TRANSFER LEARNING WITH NEURAL UCB

Anonymous authors
Paper under double-blind review

ABSTRACT

We study transfer learning in neural contextual bandits, where labeled data from a source task are available before interaction with a target task whose reward function may differ. We propose Transfer-NeuralUCB, which uses a predictor trained on the source task as the center for target adaptation and transfers source uncertainty through a design-weighted certificate for reward mismatch. The algorithm combines a mismatch-adjusted source UCB with an online UCB learned from target feedback and selects actions using their pointwise minimum. Given an RKHS transfer certificate and finite-width neural tangent kernel conditions, we establish simultaneous confidence guarantees along the source and target training iterates. The resulting regret bound is controlled by the smaller of the cumulative source and online confidence widths. Our analysis identifies residual tangent complexity as the regularization cost of adapting the source predictor and characterizes, through inverse-kernel alignment and scale, when the source prediction is simpler to correct than the target reward is to learn from scratch.Experiments on synthetic and classification-derived bandits show that Transfer-NeuralUCB improves on NeuralUCB with a cold or warm start and on LinUCB in most settings, that the gain comes from source-centered residual learning and persists over 20000 rounds, and that it turns into a loss under strong source–target mismatch.

1 INTRODUCTION

Transfer learning builds on a simple premise: when learning a new task (referred to as the target task), it is often unnecessary to start from scratch. Data or models collected from a related source task may already capture useful structure about the target task, and reusing such knowledge can reduce the amount of target data required for learning. This idea has been studied extensively in supervised learning, where transfer can take the form of reusing representations, predictors, or samples across related domains. More broadly, a central question in transfer learning is when knowledge learned from one task remains useful for another, and how this benefit depends on the similarity between the source and target tasks (Cai et al., 2024).

Contextual bandits make this question more interesting because learning and data collection are coupled. At each round, the learner chooses an action and observes only the reward of that action, so an inaccurate reward model can lead not only to poor predictions but also to less informative data collection in subsequent rounds. Classical contextual bandit methods often impose relatively simple structure on the reward function, such as linearity or membership in a prespecified function class. Neural contextual bandit methods such as NeuralUCB (Zhou et al., 2020) instead use neural networks to model complex nonlinear reward functions and combine these flexible prediction models with uncertainty-based exploration. This flexibility, however, can require many rounds of interaction when the reward model must be learned from scratch. In practice, a neural network may have been trained on related historical data or a previous task before the new bandit interaction begins. Although it is tempting to transfer knowledge from such a model, doing so raises an important question: *can such a pretrained model reduce the exploration required in the target task without making the learner overly reliant on a potentially mismatched source task?*

We develop **Transfer-NeuralUCB** to address this question. Rather than treating the source model as either fully trustworthy or uninformative, our method combines source and target information in complementary ways. The network trained on the source data serves as a starting predictor, and target learning uses it as a benchmark so the online phase learns only the remaining residual. At the same time, when transferring knowledge from the source predictor, we need to account for

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

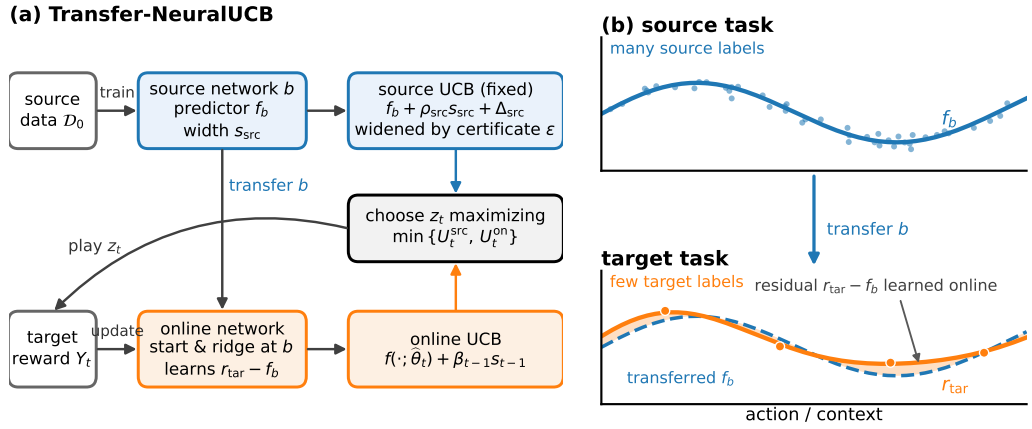


Figure 1: Transfer-NeuralUCB. (a) The source network b provides a predictor f_b and a width s_{src}; the source UCB widens it by the certificate ϵ and stays fixed. The online network starts at b, is regularized toward b, and learns the residual $r_{\text{tar}} - f_b$. Actions maximize the minimum of the two UCBs. (b) Illustration of the transfer: the source network fits the source reward from many labels; on the target task it is the starting predictor, and only the residual $r_{\text{tar}} - f_b$ (shaded) is learned from a few target labels.

its uncertainty, as well as possible mismatch between the source and target tasks. This gives two confidence bounds on the target reward: one inherited from the source task and one learned from target feedback. Transfer-NeuralUCB combines the two confidence bounds by taking their pointwise minimum (Figure 1). This allows reliable source information to reduce exploration, while target data guide learning when the source is less informative. Within this framework, we show that regret is controlled by the smaller of the cumulative source-informed and online confidence widths, and reveal when adapting the source predictor to the target task is easier than learning the target reward from scratch. These theoretical findings are also supported by extensive experiments.

Related literature is deferred to Appendix A.

2 Problem Setting

2.1 Notation and Learning Phases

Notation. For an integer $n \geq 1$ we write $[n] = \{1, \dots, n\}$ and I_n for the $n \times n$ identity; I alone denotes an identity of the dimension implied by its context. Vectors are columns, $\|\cdot\|$ is the Euclidean norm, and $\|x\|_A = \sqrt{x^\top A x}$ for positive definite A. For a positive definite operator Σ on a Hilbert space, $\|h\|_\Sigma = \left\|\Sigma^{1/2} h\right\|$. $a \otimes a$ for the rank-one operator $h \mapsto \langle a, h \rangle a$. We use $\widehat{\cdot}$ to denote a quantity computed from data, such as $\widehat{\theta}_{\text{src}}$, $\bar{\cdot}$ for the corresponding quantity under exact optimization, such as $\bar{u}_n$, and $(\cdot)*$ for an oracle quantity, such as $u*$. The subscripts src, on, and tar denote the source task, the online phase on the target task, and the target task, respectively.

Our problem consists of two phases.

Source phase. Before any interaction, the learner has access to a labeled sample $\mathcal{D}_0 = \{(Z_i^0, Y_i^0)\}_{i=1}^N$ with

$$Y_i^0 = r_{\text{src}}(Z_i^0) + \xi_i^0, \qquad i \in [N], \tag{1}$$

where ξ_i^0 is conditionally σ_{src}-sub-Gaussian along a source filtration (Assumption 4.3).

Online phase. At round $t \in [T]$, the learner observes a finite nonempty candidate set of actions $\mathcal{A}_t \in \mathcal{U}$, selects $z_t \in \mathcal{A}_t$, and observes the reward of the selected action:

$$Y_t = r_{\text{tar}}(z_t) + \eta_t, \tag{2}$$

with η_t conditionally ν-sub-Gaussian given the source data and the online history (Assumption 4.3). Let $z_t^* \in \arg\max_{z \in \mathcal{A}_t} r_{\mathrm{tar}}(z)$ denote the optimal action at round t, performance is measured by the pseudo-regret defined with respect to the target reward function:

$$R_T = \sum_{t=1}^{T} \big[r_{\mathrm{tar}}(z_t^*) - r_{\mathrm{tar}}(z_t) \big]. \tag{3}$$

2.2 Network and training objectives

Network and tangent features. We use a fully connected ReLU network $f(\cdot;\theta)$ with depth L and width m, together with the symmetric paired initialization θ_0 and normalized duplicated inputs of Zhou et al. (2020). Specifically, each context is normalized to have unit Euclidean norm, with identical first and second halves. The input dimension and m are both even. The pairing gives $f(z;\theta_0) = 0$ for every $z \in \mathcal{A}_t$. The *scaled tangent feature* at parameter θ is

$$\phi_\theta(z) = m^{-1/2} \nabla_\theta f(z;\theta).$$

Both phases train the *same realized weights*: the online phase reuses the trained source parameters and does not reinitialize the network. The initialization θ_0 is independent of $\mathcal{D}_0$, the online noises, and the candidate action set.

Source objective. Starting from θ_0, the source network runs J_{src} gradient-descent steps of size η_{src} on the *averaged* ridge objective

$$\mathcal{L}_{\mathrm{src}}(\theta) = \frac{1}{2N} \sum_{i=1}^{N} \big(f(Z_i^0;\theta) - Y_i^0 \big)^2 + \frac{m\lambda_{\mathrm{src}}}{2} \|\theta - \theta_0\|^2, \tag{4}$$

with output $\widehat{\theta}_{\mathrm{src}}$.[1] We write $b = \widehat{\theta}_{\mathrm{src}}$ and $f_b = f(\cdot; b)$ for notational simplicity.

Online objective. Target adaptation starts from the source-trained parameters, with $\widehat{\theta}_1^{\mathrm{on}} = b$. After n online observations, we run J_{on} gradient steps with step size η_{on}, using the previous output as a warm start, to minimize the following objective:

$$\mathcal{L}_n^{\mathrm{on}}(\theta) = \frac{1}{2} \sum_{s=1}^{n} \big(f(z_s;\theta) - Y_s \big)^2 + \frac{m\lambda_{\mathrm{on}}}{2} \|\theta - b\|^2, \tag{5}$$

whose output is denoted by $\widehat{\theta}_{n+1}^{\mathrm{on}}$. It is worth noting that, because the regularizer is centered at b rather than at θ_0, the online phase only needs to learn the *residual* $r_{\mathrm{tar}} - f_b$ rather than the target reward itself. We define $u_n = \sqrt{m}(\widehat{\theta}_{n+1}^{\mathrm{on}} - b)$ as the rescaled local increment, with $u_0 = 0$, and write $\gamma_{\mathrm{src}} = m\eta_{\mathrm{src}}$ and $\gamma_{\mathrm{on}} = m\eta_{\mathrm{on}}$ for the rescaled step sizes.

3 Transfer-NeuralUCB

The main idea of Transfer-NeuralUCB (Algorithm 1) can be summarized in two stages. In the initialization stage, the learner uses the source data to obtain a reward predictor and an associated measure of uncertainty, both of which are transferred to the target task (Section 3.1). In the learning stage, the learner interacts with the target task, maintains a confidence score based on the source information and another based on the target data, and updates the network using the observed target rewards (Section 3.2).

3.1 Initialization: transfer from source data

Both the source and online stages use the same training routine, TrainNN (Algorithm 2), which performs gradient descent on a ridge-regularized squared loss. The main difference from NeuralUCB (Zhou et al., 2020) is that the ridge center and the optimization initialization can be chosen

[1] Using the averaged loss in (4) keeps the ridge parameter on the same scale as in the kernel estimator (37) used in the analysis. If the loss were written as a raw sum instead, the corresponding ridge parameter and step size would be $N\lambda_{\mathrm{src}}$ and η_{src}/N, respectively.

