how learning among multiple similar tasks can improve sample effi-

ciency (Evgeniou and Pontil, 2004; Ruder, 2017; Zhang and Yang, 2017). Most of the multi-task learning analyses require an assumption on the similarity among tasks. Evgeniou and Pontil (2004) assumes that true parameters of different tasks are centered at some unknown point, which we adopted in our method. We also consider a sequential dependency among the tasks, which imposes restrictions on the unknown parameters between adjacent layers.

Analyses of multi-task learning often assume balanced sample sizes for different tasks. Otherwise, their bounds depend on the harmonic mean of sample sizes $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}$, which can be large when sample sizes are imbalanced as in funnel structures. Furthermore, most of the multi-task learning algorithms minimize the average losses over the task set, which is *not* our primary goal.

Contextual bandits. The decision making problem is modelled as a contextual bandit problem (Li et al., 2010, 2011; Beygelzimer et al., 2011) in our paper. Previous works on contextual bandits mainly focus on a single reward. Drugan and Nowe (2013); Turğay et al. (2018) study multi-objective bandits by considering a Pareto regret, which optimizes the vector of rewards for different objectives, while our work focuses on optimizing the final reward by exploiting the whole task set.

Related works This work is the first work that applies the multi-task learning to contextual bandits with funnel structure. There has been works considering contextual bandits with a sequential transfer (Lazaric et al., 2013; Soare et al., 2014), a multi-task learning approach (Deshmukh et al., 2017) and a multi-goal setting (Drugan and Nowe, 2013; Turğay et al., 2018). However, none of their algorithms adapts to the funnel structure due to the imbalance in the sample sizes of different layers. On the other side, Schwartz et al. (2017) focuses on funnel structure contextual bandits without multi-task learning. Some Reinforcement Learning algorithms with auxiliary rewards (Jaderberg et al., 2016; Lin et al., 2019) can also be applied to this problem. This line of work does not use the full information as it either learns on the output of the last layer or learns on an weighted average of different layers. Our experiments also verify that using the full information could improve performances.

Our contributions. Our main contributions are summarized below:

1. We formally formulated the important *Funnel Structure* problem from the marketing field.

2. We proposed a multi-task learning algorithm for

contextual bandits with funnel structure and analyzed the prediction errors and regret under various similarity assumptions.

3. Our prediction error bound was verified on a simulated environment.

4. Our algorithm improves previous methods on both simulated and real-data contextual bandit environment.

2 Formulation

In this section, we introduce the formulation for funnel structure and discuss how the formulation applies to our email campaign problem. We also introduce the generalized linear model and the assumptions for our theoretical analyses.

Funnel structure. A funnel, denoted by $F = \{J, \mathcal{X}, (Z_1, \ldots, Z_J)\}$, consists of the number of layers $J \in \mathbb{N}$, feature space $\mathcal{X} \in \mathbb{R}^d$ for some $d > 0$ and a sequence of J mappings $(Z_1, \ldots, Z_J)$. Each Z_j is a mapping from feature space to $[0, 1]$. On each interaction, a funnel takes an input feature $x \in \mathcal{X}$ and generates a sequence of binary variables $z_1, \ldots, z_J$ from Bernoulli distributions with parameters $Z_1(x), \ldots, Z_J(x)$, respectively. Then it returns $r_1, \ldots r_J$, for $r_j = \prod_{s=1}^{j} z_s$, to the learning agent.

Email conversion funnel. We illustrate how the formulation applies to the email conversion funnel. Our email conversion funnel, as shown in Figure 1, has 3 layers representing Open, Click and Purchase, respectively. Every email sent to a user randomly generates r_1, r_2, r_3 representing whether the email is actually opened, clicked or purchased using the mechanism described above, while z_1, z_2, z_3 are the indicators for the three events given previous events happened.

On the sparsity of the funnel, if an email is never opened, neither click or purchase could happen. Out of all the emails sent to users, 10% of them were opened, 0.4% were clicked, and 0.01% led to a purchase. More generally, when there exists a $r_j = 0$, all the successors r_i's, $i > j$, become 0, which leads to unobservable $z_{j+1}, \ldots, z_J$. On average, given a feature x, the probability of observing z_j is $P_{j-1}(x)$, which decreases **exponentially** as the layers go deeper.

Contextual bandit with funnel structure. Our contextual bandit with funnel structure is denoted by $M = \{\mathcal{A}, \mathcal{X}, P_x, \{F_a\}_{a \in \mathcal{A}}\}$, where $\mathcal{A}$ is the finite action space with $|\mathcal{A}| = A$, $\mathcal{X}$ is the common context space, P_x is the context distribution and each arm $a \in \mathcal{A}$ is assigned a funnel denoted by $F_a = \{J, \mathcal{X}, (Z_1^a, \ldots, Z_J^a)\}$.
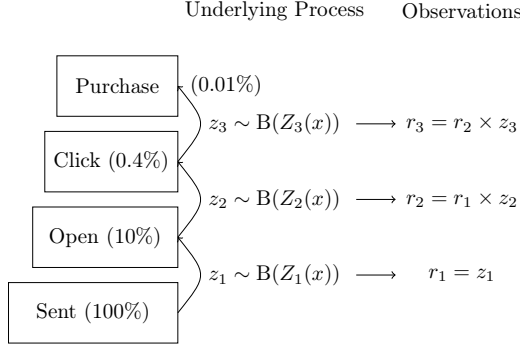
Ziping Xu, Amirhossein Meisami, Ambuj Tewari

Figure 1: An illustration of the email conversion funnel. Given any input x, the profile information of the user, the funnel generates, $z_1, \ldots z_3$, from Bernoulli distributions with parameter $Z_1(x), Z_2(x), Z_3(x)$, representing whether the email would be opened, clicked or purchased given the conversion of the previous layers happened. The observations $r_1, \ldots, r_3$ represent whether the email is actually opened, clicked or purchased, respectively.

On each round of t-th interaction, the environment generates a context $x_t \sim P_x$, the agent takes an action a_t and the funnel F_{a_t} returns the reward vector $(r_{1t}, \ldots, r_{Jt})$ taken the input context x_t based on the process described above.

Note that our setting, when $J = 1$, differs from the contextual bandit setting in Chu et al. (2011); Filippi et al. (2010), where each arm has a unique context but the same mapping from context to reward function. Filippi et al. (2010) assumes a canonical exponential family density function. We consider a binomial signal, whose density might not be in canonical exponential family.

Assumptions. Most analyses on multi-task learning assume some similarities among tasks set to allow knowledge transfer. Here we assume a generalized linear model (GLM) and a prior-known hypothesis class over the unknown parameters for all the layers. The hypothesis class characterizes the relatedness across layers.

Assumption 1 (Generalized linear model)**.** *Assume all $Z_j = \mu(x^T \theta_j^*)$ for some mean function $\mu : \mathbb{R} \mapsto [0, 1]$ and θ_j^* is the true parameter of layer j. A throughout example of this paper is the model for logistic regression, where $\mu(y) = 1/(1 + \exp(-y))$.*

We also assume that the mean function μ is Lipschitz continuous and convex.

Assumption 2. *We assume that μ is monotonically increasing and $\mu'(x) \geq c_\mu$ for all $x \in \mathcal{X}$. We also require that function μ satisfies $|\mu'(x)| \leq \kappa$.*

Assumption 3. *Assume $\mathcal{X} \subset \{x \in \mathbb{R} : \|x\| \leq d_x\}$.*

Generally, we assume that the joint parameter is from a hypothesis class. Two special cases of interest are introduced, upon which we design our practical algorithms.

Assumption 4 (Similarity assumption)**.** *Let $\boldsymbol{\theta} = (\theta_1^T, \theta_2^T, \ldots, \theta_j^T)^T \in \mathbb{R}^{dJ}$ and $\boldsymbol{\theta}^*$ is the joint vector for the true parameters. We assume $\boldsymbol{\theta}^* \in \Theta_0 \subset \mathbb{R}^{dJ}$. Throughout the paper, we discuss two special cases:*

1. ***Sequential dependency:*** $\Theta_0 := \{\boldsymbol{\theta} \in \mathbb{R}^{dJ} : \|\theta_j - \theta_{j-1}\|_2 \leq q_j, \text{ for } j > 1 \text{ and } \|\theta_1\| \leq q_1 \text{ for some } q_1, \ldots q_J \in \mathbb{R}^+.$

2. ***Clustered dependency:*** $\Theta_0 := \{\boldsymbol{\theta} \in \mathbb{R}^{dJ} : \exists \theta_0 \in \mathbb{R}^d, \|\theta_j - \theta_0\|_2 \leq q_j, \forall j \in [J]\} \text{ for some } q_1, \ldots q_J \in \mathbb{R}^+.$

For any set $\Theta \subset \mathbb{R}^{dJ}$, we denote the marginal set of task j by $\Theta[j]$ i.e., $\Theta[j] = \{\theta \in \mathbb{R}^d : \exists \boldsymbol{\theta} \in \Theta, \boldsymbol{\theta}_j = \theta\}$.

We first note that a hypothesis class over the joint parameters is a common assumption in multi-task learning literature (Maurer et al., 2016; Zhang and Yang, 2017; Pentina et al., 2015). Also, in another line of work focusing on transfer learning, the theoretical analyses often assume a discrepancy between tasks (Wang et al., 2019). We argue in Appendix A.1 that under our GLM assumptions, the discrepancy assumption is almost the same as ours.

3 Supervised learning

Before discussing the contextual bandits with funnel structure, we first consider the supervised learning scenario for a single funnel and seek a bound for the prediction error of each layer:

$$\text{PE}_j = |\mu(x^T \theta_j^*) - \mu(x^T \hat{\theta}_j)|,$$

for some estimates $\hat{\theta}_1, \ldots, \hat{\theta}_J \in \mathbb{R}^d$.

For a single funnel, our algorithm learns on the dataset $\{x_i, r_{1i}, \ldots, r_{Ji}\}_{i=1}^n$ of size n. Let $n_j = \sum_{i=1}^n \mathbf{1}(r_{j-1,i} = 1)$ be the number available observations for layer j and $j_1, \ldots j_{n_j}$ be the indices of these n_j samples, i.e. $r_{j-1,j_i} > 0$ for all $i \in [n_j]$. Let $z_{j,j_i} = r_{j,j_i}/r_{j-1,j_i}$. We denote the square loss function of layers j by

$$l_j(\theta) := \sum_{i=1}^{n_j} (z_{j,j_i} - \mu(x_{j_i}^T \theta))^2. \qquad (1)$$

3.1 Implications from a single-layered case

We first investigate a single-layered case and see how prior knowledge helps improve the upper bound on

prediction error. Lemma 1 bounds the prediction error using either prior knowledge or collected samples. The proof of Lemma 1 is provided in Appendix A.2.

Lemma 1. *Using the model defined above, assume $J = 1$ and the true parameter $\theta^* \in \Theta_0$. Let the dataset be $\{x_i, z_i\}_{i=1}^n$ and let $\tilde{\theta}$ be the solution that maximizes the L_2 function in (1) and $\hat{\theta}$ be its projection onto Θ_0. Let $q := \sup_{\theta_1, \theta_2 \in \Theta_0} \|\theta_1 - \theta_2\|_2$. Then with a probability at least $1 - \delta$, we have*

$$|\mu(x^T\hat{\theta}) - \mu(x^T\theta^*)|$$

$$\leq \kappa \min\{\sup_{\theta_1, \theta_2 \in \Theta_0} |x^T(\theta_1 - \theta_2)|, \|x\|_{M_n^{-1}} \frac{4c_\delta}{c_\mu} \sqrt{\frac{d}{n \vee 1}}\}$$

$$\leq \kappa \min\{d_x q, \|x\|_{M_n^{-1}} \frac{c_\delta}{c_\mu} \sqrt{\frac{d}{n}}\}, \tag{2}$$

furthermore, we have a confidence set on θ^*

$$\theta^* \in \{\theta : \|\hat{\theta} - \theta\|_{M_n} \leq \frac{c_\delta}{c_\mu} \sqrt{\frac{d}{n}}\}, \tag{3}$$

where $M_n = \frac{1}{n} \sum_{i=1}^n x_i x_i^T$, $c_\delta = 80 d_x \sqrt{2 \ln(8/\delta)}$.

Equation (2) bounds the prediction error with a minimum of two terms. The first term in (2) is directly derived from prior knowledge. The second term is a parametric bound without any regularization (Srebro et al., 2010). In fact, (2) is a tight upper bound on prediction error as shown in Appendix A.3.

Lemma 1 implies a **"transfer or learn"** scenario: when the sample size for the new task is not large enough, i.e. $n = o(1/q^2)$, it is more beneficial to directly apply the prior knowledge. Otherwise, one can drop the prior knowledge and use parametric bound.

3.2 Multi-task learning algorithm

Our algorithm, inspired by the "transfer or learn" idea, consists of two steps: 1) optimize the loss function for each layer within its marginal set and calculate the confidence set defined in (3); 2) take the intersection between the confidence set and Θ_0, which generates Θ_1. Then project the unconstrained solution onto the new set Θ_1. The details are shown in Algorithm 1, where $\text{Proj}_\Theta(\theta)$ denotes the projection of θ onto the set Θ.

3.3 Upper bound on prediction error

Directly applying Lemma 1 within the set Θ_1 gives us a bound on prediction error depending on the marginal set $\Theta_1[j]$ and n_j as shown in Corollary 1.

Corollary 1. *Let $\hat{\theta}_1, \ldots, \hat{\theta}_J$ be the estimates from Algorithm 1 and Θ_1 be the set defined in Algorithm 1.*

Algorithm 1 Regularized MTL for funnel structure

Input: number of layers J, hypothesis set Θ_0, dataset $\{x_i, r_{1i}, \ldots, r_{Ji}\}_{i=1}^n$ generated from the funnel, accuracy $\delta > 0$.

\# *Calculate confidence set.*

for $j = 1$ to J **do**

   Solve $\bar{\theta}_j = \text{Proj}_{\Theta_0[j]}(\arg\min l_j(\theta))$.

   Calculate $\hat{\Theta}_j$ defined in Equation (3) by

$$\hat{\Theta}_j = \{\theta : \|\theta - \bar{\theta}_j\|_{M_{j,n_j}} \leq \frac{c_\delta}{c_\mu} \sqrt{\frac{d}{n_j}}\},$$

   for $M_{j,n_j} = \sum_{i=1}^{n_j} x_{j,j_i} x_{j,j_i}^T$.

end for

\# *Re-estimate parameters.*

Calculate joint set $\hat{\Theta} = \{\boldsymbol{\theta} : \boldsymbol{\theta}_j \in \hat{\Theta}_j \text{ for all } j \in [J]\}$.

Set $\Theta_1 \leftarrow \Theta_0 \cap \hat{\Theta}$.

for $j = 1$ to J **do**

   Solve $\hat{\theta}_j = \text{Proj}_{\Theta_1[j]}(\arg\min l_j(\theta))$.

end for

return $\hat{\theta}_1, \ldots, \hat{\theta}_J$.

With a probability at least $1 - \delta$, for all $j \in [J]$, we have

$$PE_j \leq \kappa \min\{\sup_{\theta_1, \theta_2 \in \Theta_1[j]} |x^T(\theta_1 - \theta_2)|,$$

$$\|x\|_{M_n^{-1}} \frac{c_\delta}{c_\mu} \sqrt{\frac{d}{n_j}}\}. \tag{4}$$

However, it is more interesting to discuss the actual form of $\Theta_1 = \Theta_0 \cap \hat{\Theta}$ and its interactions with n_j under some special assumptions on Θ_0. As mentioned in Assumption 4, we consider two cases: sequential dependency and clustered dependency.

Recall that for the sequential dependency, we assume $\boldsymbol{\theta}^* \in \Theta_0 := \{\boldsymbol{\theta} \in \mathbb{R}^{dJ} : \|\theta_j - \theta_{j-1}\|_2 \leq q_j, \text{ for } j > 1 \text{ and } \|\theta_1\|_2 \leq q_1\}$ for some $q_1, \ldots q_J > 0$. Before presenting our results, we need an extra assumption on the distribution of covariates.

Assumption 5. *Assume the minimum eigenvalue of M_{j,n_j} is lower bounded by a constant $\lambda > 0$ for all $j \in [J]$.*

Assumption 5 guarantees that the distribution of the covariate covers all dimensions.

Theorem 1 (Prediction error under sequential dependency)**.** *For any funnel with a sequential dependency of parameters $q_1, \ldots, q_J$, let $\hat{\theta}_1, \ldots, \hat{\theta}_J$ be the estimates from Algorithm 1. If $n_{j+1} \leq n_j/4$, $q_1 \geq, \ldots, \geq q_J$ and Assumption 5 is satisfied, then with a probability at*

least $1 - \delta$, *for any* $j_0 \in [J]$, *we have*

$$
PE_j \leq \begin{cases} \kappa\|x\|_2 \frac{c_{\delta/J}}{c_\mu\lambda}\sqrt{\frac{d}{n_j}}, & \text{if } j < j_0, \\ \kappa\|x\|_2(\frac{c_{\delta/J}}{c_\mu\lambda}\sqrt{\frac{d}{n_{j_0}}} + \sum_{i=j_0+1}^{j} q_i), & \text{if } j \geq j_0, \end{cases}
$$

where we let $n_0 = \infty$. *The smallest bound of all choices of* j_0 *is achieved when* j_0 *is the smallest* $j \in [J]$, *such that*

$$
\frac{c_\delta\sqrt{d}}{c_\mu\lambda}(\frac{1}{\sqrt{n_j}} - \frac{1}{\sqrt{n_{j-1}}}) \geq q_j, \tag{5}
$$

if none of j's *in* $[J]$ *satisfies (5),* $j_0 = J + 1$.

Theorem 1 shows that for some funnel under sequential dependency assumption, there exists a threshold layer j_0, before which the bounds without multi-task learning are tighter. After j_0, we use the bounds depending on prior knowledge. For small n_j's, $j_0 = 1$ and for sufficient large n_j's, $j_0 = J + 1$. Figure 2 shows an example of how the threshold j_0 changes when total number n increases in a 5-layered funnel.
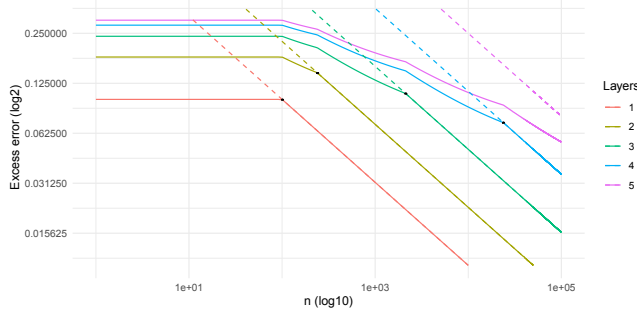

Figure 2: An example of how the prediction error bound in Theorem 1 changes when the number of observations increases in a 5-layered funnel. We set $\|x\|_2 c_\delta\sqrt{d}/(c_\mu\lambda) = 1$, $q_j = (12 - 2j)/100$ and $n_j = 0.2^{j-1}n$. Solid lines mark the prediction error bound defined in Theorem 1 and dashed lines mark the prediction error without multi-task learning (second term in (3)). Black points marked the change of j_0.

For the clustered dependency, the prediction error bound can be characterized by Theorem 2.

Theorem 2 (Prediction error under clustered dependency). *For any funnel with a clustered dependency of parameters* $q_1, \ldots, q_J$, *let* $\hat\theta_1, \ldots, \hat\theta_J$ *be the estimates from Algorithm 1. With a probability at least* $1 - \delta$,

$$
PE_j \leq \kappa\|x\|_2 \min\{\frac{c_{\delta/J}}{c_\mu\lambda}\sqrt{\frac{d}{n_{j_0}}} + q_j, \frac{c_{\delta/J}}{c_\mu\lambda}\sqrt{\frac{d}{n_j}}\},
$$

where $j_0 = \arg\min_{j\in[J]} \frac{c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n_j}} + q_j$.