Algorithm 1 Transfer-NeuralUCB

Input: number of rounds T; source sample $\mathcal{D}_0 = \{(Z_i^0, Y_i^0)\}_{i=1}^N$; ridges $\lambda_{\mathrm{src}}, \lambda_{\mathrm{on}}$; step sizes $\eta_{\mathrm{src}}, \eta_{\mathrm{on}}$; numbers of GD steps $J_{\mathrm{src}}, J_{\mathrm{on}}$; network width m and depth L; radii $\rho_{\mathrm{src}}, \{\beta_n\}$; finite-width corrections $\Delta_{\mathrm{src}} = (\alpha_N + \epsilon)b_\sigma + \zeta_\mu$ and $e_{\mathrm{on}} = e_g S_* + e_f$

Initialization: randomly initialize θ_0 as in Section 2.2, so $f(\cdot;\theta_0) = 0$
Let $b = \widehat{\theta}_{\mathrm{src}} = \textsc{TrainNN}\big(N\lambda_{\mathrm{src}}, \eta_{\mathrm{src}}/N, J_{\mathrm{src}}, m, \{Z_i^0\}_{i=1}^N, \{Y_i^0\}_{i=1}^N, \theta_0, \theta_0\big)$
Compute $G_{\mathrm{src}} = \lambda_{\mathrm{src}} I + \frac{1}{N}\sum_{i=1}^N \phi_b(Z_i^0)\phi_b(Z_i^0)^\top$
Set $\widehat{\theta}_1^{\mathrm{on}} = b$ and $V_0 = \lambda_{\mathrm{on}} I$
for $t = 1, \ldots, T$ **do**
  Observe the candidate set $\mathcal{A}_t$
  for $z \in \mathcal{A}_t$ **do**
    Compute $U_t^{\mathrm{src}}(z) = f_b(z) + \rho_{\mathrm{src}}\|\phi_b(z)\|_{G_{\mathrm{src}}^{-1}} + \Delta_{\mathrm{src}}$
    Compute $U_t^{\mathrm{on}}(z) = f(z;\widehat{\theta}_t^{\mathrm{on}}) + \beta_{t-1}\|\phi_{\widehat{\theta}_t^{\mathrm{on}}}(z)\|_{V_{t-1}^{-1}} + e_{\mathrm{on}}$
  end for
  Let $z_t \in \arg\max_{z\in\mathcal{A}_t} \min\{U_t^{\mathrm{src}}(z), U_t^{\mathrm{on}}(z)\}$
  Play z_t and observe the reward $Y_t = r_{\mathrm{tar}}(z_t) + \eta_t$
  Compute $V_t = V_{t-1} + \psi_t\psi_t^\top$ with $\psi_t = \phi_{\widehat{\theta}_t^{\mathrm{on}}}(z_t)$
  Let $\widehat{\theta}_{t+1}^{\mathrm{on}} = \textsc{TrainNN}\big(\lambda_{\mathrm{on}}, \eta_{\mathrm{on}}, J_{\mathrm{on}}, m, \{z_s\}_{s=1}^t, \{Y_s\}_{s=1}^t, b, \widehat{\theta}_t^{\mathrm{on}}\big)$
end for

Algorithm 2 $\textsc{TrainNN}\big(\lambda, \eta, J, m, \{z_i\}_{i=1}^n, \{y_i\}_{i=1}^n, \bar{\theta}, \theta^{(0)}\big)$

Input: regularization parameter λ; step size η; number of GD steps J; network width m; contexts $\{z_i\}_{i=1}^n$; rewards $\{y_i\}_{i=1}^n$; ridge center $\bar{\theta}$; initial parameter $\theta^{(0)}$

Define $\mathcal{L}(\theta) = \sum_{i=1}^n \big(f(z_i;\theta) - y_i\big)^2/2 + m\lambda\|\theta - \bar{\theta}\|^2/2$
for $j = 0, \ldots, J-1$ **do**
  $\theta^{(j+1)} = \theta^{(j)} - \eta\nabla\mathcal{L}(\theta^{(j)})$
end for
return $\theta^{(J)}$

separately. In particular, the algorithm draws a single paired initialization θ_0 and trains the network on the source data, producing the source model $b = \widehat{\theta}_{\mathrm{src}}$. The trained source model is then reused for the target task, without drawing a second initialization.

Transfer the predictor. We reiterate that the online stage starts from b and also centers its regularizer at b. Therefore, instead of learning the target reward r_{tar} from scratch, the online stage only needs to correct the source prediction f_b through the residual $r_{\mathrm{tar}} - f_b$. If b were used only as the initialization and the regularizer remained centered at θ_0, the online stage would no longer have this residual-learning interpretation.

Transfer the uncertainty. Since the source predictor is estimated from finite source data, its predictions may be more reliable for some candidate actions than for others. We therefore transfer not only the source predictor but also a measure of its uncertainty for each action. We measure this uncertainty using the tangent features of the source-trained network:

$$G_{\mathrm{src}} = \lambda_{\mathrm{src}} I + \frac{1}{N}\sum_{i=1}^N \phi_b(Z_i^0)\phi_b(Z_i^0)^\top, \qquad s_{\mathrm{src}}(z) = \|\phi_b(z)\|_{G_{\mathrm{src}}^{-1}}. \tag{6}$$

We refer to $s_{\mathrm{src}}(z)$ as the source confidence width. It reflects how well the candidate action z is covered by the source data. However, even with good source coverage, the source reward may differ from the target reward. We therefore enlarge the confidence width using the transfer certificate ϵ, formally defined in (10), to account for possible source-target mismatch. Thus, the reliability of the transferred predictor depends on both source coverage and the degree of task mismatch.

3.2 ADAPTING TO THE TARGET TASK

During online interaction, Transfer-NeuralUCB maintains two estimates of the target reward. The first is based on the source model and remains fixed throughout the online phase. The second is learned from the target rewards observed so far and is updated as new observations become available. Accordingly, each candidate action z is assigned a source-informed score and an online score

$$U_t^{\mathrm{src}}(z) = f_b(z) + \rho_{\mathrm{src}} s_{\mathrm{src}}(z) + \Delta_{\mathrm{src}}, \qquad U_t^{\mathrm{on}}(z) = f(z; \widehat{\theta}_t^{\mathrm{on}}) + \beta_{t-1} s_{t-1}(z) + e_{\mathrm{on}}. \quad (7)$$

The source score expresses what can be inferred from the source task, after accounting for source uncertainty ($\rho_{\mathrm{src}} s_{\mathrm{src}}(z)$) and task mismatch ($\Delta_{\mathrm{src}}$). The online score expresses what has been learned directly from target feedback. Here $\widehat{\theta}_t^{\mathrm{on}}$ is the network parameter trained on the first $t-1$ target rewards. To measure online uncertainty, we use the tangent features recorded when each action was selected:

$$V_n = \lambda_{\mathrm{on}} I + \sum_{s=1}^{n} \psi_s \psi_s^\top, \quad \psi_s = \phi_{\widehat{\theta}_s^{\mathrm{on}}}(z_s), \qquad s_n(z) = \left\| \phi_{\widehat{\theta}_{n+1}^{\mathrm{on}}}(z) \right\|_{V_n^{-1}}, \quad (8)$$

with $\Gamma_T = \log\{\det(V_T)/\det(\lambda_{\mathrm{on}} I)\}$ the realized log-determinant along the trajectory. The term $\beta_{t-1} s_{t-1}(z)$ is the resulting online confidence bonus, and e_{on} accounts for the approximation error.

The algorithm selects the action based on the smaller of the two upper confidence bounds:

$$z_t \in \arg\max_{z \in \mathcal{A}_t} \min\left\{U_t^{\mathrm{src}}(z),\, U_t^{\mathrm{on}}(z)\right\}. \quad (9)$$

Intuitively, when the source model is informative, the source score can rule out unnecessary exploration. When source information is unreliable or the target task differs substantially from the source task, the online score becomes more informative as target observations accumulate.

After selecting z_t and observing its reward, the algorithm updates the online model using all target observations collected so far through TRAINNN with the loss in (5). The two transfer mechanisms therefore play complementary roles: the source predictor reduces what needs to be learned from the target task, while the source confidence determines how much the source information can reduce exploration.

4 REGRET ANALYSIS

4.1 ASSUMPTIONS AND MAIN RESULT

Let k_{ntk} be the infinite-width neural tangent kernel (NTK) of the network in Section 2.2, $\mathcal{H}_{\mathrm{ntk}}$ its RKHS, and $k_z = k_{\mathrm{ntk}}(z, \cdot)$.[2] Let $\mathcal{Z}_* = \{\tilde{z}_1, \ldots, \tilde{z}_M\}$ be the distinct values of the candidate action set and H_{occ} the NTK Gram matrix of its Q occurrences.

Assumption 4.1 (Inputs and NTK; Zhou et al., 2020, Assumption 4.2)**.** Every action $z \in \mathcal{U}$ can be written as $z = (z^{(1)}, z^{(2)})$, $z^{(1)}, z^{(2)} \in \mathbb{R}^{d/2}$, with $\|z\| = 1$. $H_* = (k_{\mathrm{ntk}}(\tilde{z}_i, \tilde{z}_j))_{i,j \le M} \succeq \lambda_0 I_M$ with $\lambda_0 > 0$. The candidate action set is exogenous and $\sup_{z \in \mathcal{U}} k_{\mathrm{ntk}}(z, z) \le \kappa^2$.

Assumption 4.2 (Source reward and transfer certificate)**.** We assume that the source reward function belongs to the RKHS $\mathcal{H}_{\mathrm{ntk}}$ and satisfies $\|r_{\mathrm{src}}\|_{\mathcal{H}_{\mathrm{ntk}}} \le B_{\mathrm{src}}$. The source-target discrepancy satisfies $d = r_{\mathrm{tar}} - r_{\mathrm{src}} \in \mathcal{H}_{\mathrm{ntk}}$. In addition, let $\widehat{\Sigma}_{N,\lambda_{\mathrm{src}}} = \lambda_{\mathrm{src}} I + \frac{1}{N} \sum_{i \le N} k_{Z_i^0} \otimes k_{Z_i^0}$, we assume that

$$\|d\|^2_{\widehat{\Sigma}_{N,\lambda_{\mathrm{src}}}} = \lambda_{\mathrm{src}} \|d\|^2_{\mathcal{H}_{\mathrm{ntk}}} + \frac{1}{N} \sum_{i=1}^{N} d(Z_i^0)^2 \le \epsilon^2. \quad (10)$$

Assumption 4.3. The source noise ξ_i^0 is conditionally σ_{src}-sub-Gaussian given the source contexts and the preceding source noises ξ_1^0. The target noise η_t is conditionally ν-sub-Gaussian given θ_0, the source phase, the candidate action sets, and the online history (z_1, Y_1).

With the protocol of Appendix B.2, Assumptions 4.1–4.3 correspond to Assumptions B.1–B.2.

[2] Interested readers are referred to Jacot et al. (2018) for further details on neural tangent kernels.

Definition 4.4 (Complexities)**.** The information gain over all candidate action occurrences and the effective dimension (cf. Zhou et al., 2020, Definition 4.3) are given as follows:

$$I_{\mathrm{occ}} = \log\det\Big(I_Q + \frac{H_{\mathrm{occ}}}{\lambda_{\mathrm{on}}}\Big), \qquad \widetilde{d}^{+}_{\mathrm{occ}} = 1 + \frac{I_{\mathrm{occ}}}{\log(1 + Q\kappa^2/\lambda_{\mathrm{on}})}, \tag{11}$$

and the residual complexity is

$$S_* = (1-\bar{a})^{-1/2}\Big[\frac{\alpha_N + \epsilon}{\sqrt{\lambda_{\mathrm{src}}}} + \sqrt{\frac{M}{\lambda_0}}\,\zeta_\mu\Big], \tag{12}$$

with the source radius α_N defined in (70), the source bridge error ζ_μ defined in (71), and the Gram perturbation $\bar{a}$ defined in (73). S_* replaces the cold-start norm $\sqrt{2h^\top H^{-1}h}$ of NeuralUCB.