Under the clustered dependency, there is a single layer that gives the tightest confidence set on the unknown center, which is used by all the other layers.

4 Regret Analysis for Contextual Bandit

In this section, we bound regrets for contextual bandits with funnel structure and discuss the benefits of multi-task learning.

Extra notations. For simpler demonstration, we define $P_j : \mathcal{X} \mapsto [0, 1]$, such that $P_j(x) = \prod_{i=1}^{j} Z_i(x)$. For any $\boldsymbol\theta = (\theta_1^T, \theta_2^T, \ldots, \theta_j^T)^T \in \mathbb{R}^{dJ}$, let $P_j(x, \boldsymbol\theta) = \prod_{i=1}^{j} \mu(x^T\theta_j)$. To account for multiple funnels, we let $n_{a,j}^t$ be the number of observations for the j-th layer of funnel F_a up to step t. Let $\theta_{a,j}^*$ be the true parameters and $\boldsymbol\theta_a^*$ be the joint vector. Further we let $\lambda_{a,j}^t$ be the sample minimum eigenvalue covariance matrix of layer j of funnel F_a, $\lambda_{a,j}$ be the minimum eigenvalue of its expectation and $\bar\lambda$ be a lower bound over all a and j.

Regret of contextual bandits with funnel structure is defined as

$$
\sum_{t=1}^{T} \left[P_J(x_t, \boldsymbol\theta_{a_t^*}^*) - P_J(x_t, \boldsymbol\theta_{a_t}^*) \right],
$$

where a_t^* is the optimal action for input x_t, a_t is the the action chosen by the agent at step t and T is the total steps.

4.1 Optimistic algorithm

We propose a variation of the famous UCB (upper confidence bound) algorithm that adds bonuses based on the uncertainty of the whole funnel. The prediction error of funnel F_a of a given input x is $|P_J(x, \hat{\boldsymbol\theta}_a^t) - P_J(x, \boldsymbol\theta_a^*)|$ for $\hat{\boldsymbol\theta}_a^t$ which is the joint vector of estimates $\hat\theta_{a,j}^t$ at the step t.

From Lemma 1, we define $\Delta\mu_{a,j}^t$ by

$$
\kappa\|x_t\|_2 \min\{\sup_{\theta_1,\theta_2\in\Theta_{a,1}^t} \|\theta_1 - \theta_2\|_2, \frac{c_{\delta/3AJT}}{c_\mu\lambda_{a,j}^t}\sqrt{\frac{d}{n_{a,j}^t \vee 1}}\},
$$

where $\Theta_{a,1}^t$ is the intersection set from Algorithm 1 for funnel a at step t. Using simple Taylor's expansion, we have Lemma 2.

Lemma 2. *Using the estimates from Algorithm 1, we have, with the same probability in (4),*

$$
|P_J(x, \hat{\boldsymbol\theta}_a^t) - P_J(x, \boldsymbol\theta_a^*)| \leq
$$

$$
\sum_j \frac{P_J(x, \hat{\boldsymbol\theta}_a^t)}{\mu(x^T\hat\theta_{a,j}^t)}\Delta\mu_{a,j}^t + \sum_{i\neq j} \Delta\mu_{a,j}^t\Delta\mu_{a,i}^t =: \Delta\mu_a^t. \tag{6}
$$

Algorithm 2 Contextual Bandit with a Funnel Structure

$t \to 1$, total number of steps T, memory $\mathcal{H}_a = \{\}$ for all $a \in [A]$. Initialize $\theta_{a,\star}$ with zero vectors.

for $t = 1$ to T **do**

    Receive context x_t.

    Compute $P_J^+(x_t, \hat{\boldsymbol{\theta}}_a)$ based on (6) for all $a \in [A]$.

    Choose $a_t = \arg\max_{a \in \mathcal{A}} \hat{P}_J^+(x_t, \hat{\theta}_{a,j})$.

    Receive $r_{t,1}, \ldots, r_{t,J}$ from funnel F_{a_t}.

    Set $\mathcal{H}_{a_t} \to \mathcal{H}_{a_t} \cup \{(x_t, (r_{t,1}, \ldots, r_{t,J}))\}$.

    Update $\hat{\theta}_{a_t,\star}$ by algorithm 1 with dataset $\mathcal{H}_{a_t}$.

end for

Define $P_J^+(x, \hat{\boldsymbol{\theta}}_a^t) = P_J(x, \hat{\boldsymbol{\theta}}_a^t) + \Delta\mu_a^t$. We are able to derive an optimistic algorithm shown in Algorithm 2. Now we use the prediction error on the whole funnel to analyze the regret.

4.2 Regret analysis

Theorem 3 bounds regrets for Algorithm 2. The regret in Theorem 3 can be bounded by three terms in (7). The first term in (7) represents the normal regret without any multi-task learning with an order $\mathcal{O}(\sum_{a,j} \sqrt{n_{a,j}^T})$, which reduces to the standard $\sqrt{AT}$ when $J = 1$. Note that the impact of one layer on regret is bounded with the square root of its number of observations. The second term is a constant term that does not depend on T. The third term represents the benefits of multi-task learning. Full version of Theorem 3 is given and proved in Appendix A.5.

Theorem 3. *Using Algorithm 2, under the Assumptions 1-4, with high probability, the total regret*

$$\sum_{t=1}^T \left[P_J(x_t, \boldsymbol{\theta}_{a_t^*}^*) - P_J(x_t, \boldsymbol{\theta}_{a_t}^*) \right]$$

$$= \mathcal{O}(c_0 \sum_{a,j} \sqrt{n_{a,j}^T} + \sum_{a,j} \frac{c_0^2 J d_x^4}{\bar{p}_{a,j}^2}) - \sum_{a,j} \Delta_{a,j} \quad (7)$$

where $\mathcal{O}$ ignores all the constant terms and logarithmic terms for better demonstrations, $c_0 = (\kappa d_x c_{\delta/3AJT} \sqrt{d})/(c_\mu \bar{\lambda})$, $\bar{p}_{a,j} := \mathbb{E}_x P_{j-1}(x, \boldsymbol{\theta}_a^*)$ and

$$\Delta_{a,j} = \sum_{t=1; a_t=a}^T P_j(x_t^T \hat{\theta}_{a_t}^t) \left[c_0 \frac{1}{\sqrt{n_{a,j}^t \vee 1}} - \Delta\mu_{a,j}^t \right].$$

represents the benefits of transfer learning.

We discuss the actual form of benefits under sequential dependency. If the hypothesis class is truly sequential dependency with parameters $q_{a,1}, \ldots, q_{a,J}$ for each funnel F_a, we have

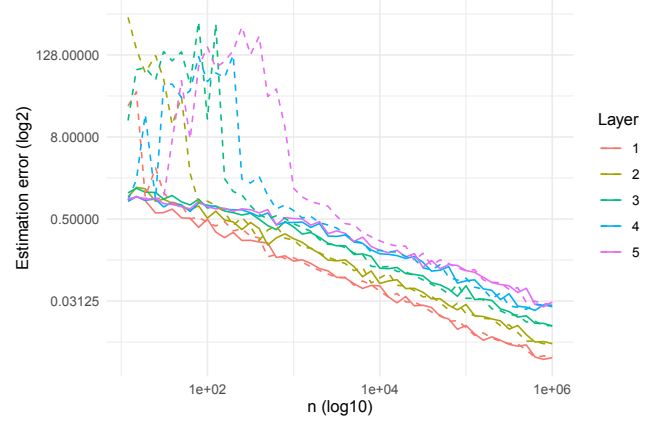$$\Delta_{a,j} = \mathcal{O}(\sum_{j' \le j} 1/q_{a,j'}),$$

Figure 3: Estimation errors (L_2 distance to θ_j^*) of $\bar{\theta}_j$ and $\hat{\theta}_j$ under different number of interactions with the funnel. Colors represents the layers. Solid (Dashed) lines represents the estimation errors of $\hat{\theta}_j$ ($\bar{\theta}_j$). Each point in the plot is an average over 10 independent runs.

for sufficient large T. Generally, for sufficient large total steps, the benefits scale with $\sum_{a,j} (J - j + 1)/q_{a,j}$.

5 Experiments

In this section, we first present a simulation on the power of multi-task learning using Algorithm 1. Then we propose a more practical contextual bandit algorithm, which is tested on a simulated environment and our real-data email campaign environment.

5.1 Simulation on supervised learning

We use a simulation to verify the bound in Theorem 1 and the curves in Figure 2. We consider a 5-layered funnel with sequential dependency. The link function is $\mu(x) = 1/(1 + \exp(-x))$. We set $d = 5$ and $\theta_{j+1} = \theta_j + u_j q_j$, where $u_j \in \mathbb{R}^d$ is a unit vector with a random direction and $q_j = 1.2 - 0.2j$. In the simulation, we apply Algorithm 1 under the sequential dependency and calculate the estimation error of the estimates without parameter transfer ($\bar{\theta}_j$) and the final estimates $\hat{\theta}_j$. The results are shown in Figure 3. We observe a similar pattern as in Figure 2 and the errors of deeper layers are controlled well despite of their small sample sizes.

5.2 Simulations on contextual bandits

Practical algorithm. Calculating the intersection between two sets is not easy when Θ_0 has a complex form. Also, we may not have access to Θ_0 in real

data analysis. We, therefore, develop practical algorithms especially for the sequential dependency and clustered dependency. Both of the algorithms reduced to optimizing parameters under a L_2 regularization. An equivalent form is to optimize under L_2 penalty. Our practical contextual bandit algorithm optimize loss function using an L_2 penalty controlled by tuned hyperparameters (Algorithm 3 in Appendix B.1). Since exploration is not the primary interest of this work, we also adopt a ϵ-greedy exploration for the simplicity of implementation.

Compared algorithms. Apart from the naive algorithm that directly learns on the signals r_J, some methods learn on the averaged rewards across the funnel. This approach is commonly used in Reinforcement Learning with auxiliary rewards (Jaderberg et al., 2016; Lin et al., 2019), where the true reward is sparse and there are some non-sparse auxiliary rewards that can accelerate learning. We call this method *Mix* in the following experiments.

Inspired by the idea of *Mix* and that of curriculum learning Bengio et al. (2009), we also test the method that learns on the signals for each layer sequentially.

To sum up, we compare the following five strategies:

1. *Target*: we train a single model that only predicts the reward from the last stage, i.e. r_J.

2. *Mix*: we train a single model that only predicts the average rewards from all the stages, i.e. $\frac{1}{J}\sum_{j=1}^{J} r_j$)

3. *Sequential*: for a total steps T, we train a single model on rewards $r_1, \ldots, r_{J-1}$ sequentially for equal number of steps $\alpha T/(J-1)$ for some constant $\alpha \in [0,1]$ and train the model on the final reward r_J for the rest of steps.

4. *Multi-layer clustered*: Algorithm 3 under clustered dependency.

5. *Multi-layer sequential*: Algorithm 3 under sequential dependency.

Simulated Environments. We first tested the performance of multi-task learning algorithms on the contextual bandit setting. In our contextual bandit setting, number of action is set to be $A = 50$, for each action, we independently generate a funnel with $J = 8$ stages. The link function is from the logistic regression, where we sample the unknown parameter $\theta_{a,j}$ sequentially. In this case, $\theta_{a,1}$ is sampled from $N(0, \sigma^2)$ and $\theta_{a,j}$ is sampled from $N(\theta_{a,j-1}, \sigma^2/j)$ for for $j > 1$. This gives us a funnel with decreasing uncertainty. Context x is sampled from a Gaussian distribution $N(0, \sigma_x^2)$.

We set the parameters for the environment to be $A = 50; J = 8; \sigma = 1; \sigma_x = 0.08; d = 45; T = 3000$.

Model setup. For *Target*, *Mix* and *Sequential*, we use a Neural Network model with one hidden layer, d-dimensional input and A-dimensional output. Each dimension on the output vector represents the predicted conversion probability for an action. The number of units for the hidden layer is searched in $\{8, 16, 32, 64\}$. *Sequential* has the hyper-parameter a, which is searched in $\{0.1, 0.2, 0.4, 0.6\}$.

For *Multi-layer clustered* and *Multi-layer sequential*, each stage is modeled with the same one-layer Neural Networks defined above. The number of units for the hidden layer is searched in $\{1, 4, 8, 16\}$. The penalty parameter λ is searched in $\{0.001, 0.005, 0.01, 0.05\}$. An ϵ-greedy exploration is applied.
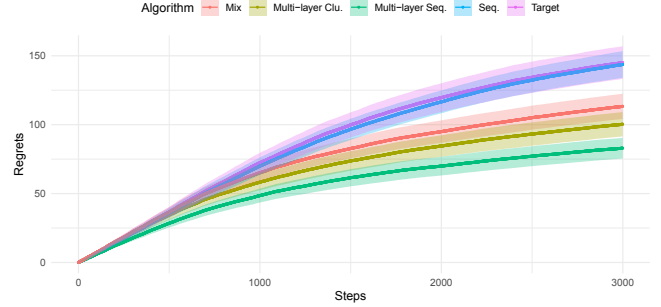

Figure 4: Cumulative regrets over 3000 steps using the best hyper-parameters for each of the five algorithms. The confidence interval is calculated from independent runs.

As shown in Figure 4, our practical algorithms beat all the other algorithms in terms of cumulative regrets. Algorithm 3 under sequential has lower regrets compared to that under clustered dependency.

5.3 Email campaign environment

We further test our algorithms on the Email Campaign problem, which aims at the act of sending a commercial message, typically to a group of people, using email.

Dataset. We randomly selected 5609706 users, who were active up to the data collection date and then tracked all the interactions of those users in the following 51 days, which adds up to 39488647 emails. Each email has a five-dimensional context, consisting of: *NumSent*, the number of emails sent to the user since 2019-12-01; *NumOpen*, the number of emails opened by the user since 2019-12-01; *NumClick*, the number of emails whose links were clicked by the user

since 2019-12-01; *BussinessGroup*, categorical variable indicating the business group of the user; and *Recency*, number of hours since last email was sent to the user, which is categorized into '0-12', '13-24', '25-36', '37-48', '49+'.

The agent takes actions of the time to send an email. The action space is divided into six blocks: 00:00-04:00, 04:00-08:00, 08:00-12:00, 12:00-16:00, 16:00-20:00, 20:00-24:00.

Three rewards are available for each email, indicating whether the email is opened, whether the email is clicked and whether the email leads to a purchase respectively. Note that we say an email leads a purchase if the user purchases in the following 30 days.

The email campaign problem defines a funnel with three layers. On average, the rates for opening, clicking and purchasing are about 10%, 0.4% and 0.01%, respectively in the dataset. As we can see, the signals for learning purchasing behaviour are sparse. We will show in our experiments that how multi-task learning can improve the sample efficiency.

Data-based environment. We first build a data-based contextual bandit environment using population distribution. At each step, the environment randomly samples a context from the dataset and the agent takes an action from the six blocks. The environment then samples a reward vector from the set of rewards with the same action and context.

	Purchase (10^{-2})	Click	Open
Target	4.09	3.88	24.6
Mix	4.23	6.63	47.0
Sequential	2.1	0.981	0.495
Multi-layer clu.	6.34	0.753	-18.1
Multi-layer seq.	1.06	2.04	3.03

Table 1: Average increases in the number of Purchase, Click or Open over 10000 steps compared to the Random policy using 20 independent runs. The standard deviations are all less than 10^{-3} for Purchase and 10^{-1} for Click and Open.

Results. We searched the same set of hyper-parameters as used in the previous experiments except for the number of hidden units. The hidden units for the two multi-layer algorithms are searched in $\{8, 16, 32, 64\}$ instead.

We first tested the decrease in prediction error for the five algorithms with actions randomly selected. Hyper-parameters with the lowest cumulative square prediction errors were selected. As shown in Figure 5, our

multi-task learning algorithms have much lower cumulative prediction errors. The total prediction errors converge after 2500 steps.
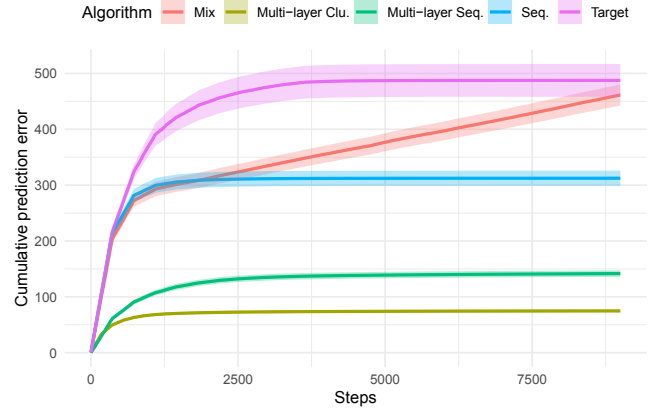

Figure 5: The cumulative square errors for five algorithms. The solid lines are averaged over 10 independent runs and regions mark the 1 standard deviation over the 10 runs.

We further applied our practical algorithm in Algorithm 3 and selected the hyper-parameters with the lowest regret. Table 1 showed the average increased number of Purchase, Click and Open within 10000 steps with respect to a purely random policy over 10 independent runs. The results are shown in Table 1. As one can see in the table, Multi-layer clustered improved Purchase rate the most by 0.0634 over 10000 steps. However, Mix improved the average number of Click or Open the most. This indicates that the three tasks are not exactly the same.

The advantages of our algorithms over the compared three algorithms are the use of full information and the appropriate regularization, which allows knowledge transfer between layers. *Target*, *Seq.* and *Mix* all use a single model for the whole funnel, where the similarities between layers are not clear and they all lose some information while processing signals. *Target* and *Seq.* learn only on a single signal from one layer on each step. *Mix* adopts a weighted average.

6   Discussion

In this paper, we formulated an important problem, funnel structure, from the marketing field. We used a multi-task learning algorithm to solve the contextual bandit problem with a funnel structure and offered its regret analysis. We verified our theorem using a simple simulation environment and tested the performances of our algorithm on both simulation and real-data environment.

Note that our bounds on prediction error in Theorem 1 and Theorem 2 do not scale with $1/\sqrt{\sum_j n_j}$ under the special case when $\theta_1 = \cdots = \theta_J$. However, the sparsity of the funnel implies that the optimal rate is only smaller than our bound by a constant factor that does not depend on J. To see this, assuming that $n_{j+1}/n_j = q$, we have $\sum_j n_j \to n/(1-q)$, which is a constant value.

In terms of the real-data environment, we adopted the population model that may lead to high variance in context and action pairs that do not have sufficient observations and our environment may not fully reflect the true environment. A better data-based environment may be proposed using rare event simulation Rubino and Tuffin (2009).

Acknowledgements

References

Barry, T. E. (1987). The development of the hierarchy of effects: An historical perspective. *Current issues and Research in Advertising*, 10(1-2):251–295.

Bengio, Y., Louradour, J., Collobert, R., and Weston, J. (2009). Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48.

Beygelzimer, A., Langford, J., Li, L., Reyzin, L., and Schapire, R. (2011). Contextual bandit algorithms with supervised learning guarantees. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 19–26.

Borrella, I., Caballero-Caballero, S., and Ponce-Cueto, E. (2019). Predict and intervene: Addressing the dropout problem in a mooc-based program. In *Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale*, pages 1–9.

Chu, W., Li, L., Reyzin, L., and Schapire, R. (2011). Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 208–214.

Clow, D. (2013). Moocs and the funnel of participation. In *Proceedings of the third international conference on learning analytics and knowledge*, pages 185–189.

Deshmukh, A. A., Dogan, U., and Scott, C. (2017). Multi-task learning for contextual bandits. In *Advances in neural information processing systems*, pages 4848–4856.

Drugan, M. M. and Nowe, A. (2013). Designing multi-objective multi-armed bandits algorithms: A study. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Evgeniou, T. and Pontil, M. (2004). Regularized multi–task learning. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 109–117.