Theorem 4.5 (Regret of Transfer-NeuralUCB)**.** *Let Assumptions 4.1–4.3 hold, $\lambda_{\mathrm{on}} \geq G^2$, and let the width m, the radius R, the step sizes, and J_{on} satisfy the deterministic conditions of Theorems D.1, D.3, D.6 and Corollary D.8; by Appendix H.3, a width polynomial in the problem constants* (160) *and a logarithmic J_{on}* (161) *suffice. With probability at least $1-\delta_{\mathrm{tot}}$,*

$$R_T \leq 2\sum_{t=1}^{T} \min\big\{\beta_{t-1}s_{t-1}(z_t) + e_{\mathrm{on}},\ \rho_{\mathrm{src}}s_{\mathrm{src}}(z_t) + \Delta_{\mathrm{src}}\big\}, \tag{13}$$

$$\leq 2\min\Big\{\beta_T\sqrt{2T\Gamma_T} + Te_{\mathrm{on}},\ \rho_{\mathrm{src}}\sum_{t=1}^{T} s_{\mathrm{src}}(z_t) + T\Delta_{\mathrm{src}}\Big\}, \qquad \beta_T = \max_{n<T}\beta_n, \tag{14}$$

where

$$\beta_T\sqrt{2T\Gamma_T} + Te_{\mathrm{on}} = O\Big(\big[\nu\sqrt{I_{\mathrm{occ}} + 2\log(1/\delta_{\mathrm{on}}) + 1} + \sqrt{\lambda_{\mathrm{on}}}S_* + 1\big]\sqrt{T(I_{\mathrm{occ}}+1)} + \sqrt{T}\Big), \tag{15}$$

which is $\widetilde{O}(\widetilde{d}^{+}_{\mathrm{occ}}\sqrt{T})$ when $\nu = O(1)$, $\log(1/\delta_{\mathrm{on}}) = O(\log T)$, and $\lambda_{\mathrm{on}}S_*^2 = O(\widetilde{d}^{+}_{\mathrm{occ}})$. The radii $\rho_{\mathrm{src}}, \beta_n$ and corrections $\Delta_{\mathrm{src}}, e_{\mathrm{on}}$ are those of (105) *and* (102)–(103)*.*

Several implications of Theorem 4.5 are worth mentioning.

• **What determines the transfer benefit?** Equation (14) shows that the benefit of transfer depends on two factors. The source-based bound is determined by the uncertainty left by the source data and the allowance for source-target mismatch. It is tighter when the source data provide good coverage and the mismatch is small, allowing source information to reduce exploration before much target data have been collected. The online bound, in contrast, depends on how difficult it is to adapt the source predictor to the target task through the residual bound S_*. A smaller valid S_* reduces the corresponding regularization term, although the information-dependent stochastic term remains.

• **Comparison with existing regret bounds.** The online branch of (15) has the same effective-dimension form as standard kernel and neural UCB bounds. In particular, NeuralUCB (Zhou et al., 2020) obtains $\widetilde{O}(\widetilde{d}\sqrt{T})$, where $\widetilde{d}$ is the effective dimension of the NTK Gram matrix. Our corresponding term has the same dependence on T and NTK information, but its regularization cost is governed by the residual complexity S_* rather than the cold-start target complexity: the online stage only needs to correct the transferred source predictor. Transfer additionally introduces a source-informed bound that accounts for both source uncertainty and source-target mismatch: $\rho_{\mathrm{src}}\sum_{t=1}^{T} s_{\mathrm{src}}(z_t) + T\Delta_{\mathrm{src}}$, so the regret is controlled by the smaller of the source-informed and online confidence widths. This minimum structure is related to offline-to-online methods such as MIN-UCB (Cheung & Lyu, 2024) and the directional linear bound of Han et al. (2026). In our setting, however, both confidence bounds are evaluated along the same source-initialized neural trajectory. Transfer can therefore help in two ways: by making target adaptation easier or by providing sufficiently informative source coverage.

We next provide a proof sketch of the main result.

4.2 Proof, part I: from a neural network to a UCB

The first step is to connect the wide-network analysis to the UCB algorithm. Under the finite-width NTK conditions, we construct two valid UCBs for the target reward: one based on the source predictor and the other based on the online predictor. The finite-width results ensure that the confidence bounds hold uniformly along the corresponding training iterates.

Lemma 4.6 (From valid scores to regret)**.** *Suppose that for every $t \le T$ and $z \in \mathcal{A}_t$,*

$$|r_{\mathrm{tar}}(z) - f_b(z)| \le \rho_{\mathrm{src}} s_{\mathrm{src}}(z) + \Delta_{\mathrm{src}}, \qquad |r_{\mathrm{tar}}(z) - f(z; \widehat{\theta}_t^{\mathrm{on}})| \le \beta_{t-1} s_{t-1}(z) + e_{\mathrm{on}}. \tag{16}$$

Then (13) *holds, and if $\lambda_{\mathrm{on}} \ge G^2$ and $\|\psi_t\| \le G$,* (14) *holds.*

4.3 Proof, Part II: Residual Learning Around the Source Model

The next two lemmas adapt the corresponding steps in the NeuralUCB analysis in Zhou et al. (2020) to the case where online learning starts from the source-trained parameters. Let $\widehat{r}_{N,\lambda_{\mathrm{src}}}$ denote the exact NTK ridge estimator on the source sample, and let $\widehat{\sigma}_N$ denote its associated confidence width, as defined in (37)–(38).[3]

Lemma 4.7 (Residual representation at b)**.** *Suppose the step-size and radius conditions* (58)–(59)*. On the event of Theorem C.1, together with the source label event* (53) *and the source self-normalized event* (74)*, if $\bar{a} < 1$, then there exists a vector u_*, measurable with respect to the source phase and the candidate action sets, such that*

$$r_{\mathrm{tar}}(z) - f_b(z) = \phi_b(z)^\top u_* \quad (z \in \mathcal{Z}_*), \qquad \|u_*\| \le S_*. \tag{17}$$

Lemma 4.8 (Online Confidence Around the Source-Trained Parameters)**.** *On the events of Lemma 4.7 and under the deterministic conditions of Theorem D.6, with conditional probability at least $1 - \delta_{y,\mathrm{on}} - \delta_{\mathrm{on}}$ given the source phase, for all $0 \le n < T$ and $z \in \mathcal{Z}_*$,*

$$|r_{\mathrm{tar}}(z) - f(z; \widehat{\theta}_{n+1}^{\mathrm{on}})| \le \underbrace{\sqrt{1 + c_V}\,\left[a_n + \sqrt{\lambda_{\mathrm{on}} + nG^2}\,E_n\right]}_{\beta_n} s_n(z) + \underbrace{e_g S_* + e_f}_{e_{\mathrm{on}}}, \tag{18}$$

where

$$a_n = \nu \sqrt{\log \frac{\det A_n}{\det(\lambda_{\mathrm{on}} I)} + 2\log(1/\delta_{\mathrm{on}})} + \sqrt{\lambda_{\mathrm{on}}} S_*, \tag{19}$$

$A_n = \lambda_{\mathrm{on}} I + \sum_{s \le n} \phi_b(z_s)\phi_b(z_s)^\top$, E_n is the optimization error (92)*, and $c_V = T(2Ge_g + e_g^2)/\lambda_{\mathrm{on}} < 1$.*

Proof sketch of Theorem 4.5. A standard argument shows that the source confidence bound holds with probability at least $1 - \delta_{\mathrm{lazy}} - \delta_y - \delta_{\mathrm{src}}$ (see (152) in Appendix H.1). Lemma 4.8 shows that the online confidence bound around the source-trained parameter holds with conditional probability at least $1 - \delta_{y,\mathrm{on}} - \delta_{\mathrm{on}}$. Taking expectations and a union bound, both hold with probability at least $1 - \delta_{\mathrm{tot}}$. Applying Lemma 4.6, we obtain (13)–(14). For the result stated in (15), we use (110) to upper bound a_n and use $I_{\mathrm{occ}} + 1$ to upper bound Γ_T. The conditions of Corollary D.8 give $\sqrt{\lambda_{\mathrm{on}} + nG^2}\,E_n \le 1$ and $Te_{\mathrm{on}} = O(\sqrt{T})$. □

5 Experiments

We compare Transfer-NeuralUCB with NeuralUCB (Zhou et al., 2020), cold and with a warm start at the source-trained parameter b, and with LinUCB (Abbasi-Yadkori et al., 2011). All neural methods use the same two-layer ReLU network and training routine, and every method's exploration multiplier is selected from the same 16-value grid on three tuning instances disjoint from the evaluation instances. Appendix I gives the training details, all tables, and the statistical tests.

5.1 Synthetic datasets

Contexts are unit vectors in $\mathbb{R}^{12}$ ($K = 10$ candidates per round from $M = 100$ values), and $N = 1000$ source contexts are labeled. The source reward is a random NTK kernel expansion, and the discrepancy d is an independent one rescaled so that the certificate (10) holds with equality at a chosen ϵ. We sweep $\epsilon \in \{0, 0.02, 0.05, 0.1, 0.2, 0.4, 0.8\}$, and vary $N \in \{125, 500, 4000\}$, $M = 1000$, and the noise level at $\epsilon = 0.1$. The network has width $m = 300$; each run has $T = 2000$ rounds, repeated over 20 instances.

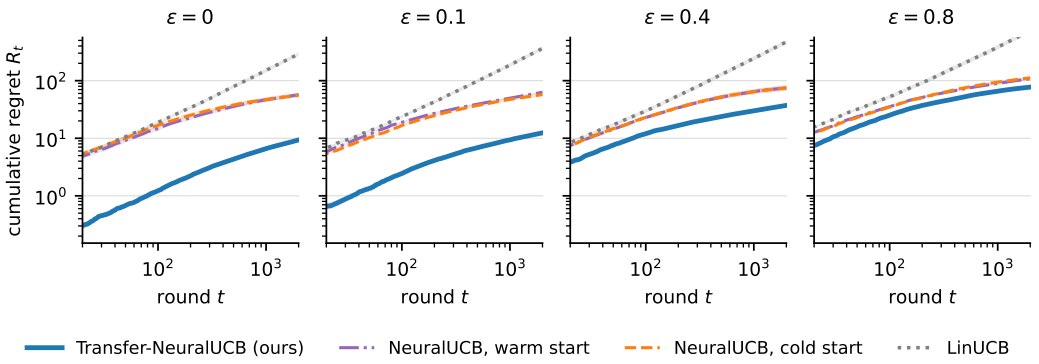


Figure 2: Synthetic cumulative regret against the round (log–log) at mismatch $\epsilon = 0$ (identical rewards), 0.1, 0.4, and 0.8. Mean and one standard error over 20 instances, $T = 2000$.

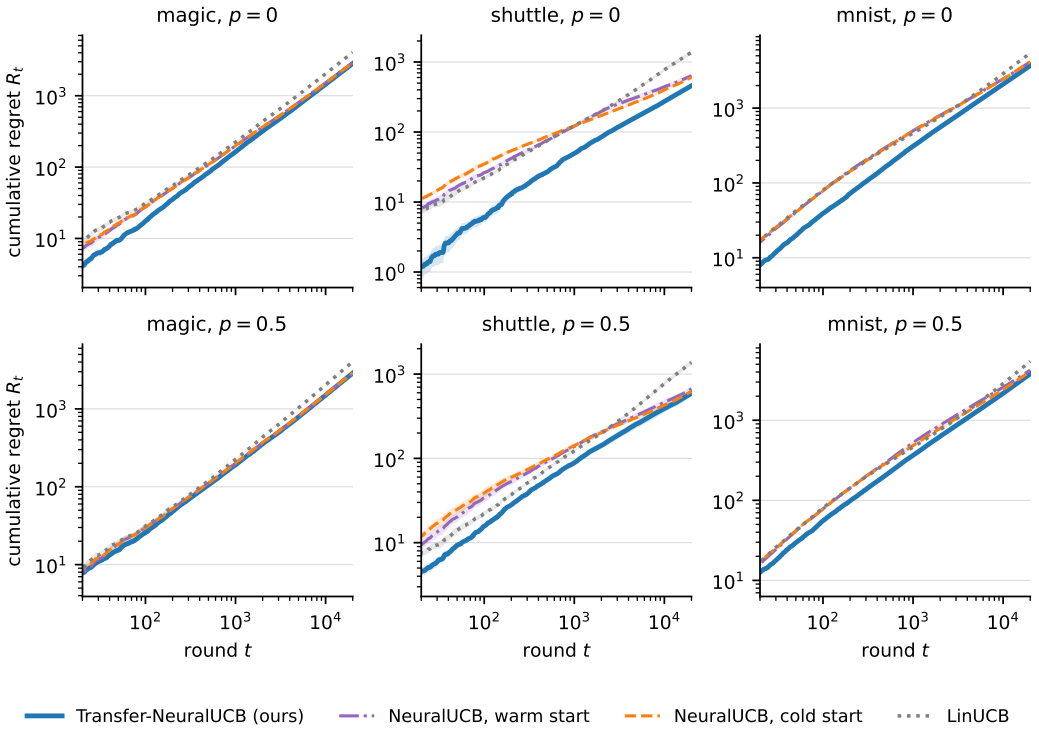


Figure 3: Classification-derived bandits over $T = 20000$ rounds (log–log), reward-mixture source with $p = 0$ (top) and $p = 0.5$ (bottom). Mean and one standard error over 10 instances.