Filippi, S., Cappe, O., Garivier, A., and Szepesvári, C. (2010). Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems*, pages 586–594.

Howard, J. Y. S. and Sheth, J. (1969). Jn (1969)" the theory of buyer behavior. *New York*.

Jaderberg, M., Mnih, V., Czarnecki, W. M., Schaul, T., Leibo, J. Z., Silver, D., and Kavukcuoglu, K. (2016). Reinforcement learning with unsupervised auxiliary tasks. *arXiv preprint arXiv:1611.05397*.

Jaksch, T., Ortner, R., and Auer, P. (2010). Near-optimal regret bounds for reinforcement learning. *Journal of Machine Learning Research*, 11(Apr):1563–1600.

Lazaric, A., Brunskill, E., et al. (2013). Sequential transfer in multi-armed bandit with finite set of models. In *Advances in Neural Information Processing Systems*, pages 2220–2228.

Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670.

Li, L., Chu, W., Langford, J., and Wang, X. (2011). Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 297–306.

Lin, X., Baweja, H., Kantor, G., and Held, D. (2019). Adaptive auxiliary task weighting for reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 4772–4783.

Manikrao, U. S. and Prabhakar, T. (2005). Dynamic selection of web services with recommendation system. In *International conference on next generation web services practices (NWESP'05)*, pages 5–pp. IEEE.

Maurer, A., Pontil, M., and Romera-Paredes, B. (2016). The benefit of multitask representation learning. *The Journal of Machine Learning Research*, 17(1):2853–2884.

Mulpuru, S. (2011). The purchase path of online buyers. *Forrester report*, 51:55.

Pentina, A., Sharmanska, V., and Lampert, C. H. (2015). Curriculum learning of multiple tasks. In

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5492–5500.

Pescher, C., Reichhart, P., and Spann, M. (2014). Consumer decision-making processes in mobile viral marketing campaigns. *Journal of interactive marketing*, 28(1):43–54.

Rubino, G. and Tuffin, B. (2009). *Rare event simulation using Monte Carlo methods*. John Wiley & Sons.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Schwartz, E. M., Bradlow, E. T., and Fader, P. S. (2017). Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, 36(4):500–522.

Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.

Soare, M., Alsharif, O., Lazaric, A., and Pineau, J. (2014). Multi-task linear bandits. In *NIPS2014 Workshop on Transfer and Multi-task Learning: Theory meets Practice*.

Srebro, N., Sridharan, K., and Tewari, A. (2010). Optimistic rates for learning with a smooth loss. *arXiv preprint arXiv:1009.3896*.

Turğay, E., Öner, D., and Tekin, C. (2018). Multi-objective contextual bandit problem with similarity information. *arXiv preprint arXiv:1803.04015*.

Wang, B., Mendez, J., Cai, M., and Eaton, E. (2019). Transfer learning via minimizing the performance gap between domains. In *Advances in Neural Information Processing Systems*, pages 10645–10655.

Zhang, Y. and Yang, Q. (2017). A survey on multi-task learning. *arXiv preprint arXiv:1707.08114*.

A   Missing Proofs

A.1   Connection to discrepancy measure

In this section, we discuss how our assumption relates to discrepancy assumptions. Consider $\mathcal{Y}$-discrepancy that measures the maximum absolute distance between the loss function: $\text{dist}(\mathcal{D}_1, \mathcal{D}_2) := \sup_{h \in \mathcal{H}} |\mathcal{L}_{\mathcal{D}_1}(h) - \mathcal{L}_{\mathcal{D}_2}(h)|$, where $\mathcal{D}_1$ and $\mathcal{D}_1$ represents the source domain and target domain and $\mathcal{L}_{\mathcal{D}_1}$ and $\mathcal{L}_{\mathcal{D}_2}$ are expected loss for two domains.

Note that under the GLM assumption, the L_2 distance in unknown parameters resembles the discrepancy using square loss. Consider a funnel with two layers and $\|\theta_1 - \theta_2\|_2 = q$. Lemma 3 indicates that $q \approx \text{dist}(\mathcal{D}_1, \mathcal{D}_2)$.

Lemma 3. *We have under square loss function,* $\text{dist}(\mathcal{D}_1, \mathcal{D}_2) \leq 4\kappa d_x q$.

Proof.

We first show the second inequality.

$$
\begin{aligned}
&\text{dist}(\mathcal{D}_1, \mathcal{D}_2) \\
&= \sup_\theta |\mathbb{E}_x(\mu(x^T\theta)) - \mu(x^T\theta_1^*))^2 - \mathbb{E}_x(\mu(x^T\theta) - \mu(x^T\theta_2^*))^2| \\
&\leq \sup_\theta |\mathbb{E}_x\mu(x^T\theta)(\mu(x^T\theta_1^*) - \mu(x^T\theta_2^*))| + |\mathbb{E}_x(\mu^2(x^T\theta_1^*) - \mu^2(x^T\theta_2^*))| \\
&\leq 4|\mathbb{E}_x(\mu(x^T\theta_1^*) - \mu(x^T\theta_2^*))| \\
&\leq 4\mathbb{E}_x|\mu(x^T\theta_1^*) - \mu(x^T\theta_2^*)| \\
&\leq 4\kappa\mathbb{E}_x|x^T(\theta_1^* - \theta_2^*)| \\
&\leq 4\kappa d_x q
\end{aligned}
$$

On the other hand, an lower bound of $\text{dist}(\mathcal{D}_1, \mathcal{D}_2)$ is also closely related to q.

$$
\begin{aligned}
&\text{dist}(\mathcal{D}_1, \mathcal{D}_2) \\
&= \sup_\theta |\mathbb{E}_x(\mu(x^T\theta)) - \mu(x^T\theta_1^*))^2 - \mathbb{E}_x(\mu(x^T\theta) - \mu(x^T\theta_2^*))^2| \\
&= \sup_\theta |\mathbb{E}_x(\mu(x^T\theta_1^*) - \mu(x^T\theta_2^*))(\mu(x^T\theta_1^*) + \mu(x^T\theta_2^*) + \mu(x^T\theta))| \\
&= \sup_\theta |\mathbb{E}_x \int_t \mu'(tx^T\theta_1^* + (1-t)x^T\theta_2^*) \, dt(x^T(\theta_1^* - \theta_2^*))(\mu(x^T\theta_1^*) + \mu(x^T\theta_2^*) + \mu(x^T\theta))| \\
&= \sup_\theta |(\theta_1^* - \theta_2^*)^T[\mathbb{E}_x x \int_t \mu'(tx^T\theta_1^* + (1-t)x^T\theta_2^*) \, dt(\mu(x^T\theta_1^*) + \mu(x^T\theta_2^*) + \mu(x^T\theta))]| \\
&\quad (\text{Let } \theta \to -\infty) \\
&\geq |(\theta_1^* - \theta_2^*)^T \nu_{\theta_1^*, \theta_2^*}| \ (\text{letting } \nu_{\theta_1^*, \theta_2^*} = [\mathbb{E}_x x \int_t \mu'(tx^T\theta_1^* + (1-t)x^T\theta_2^*) \, dt(\mu(x^T\theta_1^*) + \mu(x^T\theta_2^*))]).
\end{aligned}
$$

Let $\theta_2^* = \theta_1^* + \|\theta_1^* - \theta_2^*\|_2 \mu$, where μ is a unit vector. For sufficient small $\|\theta_1^* - \theta_2^*\|_2, \nu_{\theta_1^*, \theta_2^*} \to 2\mathbb{E}_x[x\mu'(x^T\theta_1^*)\mu(x^T\theta_1^*)] =: \nu_{\theta_1^*}$, which is a constant vector. Thus

$$
\lim_{\|\theta_1^* - \theta_2^*\|_2 \to 0} \frac{\text{dist}(\mathcal{D}_1, \mathcal{D}_2)}{\|\theta_1^* - \theta_2^*\|_2} = |\mu^T \nu_{\theta_1^*}|.
$$

For sufficient small $\|\theta_1^* - \theta_2^*\|$, discrepancy scales with $\|\theta_1^* - \theta_2^*\|$.

A.2   Proof of Lemma 1

In this subsection, we introduce the proof of Lemma 1. Many proofs could achieve a very similar bound. Here we use the idea of local Rademacher complexity.

Proof.

We discuss two cases: 1) $\hat{\theta} \in \text{int}(\Theta_0)$. 2) $\hat{\theta} \notin \text{int}(\Theta_0)$.

In both cases, one simply has

$$|\mu(x^T\hat{\theta}) - \mu(x^T\theta^*)| \leq \kappa|x^T(\hat{\theta} - \theta^*)| \leq \kappa \sup_{\theta_1,\theta_2\in\Theta_0} |x^T(\theta_1 - \theta_2)|,$$

which completes the first term in the minimum.

Now we prove the parametric bound. We first assume that case 1 holds. In this case, the constraint does not come into effects and $\hat{\theta}$ is the global minimal. By Theorem 26.5 in Shalev-Shwartz and Ben-David (2014), we have under an event, whose probability is at least $1 - \delta$,

$$L(\hat{\theta}) - L(\theta^*) \leq 2R_n(\boldsymbol{z}) + 5\sqrt{\frac{2\ln(8/\delta)}{n}}, \tag{8}$$

where $R(\boldsymbol{z})$ is the Rademacher complexity defined by

$$R_n(\boldsymbol{z}) = \mathbb{E}_{\boldsymbol{\sigma}} \frac{1}{n} \sup_{\theta\in\Theta} \sum_{i=1}^{n} \|z_i - \mu(x_i^T\theta)\|_{M_n}^2 \sigma_i,$$

and the variables in $\boldsymbol{\sigma}$ are distributed i.i.d. from Rademacher distribution. Let us call the event E_A.