5.2 Real-world datasets

We use four classification datasets, `magic`, `shuttle`, `covertype`, and `mnist` (Table 1), converted into k-armed bandits as in Zhou et al. (2020): the arms of an instance are its k block one-hot contexts, and the reward is 1 for the true class and 0 otherwise. The source sample labels all arms of about $2000/k$ disjoint instances. Two source protocols create the mismatch. The *reward-mixture* source $r_{\text{src}}(z_a) = (1-p)\mathbf{1}\{a = y\} + p\,\mathbf{1}\{a = \pi(y)\}$ permutes a fraction $p \in \{0, 0.25, 0.5, 0.75\}$ of the labels (163). The *covariate-shift* source keeps the true labels but draws the source instances from one 2-means cluster and the online stream from the other. Each run has $T = 2000$ rounds, re-

[3]These quantities are introduced only for the analysis and are not computed by the algorithm.

Table 1: Classification-derived bandits. Context dimension is $2kd_f$ after the block one-hot and duplicated-input maps; `shuttle` has only nine raw features, so $d_f = 9$ there.

	`magic`	`shuttle`	`covertype`	`mnist`
instances	19020	58000	581012	70000
classes k	2	7	7	10
context dimension	40	126	140	200

peated over 20 instances. The certificate is not known on real data; ϵ and B_{src} are plug-in estimates, so the source radius is not certified there. For the long-horizon runs of Figure 3 and Table 7, the reward-mixture protocol with $p \in \{0, 0.5\}$ is run for $T = 20000$ rounds on `magic`, `shuttle`, and `mnist`, with the multipliers selected at $T = 2000$ and 10 instances.

5.3 Results

Similar source and target tasks. When the source task is informative, transfer helps substantially. At $\epsilon = 0$ the final regret of Transfer-NeuralUCB is 9.3 ± 0.7 against 55.7 ± 4.4 for the cold start (Figure 2), and on real data with $p \leq 0.5$ it has the lowest regret on every dataset, e.g., 87 against 193 on `shuttle` at $p = 0$ (Table 5). Across the 32 settings it has the lowest regret in 26, each time significantly below all three baselines (Table 3). The warm start coincides with the cold start, so the gain comes from centering the online model at b, not from the initialization alone. The lead persists over 20000 rounds (Figure 3), although the relative gain shrinks with the horizon, from a ratio of 0.49 to the cold start at $T = 2000$ to 0.76 at $T = 20000$ on `shuttle` (Table 7).

Dissimilar tasks. The benefit decreases as the tasks drift apart, and it is not uniform. In the synthetic sweep the regret rises from 9.3 to 77 as ϵ grows to 0.8, still below the cold start's 112 (Table 4). On real data the method is significantly outperformed only under strong mismatch: when three quarters of the source labels are permuted ($p = 0.75$), the cold start is better on `magic` and `covertype` and LinUCB on `covertype` and `mnist`, and when the source instances come from a different region (covariate shift), LinUCB is better on `covertype` (Table 5). In these cases the source predictor is confidently wrong on the target instances, and the plug-in certificate, computed on the source design, does not reveal it. Avoiding this cost would require a reliable estimate of the mismatch from target feedback, which we leave to future work.

6 Conclusion

We proposed Transfer-NeuralUCB, which transfers both a predictor and its uncertainty from a source task to a neural contextual bandit. The online model is trained around the source-trained parameters, so it only has to learn the residual between the target reward and the source prediction. A source-informed confidence bound, enlarged by a certificate for task mismatch, is combined with the online bound through their pointwise minimum. The regret is controlled by the smaller of the two cumulative confidence widths, and its online part has the effective-dimension form of NeuralUCB with the residual complexity S_* in place of the complexity of the target reward. The main technical step is a confidence analysis around a center that depends on the source rewards rather than on a random initialization. In experiments, Transfer-NeuralUCB improves on NeuralUCB and LinUCB in most synthetic and classification-derived settings, and the improvement persists over 20000 rounds.

The guarantee relies on a supplied transfer certificate and on neural-tangent-kernel conditions; on real data the certificate is only a plug-in estimate. When the source predictor is confidently wrong, as under strong label mismatch or covariate shift, this estimate does not detect the problem and transfer can increase regret. At tuned radii the source score rarely attains the minimum, so the empirical gain comes from residual learning. Estimating the mismatch from target feedback, and a residual bound that reflects a small realized residual (Remark D.5), are natural next steps.

AI DISCLOSURE

In this work, we used AI tools to polish the paper and write code for generating the figures based on experimental data. We have reviewed all AI-assisted work carefully. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, volume 24, 2011.

Zeyuan Allen-Zhu, Yuanzhi Li, and Zhao Song. A convergence theory for deep learning via over-parameterization. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 242–252. PMLR, 2019.

Sanjeev Arora, Simon S. Du, Wei Hu, Zhiyuan Li, Ruslan Salakhutdinov, and Ruosong Wang. On exact computation with an infinitely wide neural net. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Ha Manh Bui, Enrique Mallada, and Anqi Liu. Variance-aware linear UCB with deep representation for neural contextual bandits. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 4213–4221. PMLR, 2025. URL https://proceedings.mlr.press/v258/bui25a.html.

Changxiao Cai, T. Tony Cai, and Hongzhe Li. Transfer learning for contextual multi-armed bandits. *The Annals of Statistics*, 52(1):207–232, 2024. doi: 10.1214/23-AOS2341. URL https://arxiv.org/abs/2211.12612.

Yuan Cao and Quanquan Gu. Generalization bounds of stochastic gradient descent for wide and deep neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Wang Chi Cheung and Lixing Lyu. Leveraging (Biased) information: Multi-armed bandits with offline data. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 8286–8309. PMLR, 2024. URL https://proceedings.mlr.press/v235/cheung24a.html.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 844–853. PMLR, 2017.

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pp. 208–214. PMLR, 2011. URL https://proceedings.mlr.press/v15/chu11a.html.

Zean Han, Ruihan Lin, Zezhen Ding, and Jiheng Zhang. Direction-aware offline-to-online learning in linear contextual bandits. *arXiv preprint arXiv:2604.24016*, 2026.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Ilja Kuzborskij and Francesco Orabona. Stability and hypothesis transfer learning. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 942–950. PMLR, 2013. URL https://proceedings.mlr.press/v28/kuzborskij13.html.

Mikhail Terekhov. Contextual bandit optimization with pre-trained neural networks. *arXiv preprint arXiv:2501.06258*, 2025. URL https://arxiv.org/abs/2501.06258.

Michal Valko, Nathaniel Korda, Rémi Munos, Ilias Flaounas, and Nelo Cristianini. Finite-time analysis of kernelised contextual bandits. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence*, 2013. URL https://arxiv.org/abs/1309.6869.

Andrew Wagenmaker and Aldo Pacchiano. Leveraging offline data in online reinforcement learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 35300–35338. PMLR, 2023. URL https://proceedings.mlr.press/v202/wagenmaker23a.html.

Ziyan Wang, Xiaoming Huo, and Hao Wang. Towards domain adaptive neural contextual bandits. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2406.09564.

Pan Xu, Zheng Wen, Handong Zhao, and Quanquan Gu. Neural contextual bandits with deep representation and shallow exploration. *arXiv preprint arXiv:2012.01780*, 2020. URL https://arxiv.org/abs/2012.01780.

Chicheng Zhang, Alekh Agarwal, Hal Daumé, III, John Langford, and Sahand N. Negahban. Warm-starting contextual bandits: Robustly combining supervised and bandit feedback. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 7335–7344. PMLR, 2019. URL https://proceedings.mlr.press/v97/zhang19b.html.

Weitong Zhang, Dongruo Zhou, Lihong Li, and Quanquan Gu. Neural Thompson sampling. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2010.00827.

Dongruo Zhou, Lihong Li, and Quanquan Gu. Neural contextual bandits with UCB-based exploration. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 11492–11502. PMLR, 2020.

A Related work

Neural contextual bandits. UCB methods tie reward prediction to exploration through a confidence width, first for linear rewards (Abbasi-Yadkori et al., 2011; Chu et al., 2011) and then for kernels (Valko et al., 2013; Chowdhury & Gopalan, 2017). NeuralUCB uses the tangent features of a wide network as the feature map (Zhou et al., 2020); Neural Thompson Sampling (Zhang et al., 2021), last-layer exploration (Xu et al., 2020), and variance-aware widths (Bui et al., 2025) follow the same template. The analyses rest on NTK and lazy-training bounds (Jacot et al., 2018; Allen-Zhu et al., 2019; Cao & Gu, 2019; Arora et al., 2019) and center the confidence set at the random initialization θ_0. We reuse these tools but center them at a source-trained parameter.

Offline-to-online learning with biased data. Unbiased offline data reduce the online exploration burden directly (Zhang et al., 2019; Wagenmaker & Pacchiano, 2023). With biased data the mismatch must be budgeted. MIN-UCB takes the pointwise minimum of an online and an offline-informed UCB under a scalar bias bound (Cheung & Lyu, 2024), and Han et al. (2026) use a directional bound in linear bandits. Our certificate (10) is the RKHS analogue of the directional bound. The minimum rule is inherited; our contribution is to make both branches valid along finite-width neural training.

Transfer learning. Statistical transfer theory derives rates in terms of a task-relatedness parameter, including for contextual bandits under covariate shift (Cai et al., 2024). Hypothesis-transfer learning corrects a given source predictor (Kuzborskij & Orabona, 2013), which is closest to our residual view. Terekhov (2025) study bandits with a pretrained network but do not model the source training, and Wang et al. (2025) let the context distribution shift rather than the reward. Here both the source sample size and the reward mismatch enter the regret, through α_N and ϵ.

B SETUP, NOTATION, AND ASSUMPTIONS

B.1 SYMBOL REFERENCE

Table 2: Symbols used in the paper. "Defined" points to the display, theorem, or section that introduces the symbol.