As for any $i \in [n]$, let $\phi_i(t) \coloneqq (z_i - \mu(t))^2$, which satisfies $|\phi_i'(t)| = |2(z_i - \mu(t))\mu'(t)| \leq \kappa$, using Contraction lemma (Shalev-Shwartz and Ben-David, 2014), we have

$$\begin{aligned}
R_n(\boldsymbol{z}) &\leq \mathbb{E}_{\boldsymbol{\sigma}} \frac{1}{n} \sup_{\theta\in\Theta} \sum_i \kappa(x_i^T\theta)\sigma_i \\
&= \kappa\mathbb{E}_{\boldsymbol{\sigma}} \frac{1}{n} \sup_{\theta\in\Theta} \sum_{i=1}^{n} x_i^T(\theta - \theta^*)\sigma_i. \\
&\leq \kappa\mathbb{E}_{\boldsymbol{\sigma}} \frac{1}{n} \sup_{\theta\in\Theta} \|\sum_i x_i\sigma_i\|_{M_n^{-1}} \|\theta - \theta^*\|_{M_n} \\
&\leq \kappa\mathbb{E}_{\boldsymbol{\sigma}} \frac{1}{n} \|\sum_i x_i\sigma_i\|_{M_n^{-1}} \sup_{\theta\in\Theta_0} \|\theta - \theta^*\|_{M_n}.
\end{aligned} \tag{9}$$

Next, using Jensen's inequality we have that

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\sigma}} \frac{1}{n} \|\sum_i x_i\sigma_i\|_{M_n^{-1}} \\
&\leq \frac{1}{n} \left(\mathbb{E}_{\boldsymbol{\sigma}} \|\sum_i x_i\sigma_i\|_{M_n^{-1}}^2 \right)^{1/2} \\
&= \frac{1}{n} \left(\mathbb{E}_{\boldsymbol{\sigma}} tr[M_n^{-1}(\sum_i x_i\sigma_i)(\sum_i x_i\sigma_i)^T] \right)^{1/2} \\
&= \frac{1}{n} \left(tr[M_n^{-1}\mathbb{E}_{\boldsymbol{\sigma}}(\sum_i x_i\sigma_i)(\sum_i x_i\sigma_i)^T] \right)^{1/2}
\end{aligned} \tag{10}$$

Finally, since the variables $\sigma_1, \ldots, \sigma_m$ are independent we have

$$\mathbb{E}_{\boldsymbol{\sigma}}(\sum_i x_i \sigma_i)(\sum_i x_i \sigma_i)^T$$

$$= \mathbb{E}_{\boldsymbol{\sigma}} \sum_{k,l \in [n]} \sigma_k \sigma_l x_k x_l^T$$

$$= \mathbb{E}_{\boldsymbol{\sigma}} \sum_{i \in [n]} \sigma_i^2 x_i x_i^T$$

$$= \sum_{i \in [n]} x_i x_i^T = n M_n.$$

Plugging this into (10), assuming M_n is full rank, we have

$$(9) \leq \sqrt{d/n} \sup_{\theta \in \Theta_0} \|\theta - \theta^*\|_{M_n}. \tag{11}$$

Lemma 4. *Under the notation in Lemma 1 and Assumption 2, if an estimate $\hat{\theta}$ satisfies $L(\hat{\theta}) \leq L(\theta^*) + b_n$, then*

$$\|\hat{\theta} - \theta^*\|_{M_n}^2 \leq \frac{d_x b_n}{c_\mu}.$$

Proof. Let $g_n(\theta) = \sum_i x_i (\mu(x_i^T \theta) - \mu(x_i^T \theta^*))$. For any θ, $\nabla g_n(\theta) = \sum_i x_i x_i^T \mu'(x_i^T \theta)$. By simple calculus,

$$g_n(\theta^*) - g_n(\hat{\theta}) = \int_0^1 \nabla g_n \left(s\theta^* + (1-s)\hat{\theta} \right) ds (\theta^* - \hat{\theta}).$$

As $\mu(t) \geq c_\mu$, we have $\int_0^1 \nabla g_n \left(s\theta^* + (1-s)\hat{\theta} \right) ds \succ c_\mu M_n$. Plugging this into the inequality above we have

$$\|\theta^* - \hat{\theta}\|_{M_n}^2 \leq \frac{1}{c_\mu} (\sum_i x_i (\mu(x_i^T \theta) - \mu(x_i^T \theta^*)))^2 = \frac{1}{c_\mu} \epsilon^T M_n \epsilon \leq \frac{d_x}{c_\mu} \epsilon^T \epsilon = \frac{d_x}{c_\mu} (L(\hat{\theta}) - L(\theta^*)),$$

where $\epsilon := (\mu(x_i^T \hat{\theta}) - \mu(x_i^T \theta^*))_{i=1}^n$.

Applying (8) and Lemma 4, we complete the proof by

$$\|\hat{\theta} - \theta^*\|_{M_n} \leq \sqrt{\frac{2 d_x \sqrt{d}}{c_\mu \sqrt{n}} \sup_{\theta \in \Theta} \|\theta - \theta^*\|_{M_n} + 5\sqrt{\frac{2 \ln(8/\delta)}{n}}}$$

$$\leq \sqrt{\frac{20 \sqrt{2 \ln(8/\delta)} d_x \sqrt{d} \sup_{\theta \in \Theta_0} \|\theta - \theta^*\|_{M_n}}{c_\mu \sqrt{n}}}. \tag{12}$$

$\square$

We apply (12) iteratively [1]. Let $\Theta_{(1)} := \Theta_0$. For any $t > 1$, let $\Theta_{(t)} = \{\theta \in \mathbb{R}^d : \|\theta - \hat{\theta}\|_{M_n} \leq \sqrt{\frac{20 \sqrt{2d \ln(8/\delta)}}{c_\mu \sqrt{n}}} \sup_{\theta \in \Theta_{(t-1)}} \|\theta - \theta^*\|_{M_n}\}$. When $t \to \infty$, we have

$$\Theta_{(\infty)} = \frac{20 d_x \sqrt{2d \ln(8/\delta)}}{c_\mu \sqrt{n}}.$$

By (12), we have $\theta^* \in \cap_{t \geq 1} \Theta_{(\infty)}$ and $\|\hat{\theta} - \theta^*\|_{M_n} \leq \frac{40 d_x \sqrt{2d \ln(8/\delta)}}{c_\mu \sqrt{n}}$, which completes the second part of Lemma 1.

---

[1]Note that (12) holds under the same event E_A as the estimates $\hat{\theta}$ keeps the same each round as it is the global minimizer.

For any $x \in \mathcal{X}$, we have

$$|\mu(x^T\hat{\theta}) - \mu(x^T\theta^*)| \leq \kappa\|x\|_{M_n^{-1}} \frac{40d_x\sqrt{2d\ln(8/\delta)}}{c_\mu\sqrt{n}}. \tag{13}$$

When case 2 holds, let $\hat{\theta}'$ be the global minimizer. Using the analysis above, we have

$$\|\hat{\theta}' - \theta^*\|_{M_n} \leq \frac{40d_x\sqrt{2d\ln(8/\delta)}}{c_\mu\sqrt{n}}.$$

Then by triangle inequality

$$\|\hat{\theta} - \theta^*\|_{M_n} \leq \|\hat{\theta} - \hat{\theta}'\|_{M_n} + \|\hat{\theta}' - \theta^*\|_{M_n} \leq \frac{80d_x\sqrt{2d\ln(8/\delta)}}{c_\mu\sqrt{n}}.$$

A.3 Tightness of Lemma 1

We use an example to show the tightness of Lemma 1. Assume a linear predictor, i.e. $\mu(t) = t$. Consider the following distribution, let X be uniform over the d-standard basis vector e_m, for $m = 1, \ldots, d$. Let $Z \mid (X = e_i) \sim Bern(r_i)$, where $r_i \in [0,1]$ is pre-determined and unknown. The optimal parameter $\theta^* = (r_1, \ldots, r_d)^T$. Let n_m be the number of samples collected for dimension m. Let $\Theta_0 := \{\theta : \|\theta\|_2 \leq q\}$.

When n is sufficiently large $n > 1/q^2$, $\hat{\theta}$ is the regularized minimizer. It can be shown that for any $\hat{\theta}$, there exists θ^* such that $\mathbb{E}[\hat{\theta}_i - \theta_i^*]^2 \geq (r_m(1 - r_m))/n_m$. Then $\mathbb{E}\|\hat{\theta} - \theta^*\|_2^2 \geq \sum_{m=1}^d \frac{r_i(1-r_i)}{n_m} \geq \frac{d^2(r_i(1-r_i))}{n} = \Omega(\frac{d^2}{n})$.

Then we also see that when n is small ($\leq \frac{1}{q^2}$), the estimation error is $\Omega(q)$. We use the same example as above. This time, we assume $\|\theta^*\| \leq \frac{q}{2}$. If we have a $\|\hat{\theta}\| = q$, then $\|\theta^* - \hat{\theta}\| \geq q/2 = \Omega(q)$. Otherwise, we use the lower bound above: $\|\theta^* - \hat{\theta}\| \geq \Omega(\frac{d}{\sqrt{n}}) = \Omega(dq)$.

The above argument corresponds to the upper bound in Lemma 1, where we use prior knowledge when n is small and use the parametric bound when n is large.

A.4 Proof of Theorem 1

In this subsection, we show the missing proof for Theorem 1.

Theorem 4 (Prediction error under sequential dependency). *For any funnel with a sequential dependency of parameters $q_1, \ldots, q_J$, let $\hat{\theta}_1, \ldots, \hat{\theta}_J$ be the estimates from Algorithm 1. If $n_{j+1} \leq n_j/4$, $q_1 \geq, \ldots, \geq q_J$ and Assumption 5 is satisfied, then with a probability at least $1 - \delta$, for any $j_0 \in [J]$, we have*

$$PE_j \leq \begin{cases} \kappa\|x\|_2 \frac{c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n_j}}, & \text{if } j < j_0, \\ \kappa\|x\|_2 (\frac{c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n_{j_0}}} + \sum_{i=j_0+1}^j q_j), & \text{if } j \geq j_0, \end{cases} \tag{14}$$

where we let $n_0 = \infty$. The bound is smallest when j_0 is the smallest $j \in [J]$, such that

$$\frac{4c_\delta\sqrt{d}}{c_\mu\lambda}\left(\frac{1}{\sqrt{n_j}} - \frac{1}{\sqrt{n_{j-1}}}\right) \geq q_j, \tag{15}$$

if none of j's in $[J]$ satisfies (15), $j_0 = J + 1$.