Symbol	Meaning	Defined
Tasks, data, and regret		
$\mathcal{U}$	finite universe of action-context vectors, $\lvert\mathcal{U}\rvert \le N + Q$	§2.1
$r_{\mathrm{src}}, r_{\mathrm{tar}}$	fixed unknown source and target reward functions	§2.1
$d = r_{\mathrm{tar}} - r_{\mathrm{src}}$	task discrepancy	(10)
$N, \mathcal{D}_0$	source sample size and labeled source sample	(1)
Z_i^0, Y_i^0, ξ_i^0	source context, label, and noise	(1)
σ_{src}	sub-Gaussian scale of the source noise	(30)
$T, \mathcal{A}_t$	online horizon and round-t candidate set	(2)
K_{arm}	$\max_{t\le T} \lvert\mathcal{A}_t\rvert$	§2.1
z_t, z_t^*	selected and optimal action at round t	(3)
Y_t, η_t, ν	online reward, noise, and its sub-Gaussian scale	(34)
R_T	pseudo-regret over T rounds	(3)
Candidate sets and kernel information		
Q, H_{occ}	occurrence-list length and its exact NTK Gram (possibly singular)	(40)
$\mathcal{Z}_*, \tilde{z}_i, M$	distinct candidate values and their number	(35)
H_*, λ_0	exact NTK Gram on $\mathcal{Z}_*$ and its eigenvalue floor	(35)
κ	uniform bound $k_{\mathrm{ntk}}(z,z) \le \kappa^2$ on $\mathcal{U}$	Ass. B.2
$K, k(z)$	exact source NTK Gram and cross-kernel vector	(36)
$I_N(\lambda_{\mathrm{src}}), I_{\mathrm{occ}}$	source and occurrence-level information	(39),(40)
$\widetilde{d}_{\mathrm{occ}}^{+}$	occurrence-level effective dimension	(114)
Γ_T	realized log-determinant $\log\{\det V_T / \det(\lambda_{\mathrm{on}} I)\}$	(8)
Network and tangent features		
$f(z;\theta), L, m$	ReLU network, depth, and width	§2.2
θ_0	symmetric paired initialization; $f(\cdot;\theta_0) \equiv 0$ on $\mathcal{U}$	§2.2
$\phi_\theta(z)$	scaled tangent feature $m^{-1/2}\nabla_\theta f(z;\theta)$	(22)
$k_{\mathrm{ntk}}, \mathcal{H}_{\mathrm{ntk}}, k_z$	infinite-width NTK, its RKHS, and $k_{\mathrm{ntk}}(z,\cdot)$	§B.2
Training and estimators		
$\mathcal{L}_{\mathrm{src}}, J_{\mathrm{src}}, \eta_{\mathrm{src}}, \gamma_{\mathrm{src}}$	source objective, steps, step size, and $\gamma_{\mathrm{src}} = m\eta_{\mathrm{src}}$	(23)
$\lambda_{\mathrm{src}}, \lambda_{\mathrm{on}}$	source and online ridge parameters	(23),(26)
$\widehat{\theta}_{\mathrm{src}} = b$, f_b	trained source parameter and its prediction	(23)
$\mathcal{L}_n^{\mathrm{on}}, J_{\mathrm{on}}, \eta_{\mathrm{on}}, \gamma_{\mathrm{on}}$	online objective, inner steps, step size, and $\gamma_{\mathrm{on}} = m\eta_{\mathrm{on}}$	(26)
$\widehat{\theta}_n^{\mathrm{on}}, u_n$	online iterate and increment $\sqrt{m}(\widehat{\theta}_{n+1}^{\mathrm{on}} - b)$	(87)
$\widehat{r}_{N,\lambda_{\mathrm{src}}}, \widehat{\sigma}_N$	exact NTK ridge predictor and posterior width (analysis only)	(37),(38)
$\bar{u}_n$	exact ridge solution in the frozen features at b	(86)
Covariances, widths, and scores		
$G_{\mathrm{src}}, s_{\mathrm{src}}$	source covariance and width	(24),(25)
ψ_s, V_n, s_n	recorded feature, dynamic covariance, and online width	(89),(90)
A_n	frozen-feature covariance at b	(85)
$U_t^{\mathrm{src}}, U_t^{\mathrm{on}}$	source and online upper confidence scores	(27),(28)
Transfer certificate and residual complexity		
B_{src}	RKHS norm bound on r_{src}	Ass. B.1
$\widehat{\Sigma}_{N,\lambda_{\mathrm{src}}}$	empirical RKHS covariance of the source design	(31)
ϵ	transfer certificate: design-weighted bound on d	(32)
h, g, q_*	target values, source predictions, and residual on $\mathcal{Z}_*$	(78)
$\Phi_{\mathrm{src},*}, K_{\mathrm{src},*}$	trained source features on $\mathcal{Z}_*$ and their Gram	(72)
$\Phi_{0,*}, K_{0,*}$	initialization-centered analogues	(77)
u_*, u_{tar}	residual representer and cold-start target representer	(79)
S_*	certified upper bound on $\lVert u_*\rVert$	(76)
$\langle\cdot,\cdot\rangle_*, \lVert\cdot\rVert_*$	inverse-kernel metric $a^\top H_*^{-1} b$	(80)
τ, ρ	relative scale and alignment of g against h	(80)

continued on next page

Symbol	Meaning	Defined
Calibration radii		
α_N	source statistical confidence radius	(70)
$\rho_{\mathrm{src}}, \Delta_{\mathrm{src}}$	source score radius and additive correction	(105)
a_n	self-normalized residual radius	(101)
β_n	online score radius	(102)
e_{on}	online additive correction $e_g S_* + e_f$	(103)
Finite-width and optimization constants		
R	lazy-ball radius, $\|u\| \le R$	Thm C.1
G, e_g, e_f	feature-norm, feature-drift, and Taylor bounds on the ball	(43)–(45)
ϵ_{ntk}	initial kernel accuracy at θ_0	(46)
ζ_μ, ζ_σ	source prediction and width bridge errors	(71),(67)
$D_z, \rho_A, a_\sigma, b_\sigma$	kernel and covariance perturbations in the source bridge	(62),(64),(104)
$\bar{d}_{\mathrm{src}}, \bar{a}, \bar{a}_0$	candidate Gram perturbation and its relative versions	(73),(81)
c_V	covariance-sandwich constant	(97)
c	exploration-scale multiplier in the experiments (not a theory constant)	§5
B_0, E_{src}	source label bound and source optimization error	(52),(60)
$B_y, R_1, R_{\mathrm{out}}$	online label bound and online trajectory radii	(83),(84)
q, D_n, E_n	contraction factor and online optimization errors	(91)–(93)
Failure probabilities		
δ_{lazy}	finite-width uniform event	Thm C.1
$\delta_y, \delta_{y,\mathrm{on}}$	source and online label-boundedness events	(53),(94)
$\delta_{\mathrm{src}}, \delta_{\mathrm{on}}$	source and online self-normalized events	(74), Thm D.6
δ_{tot}	their sum, assumed < 1	(106)

B.2 Protocol, network, and algorithm

Protocol. Source sample and online rounds $t \in [T]$:

$$Y_i^0 = r_{\mathrm{src}}(Z_i^0) + \xi_i^0 \quad (i \le N), \qquad Y_t = r_{\mathrm{tar}}(z_t) + \eta_t, \quad z_t \in \mathcal{A}_t, \tag{20}$$

$$R_T = \sum_{t=1}^{T} \big[r_{\mathrm{tar}}(z_t^*) - r_{\mathrm{tar}}(z_t) \big], \qquad z_t^* \in \arg\max_{z \in \mathcal{A}_t} r_{\mathrm{tar}}(z). \tag{21}$$

$Z_i^0 \in \mathcal{U}$ and every candidate lies in the finite set $\mathcal{U}$; the candidate list $(\mathcal{A}_1, \dots, \mathcal{A}_T)$ is exogenous, has $Q = \sum_t |\mathcal{A}_t| \le T K_{\mathrm{arm}}$ occurrences, and $|\mathcal{U}| \le N + Q$. $N, T \ge 1$, $L \ge 2$, $\lambda_{\mathrm{src}}, \lambda_{\mathrm{on}}, \eta_{\mathrm{src}}, \eta_{\mathrm{on}} > 0$, $J_{\mathrm{src}}, J_{\mathrm{on}} \in \mathbb{N}$, and $\delta_{\mathrm{lazy}} + \delta_y + \delta_{\mathrm{src}} + \delta_{y,\mathrm{on}} + \delta_{\mathrm{on}} < 1$.

Network. Depth-L, width-m ReLU network with the paired initialization of Zhou et al. (2020); inputs are unit vectors with equal halves, and d and m are even. θ_0 is independent of $\mathcal{D}_0$, of the online noises, and of the candidate list, and $f(z; \theta_0) = 0$ for $z \in \mathcal{U}$.

$$\phi_\theta(z) = m^{-1/2} \nabla_\theta f(z; \theta), \qquad k_{\mathrm{ntk}} = \text{infinite-width NTK}, \quad \mathcal{H}_{\mathrm{ntk}} = \text{its RKHS}, \quad k_z = k_{\mathrm{ntk}}(z, \cdot). \tag{22}$$

Source phase. J_{src} gradient steps of size η_{src} from θ_0 on

$$\mathcal{L}_{\mathrm{src}}(\theta) = \frac{1}{2N} \sum_{i=1}^{N} \big(f(Z_i^0; \theta) - Y_i^0 \big)^2 + \frac{m \lambda_{\mathrm{src}}}{2} \|\theta - \theta_0\|^2, \qquad b := \widehat{\theta}_{\mathrm{src}}, \quad f_b := f(\cdot; b), \tag{23}$$

$$G_{\mathrm{src}} = \lambda_{\mathrm{src}} I + \frac{1}{N} \sum_{i=1}^{N} \phi_b(Z_i^0) \phi_b(Z_i^0)^\top, \tag{24}$$

$$s_{\mathrm{src}}(z) = \|\phi_b(z)\|_{G_{\mathrm{src}}^{-1}}. \tag{25}$$

A raw-sum source loss is equivalent with ridge $N\lambda_{\mathrm{src}}$ and step η_{src}/N.

Online phase. $\widehat{\theta}_1^{\mathrm{on}} = b$; after n rewards, J_{on} gradient steps of size η_{on} started at $\widehat{\theta}_n^{\mathrm{on}}$ on

$$\mathcal{L}_n^{\mathrm{on}}(\theta) = \frac{1}{2}\sum_{s=1}^{n}\big(f(z_s;\theta) - Y_s\big)^2 + \frac{m\lambda_{\mathrm{on}}}{2}\,\|\theta - b\|^2 \tag{26}$$

return $\widehat{\theta}_{n+1}^{\mathrm{on}}$. With V_{t-1} and s_{t-1} of (89)–(90),

$$U_t^{\mathrm{src}}(z) = f_b(z) + \rho_{\mathrm{src}} s_{\mathrm{src}}(z) + \Delta_{\mathrm{src}}, \tag{27}$$

$$U_t^{\mathrm{on}}(z) = f(z;\widehat{\theta}_t^{\mathrm{on}}) + \beta_{t-1} s_{t-1}(z) + e_{\mathrm{on}}, \tag{28}$$

$$z_t \in \arg\max_{z\in\mathcal{A}_t} U_t(z), \qquad U_t = \min\{U_t^{\mathrm{src}}, U_t^{\mathrm{on}}\}. \tag{29}$$

B.3 Assumptions and exact kernel objects

Assumption B.1 (Source RKHS and transfer certificate)**.** $B_{\mathrm{src}}, \sigma_{\mathrm{src}}, \epsilon \geq 0$. Conditional on the source contexts, there is a filtration $(\mathcal{F}_i^{\mathrm{src}})_{i\leq N}$ with ξ_i^0 $\mathcal{F}_i^{\mathrm{src}}$-measurable and

$$\mathbb{E}[e^{a\xi_i^0} \mid \mathcal{F}_{i-1}^{\mathrm{src}}] \leq e^{a^2\sigma_{\mathrm{src}}^2/2} \quad (a\in\mathbb{R}). \tag{30}$$

$r_{\mathrm{src}} \in \mathcal{H}_{\mathrm{ntk}}$ with $\|r_{\mathrm{src}}\|_{\mathcal{H}_{\mathrm{ntk}}} \leq B_{\mathrm{src}}$, and $d \in \mathcal{H}_{\mathrm{ntk}}$ with

$$\widehat{\Sigma}_{N,\lambda_{\mathrm{src}}} = \lambda_{\mathrm{src}} I + \frac{1}{N}\sum_{i=1}^{N} k_{Z_i^0}\otimes k_{Z_i^0}, \tag{31}$$

$$\|d\|_{\widehat{\Sigma}_{N,\lambda_{\mathrm{src}}}} \leq \epsilon, \qquad \|h\|_{\Sigma} := \Big\|\Sigma^{1/2}h\Big\|_{\mathcal{H}_{\mathrm{ntk}}}. \tag{32}$$

For a linear kernel $k(z,z') = \varphi(z)^\top\varphi(z')$ with $r_\bullet = \varphi^\top\theta_\bullet$, (32) reads $\|\theta_{\mathrm{tar}} - \theta_{\mathrm{src}}\|_{\lambda_{\mathrm{src}}I + N^{-1}\sum_i \varphi(Z_i^0)\varphi(Z_i^0)^\top} \leq \epsilon$, the direction-aware condition of Han et al. (2026).

Assumption B.2 (Online filtration and kernel nondegeneracy)**.** Let

$$\mathcal{S}_{\mathrm{src}} = \sigma\big(\theta_0, \mathcal{D}_0, \text{source phase, candidate list}\big), \qquad \mathcal{F}_{t-1} = \mathcal{S}_{\mathrm{src}} \vee \sigma(z_s, Y_s : s < t). \tag{33}$$

z_t is $\mathcal{F}_{t-1}$-measurable (measurable tie-breaking) and

$$\mathbb{E}[e^{a\eta_t} \mid \mathcal{F}_{t-1}] \leq e^{a^2\nu^2/2} \quad (a\in\mathbb{R}) \text{ a.s.} \tag{34}$$

With $\mathcal{Z}_* = \{\tilde{z}_1, \ldots, \tilde{z}_M\}$ the distinct values of the candidate list (every $z_t \in \mathcal{Z}_*$),

$$H_* = \big(k_{\mathrm{ntk}}(\tilde{z}_i, \tilde{z}_j)\big)_{i,j\leq M} \succeq \lambda_0 I_M, \qquad \lambda_0 > 0, \qquad \sup_{z\in\mathcal{U}} k_{\mathrm{ntk}}(z,z) \leq \kappa^2. \tag{35}$$

Exact kernel objects. With $Y^0 = (Y_i^0)_{i\leq N}$,

$$K = \big(k_{\mathrm{ntk}}(Z_i^0, Z_j^0)\big)_{i,j\leq N}, \qquad k(z) = \big(k_{\mathrm{ntk}}(Z_i^0, z)\big)_{i\leq N}, \tag{36}$$

$$\widehat{r}_{N,\lambda_{\mathrm{src}}}(z) = k(z)^\top (K + N\lambda_{\mathrm{src}} I)^{-1} Y^0, \tag{37}$$

$$\widehat{\sigma}_N(z)^2 = \lambda_{\mathrm{src}}^{-1}\big[k_{\mathrm{ntk}}(z,z) - k(z)^\top (K + N\lambda_{\mathrm{src}} I)^{-1} k(z)\big] = \|k_z\|^2_{\widehat{\Sigma}_{N,\lambda_{\mathrm{src}}}^{-1}}, \tag{38}$$

$$I_N(\lambda_{\mathrm{src}}) = \tfrac{1}{2}\log\det\big(I_N + K/(N\lambda_{\mathrm{src}})\big). \tag{39}$$

The algorithm never computes (37)–(38). H_{occ} is the exact NTK Gram matrix of the Q candidate occurrences (repeated values kept, possibly singular) and

$$I_{\mathrm{occ}} = \log\det(I_Q + H_{\mathrm{occ}}/\lambda_{\mathrm{on}}), \qquad \widetilde{d}_{\mathrm{occ}}^{+} = 1 + \frac{I_{\mathrm{occ}}}{\log(1 + Q\kappa^2/\lambda_{\mathrm{on}})}. \tag{40}$$

C IMPORTED RESULTS

Theorem C.1 (Uniform finite-width event; Zhou–Li–Gu, Lemmas B.1, B.4–B.6)**.** *There are universal constants $C, c > 0$ such that the following holds. Fix $R > 0$ and $\epsilon_{\mathrm{ntk}} > 0$ before drawing θ_0 and let $\tau = R/\sqrt{m}$. Suppose*

$$\tau \geq Cm^{-3/2}L^{-3/2}\max\{(\log m)^{-3/2}, \{\log(4|\mathcal{U}|L^2/\delta_{\mathrm{lazy}})\}^{3/2}\},$$
$$\tau \leq c\min\{L^{-6}(\log m)^{-3/2}, L^{-9/2}(\log m)^{-3}\}, \tag{41}$$
$$m \geq CL^6|\mathcal{U}|^4\epsilon_{\mathrm{ntk}}^{-4}\log(4|\mathcal{U}|L/\delta_{\mathrm{lazy}}). \tag{42}$$

With probability at least $1 - \delta_{\mathrm{lazy}}$ over θ_0, for all $z, z' \in \mathcal{U}$ and all $\|u\|, \|v\| \leq R$,

$$\left\|\phi_{\theta_0+u/\sqrt{m}}(z)\right\| \leq G, \tag{43}$$
$$\left\|\phi_{\theta_0+u/\sqrt{m}}(z) - \phi_{\theta_0+v/\sqrt{m}}(z)\right\| \leq e_g, \tag{44}$$
$$\left|f(z;\theta_0+u/\sqrt{m}) - f(z;\theta_0+v/\sqrt{m}) - \phi_{\theta_0+v/\sqrt{m}}(z)^\top(u-v)\right| \leq e_f, \tag{45}$$
$$\left|\phi_{\theta_0}(z)^\top\phi_{\theta_0}(z') - k_{\mathrm{ntk}}(z,z')\right| \leq \epsilon_{\mathrm{ntk}}, \tag{46}$$

with

$$G = C\sqrt{L}, \qquad e_g = CGL^3\sqrt{\log(m|\mathcal{U}|/\delta_{\mathrm{lazy}})}\,R^{1/3}m^{-1/6}, \tag{47}$$
$$e_f = CL^3\sqrt{\log(m|\mathcal{U}|/\delta_{\mathrm{lazy}})}\,R^{4/3}m^{-1/6}. \tag{48}$$

Remark C.2 (Sources and scope). Each of Lemmas B.1, B.4, B.5, B.6 of Zhou et al. (2020) is applied with failure probability $\delta_{\mathrm{lazy}}/4$, followed by a union bound; Lemma B.1 is applied with accuracy $\epsilon_{\mathrm{ntk}}/|\mathcal{U}|$, so its Frobenius error is at most ϵ_{ntk}. The powers in (47) follow from $\tau = R/\sqrt{m}$ and $\phi = m^{-1/2}\nabla f$. The event is uniform over the ball, so it may be evaluated at the data-dependent b once b is shown to lie in the ball. Theorem 4.5 of Zhou et al. (2020) is not used.

Theorem C.3 (Self-normalized concentration; Abbasi-Yadkori et al., 2011, Thm. 1)**.** *Let $(\mathcal{G}_t)_{t\geq 0}$ be a filtration, $x_t \in \mathbb{R}^d$ $\mathcal{G}_{t-1}$-measurable, ε_t $\mathcal{G}_t$-measurable with $\mathbb{E}[e^{a\varepsilon_t} \mid \mathcal{G}_{t-1}] \leq e^{a^2\sigma^2/2}$, and $V_n = \lambda I + \sum_{t\leq n} x_t x_t^\top$. Then a.s.*

$$\Pr\Big(\exists n : \ \Big\|\sum_{t=1}^{n} x_t\varepsilon_t\Big\|_{V_n^{-1}} > \sigma\sqrt{2\log\frac{\det(V_n)^{1/2}}{\det(\lambda I)^{1/2}\delta}} \ \Big|\ \mathcal{G}_0\Big) \leq \delta. \tag{49}$$

The conditional form holds because the mixture supermartingale M_n of the original proof satisfies $\mathbb{E}[M_n \mid \mathcal{G}_0] \leq 1$.

Corollary C.4 (Fixed-design RKHS version)**.** *For fixed $x_1, \ldots, x_n$ in an RKHS,* (49) *holds at this n with $V_n = \lambda I + \sum_i x_i \otimes x_i$ and*

$$\frac{\det V_n}{\det(\lambda I)} := \det\big(I_n + \lambda^{-1}(\langle x_i, x_j\rangle)_{i,j\leq n}\big). \tag{50}$$

Proof. Restrict to $\mathrm{span}\{x_i\}$, apply Theorem C.3 in an orthonormal basis, and use $\det(\lambda I_p + XX^\top)/\lambda^p = \det(I_n + X^\top X/\lambda)$. □

Lemma C.5 (Elliptical potential; Abbasi-Yadkori et al., 2011, Lemma 11)**.** *If $V_t = V_{t-1} + x_t x_t^\top$ and $\|x_t\|^2_{V_{t-1}^{-1}} \leq 1$, then*

$$\sum_{t=1}^{T}\|x_t\|^2_{V_{t-1}^{-1}} \leq 2\log\frac{\det V_T}{\det V_0}. \tag{51}$$

D FORMAL STATEMENTS

Throughout, $u = \sqrt{m}(\theta - \theta_0)$ is the scaled displacement from the initialization and the *lazy ball* is $\{\|u\| \leq R\}$.

D.1 SOURCE PHASE

$$B_0 = \kappa B_{\mathrm{src}} + \sigma_{\mathrm{src}}\sqrt{2\log(2N/\delta_y)}, \tag{52}$$

$$\mathcal{E}_{y,\mathrm{src}} = \Big\{\max_{i\le N}|Y_i^0| \le B_0\Big\}, \qquad \Pr(\mathcal{E}_{y,\mathrm{src}}^c) \le \delta_y, \tag{53}$$

$$\Phi_0 = [\phi_{\theta_0}(Z_1^0), \dots, \phi_{\theta_0}(Z_N^0)], \tag{54}$$

$$\bar{u} = \big(\lambda_{\mathrm{src}} I + N^{-1}\Phi_0\Phi_0^\top\big)^{-1} N^{-1}\Phi_0 Y^0, \tag{55}$$

$$K_0 = \Phi_0^\top\Phi_0, \qquad k_0(z) = \Phi_0^\top\phi_{\theta_0}(z), \qquad k_0(z,z) = \|\phi_{\theta_0}(z)\|^2, \tag{56}$$

$$\epsilon_K \ge \|K_0 - K\|_{\mathrm{op}}, \qquad \epsilon_k \ge \|k_0(z) - k(z)\|, \qquad \epsilon_d \ge |k_0(z,z) - k_{\mathrm{ntk}}(z,z)|. \tag{57}$$

On the event of Theorem C.1, $\epsilon_K \le N\epsilon_{\mathrm{ntk}}$, $\epsilon_k \le \sqrt{N}\epsilon_{\mathrm{ntk}}$, and $\epsilon_d \le \epsilon_{\mathrm{ntk}}$ are admissible.