Proof. First we reshape the ellipsoid in (3) to a ball.

Lemma 5 (Reshape). *For any vector $x \in \mathbb{R}^d$ and any matrix $M \succ 0 \in \mathbb{R}^{d \times d}$, $\|x\|_2 \leq \frac{1}{\lambda}\|x\|_M$, where λ is the minimum eigenvalue of M.*

Proof. We directly use the definition of positive definite matrix: $\lambda^2\|x\|_2^2 - \|x\|_M^2 = x^T(\lambda^2 I - M)x \leq 0$. Thus, $\|x\|_2 \leq \frac{1}{\lambda^2}\|x\|_M$. #

Using Lemma 5 and Assumption 5, we have $\|\bar{\theta}_j - \theta_j^*\|_2 \leq \frac{1}{\lambda}\|\bar{\theta}_j - \theta_j^*\|_{M_n} \leq \frac{4c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n}}$. Thus the set $\hat{\Theta}_j \subset \{\theta : \|\theta - \bar{\theta}_j\|_2 \leq \frac{4c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n}}\} =: \hat{\Theta}_j^{ball}$.

For every j, one can derive two bounds. First we can directly apply Corollary 1 and get $PE_j \leq \kappa\|x\|_2 \frac{4c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n_j}}$. Second, for any j_0, we have $\theta_j^* \in \Theta_1[j] \subset \{\theta : \|\bar{\theta}_{j_0} - \theta\|_2 \leq \frac{4c_s}{c_\mu\lambda}\sqrt{\frac{d}{n_{j_0}}} + \sum_{j_0+1\leq i\leq j} q_i\}$ and get $PE_j \leq \kappa\|x\|_2(\frac{c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n_{j_0}}} + \sum_{i=j_0+1}^{j} q_j)$.

Now we show the second argument: of all those bounds the one defined in (14) with j_0 defined in (15) is the smallest. For any $j \leq j_0$ and $j_1 \leq j$, we have

$$\frac{4c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n_j}} = \frac{4c_\delta\sqrt{d}}{c_\mu\lambda}\left(\sum_{i=j_1+1}^{j}(\frac{1}{\sqrt{n_i}} - \frac{1}{\sqrt{n_{i-1}}}) + \frac{1}{\sqrt{n_{j_1}}}\right) \leq \frac{4c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n_{j_1}}} + \sum_{i=j_1+1}^{j} q_i. \tag{16}$$

The second inequality is given by $(\frac{1}{\sqrt{n_i}} - \frac{1}{\sqrt{n_{i-1}}}) \leq q_i$ for all $i < j_0$. For any $j \geq j_0$ and $j_1 \leq j_0$, by (16), we have

$$\frac{4c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n_{j_0}}} + \sum_{i=j_0+1}^{j} q_i \leq \frac{4c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n_{j_1}}} + \sum_{i=j_1+1}^{j} q_i.$$

Now we prove that for all $i \geq j_0$,

$$\frac{4c_\delta\sqrt{d}}{c_\mu\lambda}(\frac{1}{\sqrt{n_i}} - \frac{1}{\sqrt{n_{i-1}}}) \geq q_i. \tag{17}$$

We use induction. Assume for some i_1, (17) is satisfied. Under the assumption that $n_{i_1-1} \leq n_{i_1}/4$ and $q_{i_1} \geq q_{i_1+1}$, we have

$$\frac{4c_\delta\sqrt{d}}{c_\mu\lambda}(\frac{1}{\sqrt{n_{i_1+1}}} - \frac{1}{\sqrt{n_{i_1}}}) = \frac{4c_\delta\sqrt{d}}{c_\mu\lambda}(\frac{1}{\sqrt{n_{i_1+1}}} + \frac{1}{\sqrt{n_{i_1-1}}} - \frac{2}{\sqrt{n_{i_1}}} + \frac{1}{\sqrt{n_{i_1}}} - \frac{1}{\sqrt{n_{i_1-1}}})$$

$$\geq \frac{4c_\delta\sqrt{d}}{c_\mu\lambda}(\frac{1}{\sqrt{n_{i_1+1}}} + \frac{2}{\sqrt{n_{i_1}}} - \frac{2}{\sqrt{n_{i_1}}} + \frac{1}{\sqrt{n_{i_1}}} - \frac{1}{\sqrt{n_{i_1-1}}})$$

$$\geq \frac{4c_\delta\sqrt{d}}{c_\mu\lambda}(+\frac{1}{\sqrt{n_{i_1}}} - \frac{1}{\sqrt{n_{i_1-1}}})$$

$$\geq q_{i_1} \geq q_{i_1+1}.$$

Using 17, for any $j \geq j_1 > j_0$,

$$\frac{4c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n_{j_0}}} + \sum_{i=j_0+1}^{j} q_i = \frac{4c_\delta\sqrt{d}}{c_\mu\lambda}\left(\sum_{i=j_0+1}^{j_1}(\frac{1}{\sqrt{n_{i-1}}} - \frac{1}{\sqrt{n_i}}) + \frac{1}{\sqrt{n_{j_1}}}\right) + \sum_{i=j_0+1}^{j} q_i \leq \frac{4c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n_{j_1}}} + \sum_{i=j_1+1}^{j} q_i.$$

Finally, we conclude that j_0 gives the smallest bound. # $\qquad\square$

Similar argument can be used to show Theorem 2. For any $j_0 \in [J]$, we have

$$PE_j \leq \kappa\|x\|_2 \min\left\{\frac{c_{\delta/J}}{c_\mu\lambda}\sqrt{\frac{d}{n_{j0}}} + q_j, \frac{c_{\delta/J}}{c_\mu\lambda}\sqrt{\frac{d}{n_j}}\right\}.$$

Out of all the choices of j_0, the best one is achieved by $j_0 = \arg\min_{j\in[J]} \frac{c_\delta}{c_\mu\lambda}\sqrt{\frac{d}{n_j}} + q_j$.

A.5   Proof of Theorem 3

Theorem 5. *Using Algorithm 2, under the Assumptions 1-4, with a probability at least $1 - \delta$, the total regret*

$$\sum_{t=1}^{T} \left[P(x_t, \boldsymbol{\theta}^*_{a^*_t}) - P(x_t, \boldsymbol{\theta}^*_{a_t}) \right]$$

$$\leq 2\sqrt{2}c_0 \sum_{a,j} \sqrt{n^T_{a,j}} + \sum_{a,j} \frac{8c_0^2 J d_x^4 \log(6AJT/\delta)}{\bar{p}_{a,j}^2} - \sum_{a,j} \Delta_{a,j}. \tag{18}$$

where $\mathcal{O}$ ignores all the constant terms and logarithmic terms for better demonstrations, $c_0 = (\kappa d_x c_{\delta/AJT}\sqrt{d})/(c_\mu \bar{\lambda})$, $\bar{p}_a := \mathbb{E}_x P_{J-1}(x^T \boldsymbol{\theta}^*_a)$ and

$$\Delta_{a,j} = \sum_{t=1;a_t=a}^{T} P_j(x_t^T \hat{\theta}_{a_t}^t) \left[c_0 \frac{1}{\sqrt{n_{a,j}^t \vee 1}} - \Delta\mu_{a,j}^t \right].$$

represents the benefits of transfer learning.

Let $\bar{p}_{a,j} := \mathbb{E}_x P_{j-1}\left(x^T \theta^*_a\right)$. We first show that upper bound the number of steps t with $\lambda^t_{a_t,j} \leq \bar{\lambda}/2$ or $n^t_{a,j} \leq \frac{1}{2}n^t_{a,1}\bar{p}_{a,j}$. These steps are considered bad events.

Lemma 6 shows that with high probability, the number of observations for each layer is close to its expectation.

Lemma 6. *With a probability at least $1 - \delta$, we have $n^t_{a,j} \geq n^t_{a,1}\bar{p}_{a,j} - \sqrt{2n^t_{a,1}\log(1/\delta)}$. Especially, when $n^t_{a,1} > 8\log(1/\delta)/\bar{p}^2_{a,j} =: c_{n,a}$, we have $n^t_{a,j} \geq \frac{1}{2}n^t_{a,1}\bar{p}_{a,j}$.*

Proof. This is a direct application of Hoeffding inequality. ☐

Lemma 7. *For any $x_1, \ldots, x_n$ i.i.d, $\|x_i\| \leq d_x$, let λ_n be the minimum eigenvalue of $\sum_i x_i x_i^T/n$ and $\bar{\lambda}$ be the minimum eigenvalue of its expectation. We have $\lambda_n \geq \bar{\lambda}/2$, when $n > d_x^4 \log(1/\delta)/\bar{\lambda}^2$.*

Proof. For all $x_1, \ldots, x_n$, write $x_i = \sum_{s=1}^{d} \nu_{s,i}\tilde{x}_s$, where $\tilde{x}_1, \ldots, \tilde{x}_d$ are any basis of $\mathbb{R}^d$. We have $\mathbb{E}\nu^2_{s,i} \geq \bar{\lambda}$. For Hoeffding's inequality, since $\nu_{s,i} \leq d_x$, with a probability $1 - \delta$, we have

$$\frac{1}{n}\sum_i \nu^2_{s,i} \geq \mathbb{E}\nu^2_{s,1} - d_x^2\sqrt{\frac{\log(1/\delta)}{n}} \geq \bar{\lambda} - d_x^2\sqrt{\frac{\log(1/\delta)}{n}}.$$

For $n > d_x^4 \log(1/\delta)/\bar{\lambda}^2$, we have $\frac{1}{n}\sum_i \nu^2_{s,i} \geq \bar{\lambda}/2$. There exists a choice of $\tilde{x}_1, \ldots, \tilde{x}_d$ such that $\lambda_n = \frac{1}{n}\sum_i \nu^2_{s,i}$. ☐

Combining Lemma 6 and Lemma 7, we have with a probability at least $1 - \delta/3$, $\#\{t : \exists j, \lambda^t_{a_t,j} \leq \bar{\lambda}/2 \text{ or } n^t_{a,j} \leq \frac{1}{2}n^t_{a,1}\bar{p}_{a,j}\}$ can be upper bounded by

$$\sum_{a,j} \max\left\{8\log(6AJT/\delta)/\bar{p}^2_{a,j}, 2d_x^4 \log(6AJT/\delta)/(\bar{\lambda}^2\bar{p}_{a,j})\right\}. \tag{19}$$

In the following proof, we assume for all t, $\lambda^t_{a,j} \geq \bar{\lambda}/2$ and $n^t_{a,j} \geq \frac{1}{2}n^t_{a,1}\bar{p}_{a,j}$. We also assume the event in Lemma 1 happens for all $a \in [A], j \in [J]$ and $t < T$. The probability is at least $1 - \delta/3$ as each probability is at least $1 - \delta/(3AJT)$.