Theorem D.1 (Source TrainNN-to-NTK bridge)**.** *Work on the event of Theorem C.1 and on $\mathcal{E}_{y,\mathrm{src}}$. Let $u_{J_{\mathrm{src}}} = \sqrt{m}(b - \theta_0)$, $\gamma_{\mathrm{src}} = m\eta_{\mathrm{src}}$, and suppose*

$$0 < \gamma_{\mathrm{src}} \le (\lambda_{\mathrm{src}} + G^2)^{-1}, \qquad c_g := \frac{e_g G}{\lambda_{\mathrm{src}}} < 1, \tag{58}$$

$$R \ge \frac{2B_0/\sqrt{\lambda_{\mathrm{src}}} + (e_g B_0 + G e_f)/\lambda_{\mathrm{src}}}{1 - c_g}. \tag{59}$$

Then (a) every source iterate lies in the lazy ball and

$$\|u_{J_{\mathrm{src}}} - \bar{u}\| \le E_{\mathrm{src}} := (1 - \gamma_{\mathrm{src}}\lambda_{\mathrm{src}})^{J_{\mathrm{src}}}\frac{B_0}{\sqrt{\lambda_{\mathrm{src}}}} + \frac{e_g(GR + B_0) + G e_f}{\lambda_{\mathrm{src}}}. \tag{60}$$

(b) For every $z \in \mathcal{U}$, with

$$\zeta_\mu(z) = e_f + G E_{\mathrm{src}} + \sqrt{N}B_0\Big[\frac{\epsilon_k}{N\lambda_{\mathrm{src}}} + \frac{\|k(z)\|\,\epsilon_K}{(N\lambda_{\mathrm{src}})^2}\Big], \tag{61}$$

$$D_z = \frac{1}{\lambda_{\mathrm{src}}}\Big[\epsilon_d + \frac{\epsilon_k(2\,\|k(z)\| + \epsilon_k)}{N\lambda_{\mathrm{src}}} + \frac{\|k(z)\|^2\,\epsilon_K}{(N\lambda_{\mathrm{src}})^2}\Big], \tag{62}$$

$$|f_b(z) - \widehat{r}_{N,\lambda_{\mathrm{src}}}(z)| \le \zeta_\mu(z). \tag{63}$$

(c) With

$$G_0 = \lambda_{\mathrm{src}} I + N^{-1}\Phi_0\Phi_0^\top, \qquad s_0(z) = \|\phi_{\theta_0}(z)\|_{G_0^{-1}}, \qquad \rho_A = \frac{2Ge_g + e_g^2}{\lambda_{\mathrm{src}}}, \tag{64}$$

for every $z \in \mathcal{U}$,

$$\widehat{\sigma}_N(z) \le \sqrt{1+\rho_A}\, s_{\mathrm{src}}(z) + \frac{e_g}{\sqrt{\lambda_{\mathrm{src}}}} + \sqrt{D_z}, \tag{65}$$

$$s_{\mathrm{src}}(z) \le \sqrt{1+\rho_A}\Big[\widehat{\sigma}_N(z) + \sqrt{D_z} + \frac{e_g}{\sqrt{\lambda_{\mathrm{src}}}}\Big]. \tag{66}$$

Corollary D.2 (Two initializations)**.** *Fix one source sample in $\mathcal{E}_{y,\mathrm{src}}$ and draw two independent initializations $\theta_0^{(1)}, \theta_0^{(2)}$ satisfying the hypotheses of Theorem C.1 with common $\mathcal{U}, R, m, \epsilon_{\mathrm{ntk}}$ and $\delta_{\mathrm{lazy}} < 1/2$; train both source networks with the same data and hyperparameters, and suppose the conditions of Theorem D.1 hold in both runs. Use $\epsilon_K = N\epsilon_{\mathrm{ntk}}$, $\epsilon_k = \sqrt{N}\epsilon_{\mathrm{ntk}}$, $\epsilon_d = \epsilon_{\mathrm{ntk}}$ and*

$$a_\sigma = \sqrt{1+\rho_A}, \qquad b_{\sigma,z} = \frac{e_g}{\sqrt{\lambda_{\mathrm{src}}}} + \sqrt{D_z},$$

$$\zeta_\sigma(z) = \max\Big\{(a_\sigma - 1)\frac{G}{\sqrt{\lambda_{\mathrm{src}}}} + b_{\sigma,z},\ (a_\sigma - 1)\frac{\kappa}{\sqrt{\lambda_{\mathrm{src}}}} + a_\sigma b_{\sigma,z}\Big\}. \tag{67}$$

With probability at least $1 - 2\delta_{\mathrm{lazy}}$, for every $z \in \mathcal{U}$,

$$|f(z; b^{(1)}) - f(z; b^{(2)})| \le 2\zeta_\mu(z), \tag{68}$$

$$|s_{\mathrm{src}}^{(1)}(z) - s_{\mathrm{src}}^{(2)}(z)| \le 2\zeta_\sigma(z). \tag{69}$$

D.2 Residual representation on the candidate set

$$\alpha_N = \frac{\sigma_{\mathrm{src}}}{\sqrt{N}}\sqrt{2\big(I_N(\lambda_{\mathrm{src}}) + \log(1/\delta_{\mathrm{src}})\big)} + \sqrt{\lambda_{\mathrm{src}}}B_{\mathrm{src}}, \tag{70}$$

$$\zeta_\mu = \sup_{z\in\mathcal{U}} \zeta_\mu(z), \tag{71}$$

$$\Phi_{\mathrm{src},*} = [\phi_b(\tilde{z}_1), \dots, \phi_b(\tilde{z}_M)], \qquad K_{\mathrm{src},*} = \Phi_{\mathrm{src},*}^\top \Phi_{\mathrm{src},*}, \tag{72}$$

$$\bar{d}_{\mathrm{src}} = M(\epsilon_{\mathrm{ntk}} + 2Ge_g + e_g^2), \qquad \bar{a} = \bar{d}_{\mathrm{src}}/\lambda_0, \tag{73}$$

$$\mathcal{E}_{\mathrm{sn,src}} = \Big\{ \Big\| \frac{1}{N}\sum_{i=1}^N k_{Z_i^0}\xi_i^0 \Big\|_{\widehat{\Sigma}_{N,\lambda_{\mathrm{src}}}^{-1}} \le \frac{\sigma_{\mathrm{src}}}{\sqrt{N}}\sqrt{2\big(I_N(\lambda_{\mathrm{src}}) + \log(1/\delta_{\mathrm{src}})\big)} \Big\}. \tag{74}$$

By Corollary C.4 with $x_i = k_{Z_i^0}/\sqrt{N}$ and $\varepsilon_i = \xi_i^0/\sqrt{N}$, $\Pr(\mathcal{E}_{\mathrm{sn,src}}^c \mid Z_{1:N}^0) \le \delta_{\mathrm{src}}$.

Theorem D.3 (Exact residual tangent representation)**.** *Suppose Assumptions B.1–B.2, the conclusions of Theorem D.1, and $\bar{a} < 1$. On $\mathcal{E}_{\mathrm{sn,src}} \cap \mathcal{E}_{y,\mathrm{src}}$ and the event of Theorem C.1, there is u_* with*

$$r_{\mathrm{tar}}(z) - f_b(z) = \phi_b(z)^\top u_* \qquad (z \in \mathcal{Z}_*), \tag{75}$$

$$\|u_*\| \le S_* := (1-\bar{a})^{-1/2}\Big[\frac{\alpha_N + \epsilon}{\sqrt{\lambda_{\mathrm{src}}}} + \sqrt{\frac{M}{\lambda_0}}\,\zeta_\mu\Big], \tag{76}$$

and u_* is $\mathcal{S}_{\mathrm{src}}$-measurable.

Cold-start comparison. With

$$\Phi_{0,*} = [\phi_{\theta_0}(\tilde{z}_1), \dots, \phi_{\theta_0}(\tilde{z}_M)], \qquad K_{0,*} = \Phi_{0,*}^\top \Phi_{0,*}, \tag{77}$$

$$h = \big(r_{\mathrm{tar}}(\tilde{z}_i)\big)_{i\le M}, \qquad g = \big(f_b(\tilde{z}_i)\big)_{i\le M}, \qquad q_* = h - g, \tag{78}$$

$$u_* = \Phi_{\mathrm{src},*}K_{\mathrm{src},*}^{-1}q_*, \qquad u_{\mathrm{tar}} = \Phi_{0,*}K_{0,*}^{-1}h, \tag{79}$$

u_{tar} is the minimum-norm u with $\phi_{\theta_0}(z)^\top u = r_{\mathrm{tar}}(z)$ on $\mathcal{Z}_*$, the quantity that replaces S_* for a run centered at θ_0.

Corollary D.4 (Residual versus target complexity)**.** *Under Theorem D.3, let $h, g \neq 0$,*

$$\langle a, b\rangle_* = a^\top H_*^{-1} b, \qquad \|a\|_* = \langle a, a\rangle_*^{1/2}, \qquad \bar{a}_0 = \frac{M\epsilon_{\mathrm{ntk}}}{\lambda_0} \le \bar{a},$$
$$\tau = \frac{\|g\|_*}{\|h\|_*}, \qquad \rho = \frac{\langle g, h\rangle_*}{\|g\|_*\,\|h\|_*}. \tag{80}$$

Then

(i) $\|q_*\|_* / \|h\|_* = \sqrt{1 - 2\rho\tau + \tau^2}$*;*

(ii) $\|q_*\|_* < \|h\|_* \iff \tau < 2\rho$*;*

(iii) $\min_{c\in\mathbb{R}} \|h - cg\|_* / \|h\|_* = \sqrt{1-\rho^2}$*, attained at* $c = \rho\,\|h\|_* / \|g\|_*$*;*

(iv)

$$\frac{\|u_*\|}{\|u_{\mathrm{tar}}\|} \le \sqrt{\frac{1+\bar{a}_0}{1-\bar{a}}}\,\sqrt{1 - 2\rho\tau + \tau^2}. \tag{81}$$

Remark D.5 (The certified bound does not see the gain). Since $\alpha_N \ge \sqrt{\lambda_{\mathrm{src}}}B_{\mathrm{src}}$,

$$S_* \ge \frac{\alpha_N + \epsilon}{\sqrt{\lambda_{\mathrm{src}}}} \ge B_{\mathrm{src}}, \tag{82}$$

so $S_* < \|u_{\mathrm{tar}}\|$ requires $\|h\|_* > \sqrt{(1+\bar{a}_0)/(1-\bar{a})}\,B_{\mathrm{src}}$. The loss comes from $\widehat{\Sigma}_{N,\lambda_{\mathrm{src}}} \succeq \lambda_{\mathrm{src}} I$ in (130).

D.3 Online phase

Fix $b_{\text{slack}} > 0$ and set, for $n \le T$,

$$B_y = GS_* + \nu\sqrt{2\log(2T/\delta_{y,\text{on}})} + b_{\text{slack}}, \tag{83}$$

$$R_1 = 7B_y\sqrt{T/\lambda_{\text{on}}}, \qquad R_{\text{out}} = 4B_y\sqrt{T/\lambda_{\text{on}}}, \tag{84}$$

$$A_n = \lambda_{\text{on}} I + \sum_{s=1}^{n} \phi_b(z_s)\phi_b(z_s)^\top, \tag{85}$$

$$\bar{u}_n = A_n^{-1} \sum_{s=1}^{n} \phi_b(z_s)\big(Y_s - f_b(z_s)\big), \tag{86}$$

$$u_n = \sqrt{m}(\widehat{\theta}^{\text{on}}_{n+1} - b), \qquad u_0 = 0, \tag{87}$$

$$\chi_n(z) = \phi_{b+u_n/\sqrt{m}}(z), \tag{88}$$

$$\psi_s = \chi_{s-1}(z_s), \qquad V_n = \lambda_{\text{on}} I + \sum_{s=1}^{n} \psi_s\psi_s^\top, \tag{89}$$

$$s_n(z) = \|\chi_n(z)\|_{V_n^{-1}}, \tag{90}$$

$$D_n = n\big\{e_g(GR_1 + e_f + B_y) + Ge_f\big\}, \tag{91}$$

$$E_n = q^{J_{\text{on}}}\big(R_{\text{out}} + B_y\sqrt{n/\lambda_{\text{on}}}\big) + D_n/\lambda_{\text{on}}, \tag{92}$$

$$q = 1 - \gamma_{\text{on}}\lambda_{\text{on}}, \qquad \gamma_{\text{on}} = m\eta_{\text{on}}. \tag{93}$$

Theorem D.6 (Source-centered warm-start NeuralUCB confidence)**.** *Work on the events of Theorems D.1–D.3. Let*

$$\mathcal{E}_{y,\text{on}} = \Big\{\max_{t\le T}|\eta_t| \le \nu\sqrt{2\log(2T/\delta_{y,\text{on}})}\Big\}, \qquad \Pr(\mathcal{E}^c_{y,\text{on}} \mid \mathcal{S}_{\text{src}}) \le \delta_{y,\text{on}}. \tag{94}$$

Suppose

$$R \ge \frac{2B_0/\sqrt{\lambda_{\text{src}}} + R_1 + (e_gB_0 + Ge_f)/\lambda_{\text{src}}}{1 - c_g}, \tag{95}$$

$$0 < \gamma_{\text{on}} \le (\lambda_{\text{on}} + TG^2)^{-1}, \qquad q^{J_{\text{on}}} \le 1/8, \tag{96}$$

$$D_T/\lambda_{\text{on}} \le B_y\sqrt{T/\lambda_{\text{on}}}, \qquad c_V := \frac{T(2Ge_g + e_g^2)}{\lambda_{\text{on}}} < 1. \tag{97}$$