The total regret is

$$\sum_{t=1}^{T} \left[P(x_t, \boldsymbol{\theta}^*_{a^*_t}) - P(x_t, \boldsymbol{\theta}^*_{a_t}) \right]$$

$$\leq \sum_{t=1}^{T} \left[P(x_t, \boldsymbol{\theta}^*_{a^*_t}) - P^+(x_t, \hat{\boldsymbol{\theta}}_{a_t}) + P^+(x_t, \hat{\boldsymbol{\theta}}_{a_t}) - P(x_t, \boldsymbol{\theta}^*_{a_t}) \right]$$

(Using $P(x_t, \boldsymbol{\theta}^*_{a^*_t}) - P^+(x_t, \hat{\boldsymbol{\theta}}_{a_t}) \leq 0$)

$$\leq \sum_{t=1}^{T} \left[P^+(x_t, \hat{\boldsymbol{\theta}}^t_{a_t}) - P(x_t, \boldsymbol{\theta}^*_{a_t}) \right]$$

(Using Lemma 2)

$$\leq \sum_{t=1}^{T} \left[\sum_j \frac{P_J\left(x, \hat{\theta}^t_{a_t}\right)}{\mu\left(x^T \hat{\theta}^t_{a_t,j}\right)} \Delta\mu^t_{a_t,j} + \sum_{i \neq j} \Delta\mu^t_{a_t,j} \Delta\mu^t_{a_t,i} \right]$$

$$\leq \sum_{t=1}^{T} \left[\sum_j P_j(x_t, \hat{\theta}^t_{a_t}) \Delta\mu^t_{a_t,j} + \sum_{i \neq j} \Delta\mu^t_{a_t,j} \Delta\mu^t_{a_t,i} \right]$$

$$= \sum_{t=1}^{T} \left[\sum_j (P_j(x_t, \theta^*_{a_t}) + P_j(x_t, \hat{\theta}^t_{a_t}) - P_j(x_t, \theta^*_{a_t})) \Delta\mu^t_{a_t,j} + \sum_{i \neq j} \Delta\mu^t_{a_t,j} \Delta\mu^t_{a_t,i} \right]$$

$$\leq \underbrace{\sum_{t=1}^{T} \sum_j P_j\left(x_t, \theta^*_{a_t}\right) \frac{c_0}{\sqrt{n^t_{a_t,j}}}}_{\text{①}} - \underbrace{\sum_{t=1}^{T} \sum_j P_j\left(x_t, \theta^*_{a_t}\right) \left(\frac{c_0}{\sqrt{n^t_{a_t,j}}} - \Delta\mu^t_{a_t,i}\right)}_{\text{②}} + \underbrace{\sum_{t=1}^{T} \sum_{i \neq j} \Delta\mu^t_{a_t,j} \Delta\mu^t_{a_t,i}}_{\text{③}}$$

$$+ \underbrace{\sum_{t=1}^{T} \left[\sum_j (P_j(x_t, \hat{\theta}^t_{a_t}) - P_j(x_t, \theta^*_{a_t})) \Delta\mu^t_{a_t,j} \right]}_{\text{④}}.$$

We further bound the terms separately. The first term ① represents the bound one could have without multi-task learning.

$$\sum_{t=1}^{T} \sum_j P_j(x_t, \theta^*_{a_t}) \frac{c_0}{\sqrt{n^t_{a_t,j}}}$$

$$\leq \sum_{t=1}^{T} \sum_j \mathbf{1}(r_{t,j-1} = 1) \frac{c_0}{\sqrt{n^t_{a_t,j}}} + \sum_{t=1}^{T} \sum_j (P_j(x_t, \theta^*_{a_t}) - \mathbf{1}(r_{t,j-1} = 1)) \frac{c_0}{\sqrt{n^t_{a_t,j}}}$$

(Using Lemma 19 in Jaksch et al. (2010))

$$\leq c_0 2\sqrt{2} \sum_{a,j} \sqrt{n^T_{a,j}} + \sum_{t=1}^{T} \sum_j (P_j(x_t, \theta^*_{a_t}) - \mathbf{1}(r_{t,j-1} = 1)) \frac{c_0}{\sqrt{n^t_{a_t,j}}} \tag{20}$$

As $\mathbb{E}[P_j(x_t, \theta^*_{a_t}) - \mathbf{1}(r_{t,j-1} = 1)] = 0$, the second term in (20) is a martingale. Using Azuma-Hoeffding inequality, with a probability at least $1 - \delta/3$, for all T,

$$\sum_{t=1}^{T} \sum_j (P_j(x_t, \theta^*_{a_t}) - \mathbf{1}(r_{t,j-1} = 1)) \frac{c_0}{\sqrt{n^t_{a_t,j}}} \leq c_0 \sqrt{2 \log(3TJ/\delta)}. \tag{21}$$

Combined with (20),

$$① \leq 2\sqrt{2}c_0 \sum_{a,j} \sqrt{n_{a,j}^T} + c_0\sqrt{2\log(3TJ/\delta)}. \tag{22}$$

Next we bound ③. We notice that this is a quadratic term. We first show Lemma 6 that lower bounds the number of observations for each layer. Lemma 6 is a direct application of Hoeffding's inequality.

For any pair i, j, we have

$$\sum_{t=1}^{T} \Delta\mu_{a_t,j}^t \Delta\mu_{a_t,i}^t$$

$$\leq c_0^2 \sum_{t=1}^{T} \frac{1}{\sqrt{n_{a_t,i}^t}} \frac{1}{\sqrt{n_{a_t,j}^t}}$$

$$\leq c_0^2 \sum_{t=1}^{T} \left[\mathbf{1}(n_{a_t,1}^t \leq c_{n,a_t}) \frac{1}{\sqrt{n_{a_t,i}^t}} \frac{1}{\sqrt{n_{a_t,j}^t}} + \mathbf{1}(n_{a_t,1}^t > c_{n,a_t}) \frac{1}{\sqrt{n_{a_t,i}^t}} \frac{1}{\sqrt{n_{a_t,j}^t}} \right]$$

$$\leq c_0^2 \sum_{a} c_{n,a} + c_0^2 \sum_{t} \frac{4}{\bar{p}_a^2 n_{a_t,1}^t}$$

$$\leq c_0^2 \sum_{a} c_{n,a} + c_0^2 \sum_{a} \frac{4\log(n_{a,1}^T)}{\bar{p}_a^2}$$

$$\leq 4c_0^2 \sum_{a} \frac{\log(n_{a,1}^T A/\delta)}{\bar{p}_a^2}. \tag{23}$$

where we let $\bar{p}_a := \mathbb{E}_x P_J \left(x^T \theta_a^* \right)$.

Thus, ③ is upper bounded by $4c_0^2 J^2 \sum_a \frac{\log(n_{a,1}^T A/(3\delta))}{\bar{p}_a^2}$.

Finally we bound term ④. Using Lemma 2 on only first j layers, we have

$$④ \leq \sum_t \sum_j [\sum_i \Delta\mu_{a_t,i}^t + \sum_{i,k} \Delta\mu_{a_t,k}^t \Delta\mu_{a_t,i}^t]\Delta\mu_{a_t,j}^t \leq (J+1) \times ③. \tag{24}$$

The proof is completed by combining Equations (19), (21), (22), (23) and (24).

B   Experiments

B.1   Practical algorithm

Algorithm 3 Practical Algorithm for Contextual Bandit with a Funnel Structure

---

$t \to 1$, total number of steps T, memory $\mathcal{H}_a = \{\}$ for all $a \in [A]$. Initialize $\hat{\theta}_{a,\star}$ with zero vectors.
$\hat{\theta}_{a,0} \to 0$.
for $t = 1$ to T **do**
    Receive context x_t.
    Choose $a_t = \arg\max_{a \in \mathcal{A}} \hat{P}_J(x_t, \hat{\theta}_{a,j})$.
    Set $a_t = \text{Unif}([A])$ with probability ϵ.
    Receive $r_{t,1}, \ldots, r_{t,J}$ from funnel F_{a_t}.
    Set $\mathcal{H}_{a_t} \to \mathcal{H}_{a_t} \cup \{(x_t, (r_{t,1}, \ldots, r_{t,J}))\}$.
    for $j = 1, \ldots, J$ **do**
      # For sequential dependency

$$\hat{\theta}_{a_t,j} \to \arg\min_{\theta} l(\theta, \mathcal{H}_{a_t}) + \lambda_j \|\theta - \hat{\theta}_{a_t,j-1}\|_2$$

      # For clustered dependency

$$\hat{\theta}_{a_t,j} \to \arg\min_{\theta} l(\theta, \mathcal{H}_{a_t}) + \lambda_j \|\theta - \frac{1}{J} \sum_i \hat{\theta}_{a_t,i}\|_2$$

    end for
end for

B.2   Tuned hyper-parameters

Simulated environment.

1. Target: units 16

2. Mix: units 32

3. Sequential: units 32

4. Multi-layer Clustered: units 4; λ 0.001

5. Multi-layer Sequential: units 8; λ 0.001

Data-based environment.

1. Target: units 64

2. Mix: units 64

3. Sequential: units 64

4. Multi-layer Clustered: units 64; λ 0.005

5. Multi-layer Sequential: units 16; λ 0.001