On $\mathcal{E}_{y,\text{on}}$: every inner iterate lies in the lazy ball, $\|u_n\| \le R_{\text{out}}$,

$$\|u_n - \bar{u}_n\| \le E_n, \tag{98}$$

$$(1 - c_V)A_n \preceq V_n \preceq (1 + c_V)A_n. \tag{99}$$

There is an event $\mathcal{E}_{\text{sn,on}}$ with $\Pr(\mathcal{E}^c_{\text{sn,on}} \mid \mathcal{S}_{\text{src}}) \le \delta_{\text{on}}$ such that on $\mathcal{E}_{y,\text{on}} \cap \mathcal{E}_{\text{sn,on}}$, for all $0 \le n < T$ and $z \in \mathcal{Z}_*$,

$$|r_{\text{tar}}(z) - f(z; b + u_n/\sqrt{m})| \le \beta_n s_n(z) + e_{\text{on}}, \tag{100}$$

$$a_n = \nu\sqrt{2\log\frac{\det(A_n)^{1/2}}{\det(\lambda_{\text{on}}I)^{1/2}\delta_{\text{on}}}} + \sqrt{\lambda_{\text{on}}}S_*, \tag{101}$$

$$\beta_n = \sqrt{1 + c_V}\big[a_n + \sqrt{\lambda_{\text{on}} + nG^2}\,E_n\big], \tag{102}$$

$$e_{\text{on}} = e_gS_* + e_f. \tag{103}$$

D.4 Regret

$$a_\sigma = \sqrt{1 + \rho_A}, \qquad b_\sigma = \sup_{z\in\mathcal{U}}\Big(\frac{e_g}{\sqrt{\lambda_{\text{src}}}} + \sqrt{D_z}\Big), \tag{104}$$

$$\rho_{\text{src}} = a_\sigma(\alpha_N + \epsilon), \qquad \Delta_{\text{src}} = (\alpha_N + \epsilon)b_\sigma + \zeta_\mu. \tag{105}$$

Theorem D.7 (Offline-to-online NeuralUCB regret)**.** *Suppose Assumptions B.1–B.2, the hypotheses of Theorem C.1, the deterministic conditions of Theorems D.1, D.3, and D.6, and $\lambda_{\mathrm{on}} \geq G^2$. With probability at least*

$$1 - \delta_{\mathrm{tot}}, \qquad \delta_{\mathrm{tot}} = \delta_{\mathrm{lazy}} + \delta_y + \delta_{\mathrm{src}} + \delta_{y,\mathrm{on}} + \delta_{\mathrm{on}}, \tag{106}$$

$$R_T \leq 2 \sum_{t=1}^{T} \min\bigl\{\beta_{t-1} s_{t-1}(z_t) + e_{\mathrm{on}},\ \rho_{\mathrm{src}} s_{\mathrm{src}}(z_t) + \Delta_{\mathrm{src}}\bigr\}, \tag{107}$$

$$R_T \leq 2 \min\Bigl\{\beta_T \sqrt{2T\Gamma_T} + T e_{\mathrm{on}},\ \rho_{\mathrm{src}} \sum_{t=1}^{T} s_{\mathrm{src}}(z_t) + T\Delta_{\mathrm{src}}\Bigr\}, \qquad \beta_T = \max_{n<T} \beta_n. \tag{108}$$

Corollary D.8 (Occurrence-level rate)**.** *Suppose*

$$\delta_{\mathrm{ker}} := \epsilon_{\mathrm{ntk}} + 2Ge_g + e_g^2 \leq \frac{\lambda_{\mathrm{on}}}{T}\bigl(e^{1/T} - 1\bigr). \tag{109}$$

Then for every $n \leq T$

$$\log \frac{\det A_n}{\det(\lambda_{\mathrm{on}} I)} \leq I_{\mathrm{occ}} + 1, \qquad \Gamma_T \leq I_{\mathrm{occ}} + 1. \tag{110}$$

If moreover, for a constant $C_{\mathrm{on}} > 0$,

$$q^{J_{\mathrm{on}}}\bigl(R_{\mathrm{out}} + B_y\sqrt{T/\lambda_{\mathrm{on}}}\bigr) \leq \frac{1}{2\sqrt{\lambda_{\mathrm{on}} + TG^2}}, \qquad \frac{D_T}{\lambda_{\mathrm{on}}} \leq \frac{1}{2\sqrt{\lambda_{\mathrm{on}} + TG^2}}, \qquad e_{\mathrm{on}} \leq C_{\mathrm{on}} T^{-1/2}, \tag{111}$$

then

$$\mathfrak{R}_T^{\mathrm{on}} := 2\bigl[\beta_T \sqrt{2T\Gamma_T} + T e_{\mathrm{on}}\bigr] \tag{112}$$

satisfies

$$\mathfrak{R}_T^{\mathrm{on}} = O\Bigl(\bigl[\nu\sqrt{I_{\mathrm{occ}} + 2\log(1/\delta_{\mathrm{on}}) + 1} + \sqrt{\lambda_{\mathrm{on}}} S_* + 1\bigr]\sqrt{T(I_{\mathrm{occ}} + 1)} + \sqrt{T}\Bigr), \tag{113}$$

and, with

$$\widetilde{d}_{\mathrm{occ}} = \frac{I_{\mathrm{occ}}}{\log(1 + Q\kappa^2/\lambda_{\mathrm{on}})}, \qquad \widetilde{d}_{\mathrm{occ}}^{+} = 1 + \widetilde{d}_{\mathrm{occ}}, \tag{114}$$

$\nu = O(1)$, $\log(1/\delta_{\mathrm{on}}) = O(\log T)$, *and* $\lambda_{\mathrm{on}} S_*^2 = O(\widetilde{d}_{\mathrm{occ}}^{+})$,

$$\mathfrak{R}_T^{\mathrm{on}} = \widetilde{O}(\widetilde{d}_{\mathrm{occ}}^{+}\sqrt{T}). \tag{115}$$

E Proof of Theorem D.1 and Corollary D.2

E.1 Part (a): containment and optimization error

Write $\gamma = \gamma_{\mathrm{src}}$, $\lambda = \lambda_{\mathrm{src}}$, and for $u \in \mathbb{R}^p$

$$F(u) = \bigl(f(Z_i^0; \theta_0 + u/\sqrt{m})\bigr)_{i \leq N}, \qquad J(u) = \bigl[\phi_{\theta_0 + u/\sqrt{m}}(Z_1^0), \ldots, \phi_{\theta_0 + u/\sqrt{m}}(Z_N^0)\bigr].$$

Gradient descent on (23) in the coordinate $u = \sqrt{m}(\theta - \theta_0)$ reads

$$u_{j+1} = u_j - \gamma\bigl[N^{-1} J(u_j)\bigl(F(u_j) - Y^0\bigr) + \lambda u_j\bigr], \qquad u_0 = 0. \tag{116}$$

For $\|u\| \leq R$, Theorem C.1 and $f(\cdot; \theta_0) = 0$ give

$$F(u)_i = \int_0^1 \phi_{\theta_0 + tu/\sqrt{m}}(Z_i^0)^\top u \, dt, \qquad \|F(u)\| \leq \sqrt{N}\, GR, \qquad \bigl\|F(u) - \Phi_0^\top u\bigr\| \leq \sqrt{N}\, e_f,$$
$$\|J(u) - \Phi_0\|_{\mathrm{op}} \leq \sqrt{N}\, e_g, \qquad \|\Phi_0\|_{\mathrm{op}} \leq \sqrt{N}\, G, \qquad \bigl\|Y^0\bigr\| \leq \sqrt{N}\, B_0.$$

Hence, with $r(u) := N^{-1} J(u)(F(u) - Y^0) - N^{-1}\Phi_0(\Phi_0^\top u - Y^0) = N^{-1}(J(u) - \Phi_0)(F(u) - Y^0) + N^{-1}\Phi_0(F(u) - \Phi_0^\top u)$,

$$\|r(u)\| \leq e_g(GR + B_0) + Ge_f =: \zeta_{\mathrm{grad}} \qquad (\|u\| \leq R). \tag{117}$$

By (55), $G_0\bar{u} = N^{-1}\Phi_0 Y^0$ and

$$\lambda I \preceq G_0 \preceq (\lambda + G^2)I, \qquad \|I - \gamma G_0\|_{\mathrm{op}} \le 1 - \gamma\lambda, \qquad \|\bar{u}\| \le \frac{\|Y^0\|}{\sqrt{N\lambda}} \le \frac{B_0}{\sqrt{\lambda}}. \tag{118}$$

Subtracting the normal equation from (116),

$$u_{j+1} - \bar{u} = (I - \gamma G_0)(u_j - \bar{u}) - \gamma r(u_j). \tag{119}$$

Claim $C(j)$: $\|u_i\| \le R$ for $i \le j$, and

$$\|u_j - \bar{u}\| \le (1 - \gamma\lambda)^j \|\bar{u}\| + \zeta_{\mathrm{grad}}/\lambda. \tag{120}$$

$C(0)$ holds since $u_0 = 0$. If $C(j)$ holds, then (117) applies at u_j, and (119) gives

$$\|u_{j+1} - \bar{u}\| \le (1 - \gamma\lambda)^{j+1}\|\bar{u}\| + (1 - \gamma\lambda)\frac{\zeta_{\mathrm{grad}}}{\lambda} + \gamma\zeta_{\mathrm{grad}} = (1 - \gamma\lambda)^{j+1}\|\bar{u}\| + \frac{\zeta_{\mathrm{grad}}}{\lambda},$$

$$\|u_{j+1}\| \le 2\|\bar{u}\| + \frac{\zeta_{\mathrm{grad}}}{\lambda} \le \frac{2B_0}{\sqrt{\lambda}} + \frac{e_g(GR + B_0) + Ge_f}{\lambda} \le R, \tag{121}$$

where the last inequality is (59). So $C(j+1)$ holds. $C(J_{\mathrm{src}})$ with (118) is (60). □

E.2 Part (b): mean

Let $R_0 = (K_0 + N\lambda I)^{-1}$ and $R_K = (K + N\lambda I)^{-1}$, so $\|R_0\|, \|R_K\| \le (N\lambda)^{-1}$ and $R_0 - R_K = R_0(K - K_0)R_K$. By the push-through identity $(\lambda I + N^{-1}\Phi_0\Phi_0^\top)^{-1}N^{-1}\Phi_0 = \Phi_0 R_0$,

$$\phi_{\theta_0}(z)^\top \bar{u} = k_0(z)^\top R_0 Y^0 =: \widehat{r}_0(z). \tag{122}$$

$$|\widehat{r}_0(z) - \widehat{r}_{N,\lambda}(z)| = \left|(k_0 - k)^\top R_0 Y^0 + k^\top R_0(K - K_0)R_K Y^0\right| \le \|Y^0\|\left[\frac{\epsilon_k}{N\lambda} + \frac{\|k(z)\|\,\epsilon_K}{(N\lambda)^2}\right]. \tag{123}$$

With (45) at $(u, v) = (u_{J_{\mathrm{src}}}, 0)$,

$$|f_b(z) - \widehat{r}_{N,\lambda}(z)| \le \underbrace{|f_b(z) - \phi_{\theta_0}(z)^\top u_{J_{\mathrm{src}}}|}_{\le e_f} + \underbrace{|\phi_{\theta_0}(z)^\top(u_{J_{\mathrm{src}}} - \bar{u})|}_{\le GE_{\mathrm{src}}} + |\widehat{r}_0(z) - \widehat{r}_{N,\lambda}(z)| \le \zeta_\mu(z).$$

□

E.3 Part (c): width

By Woodbury,

$$s_0(z)^2 = \lambda^{-1}\left[k_0(z,z) - k_0(z)^\top R_0 k_0(z)\right]. \tag{124}$$

With (38), $R_0 = R_0^\top$, and $\|k_0\| \le \|k\| + \epsilon_k$,

$$\begin{aligned}\lambda|s_0^2 - \widehat{\sigma}_N^2| &\le |k_0(z,z) - k_{\mathrm{ntk}}(z,z)| + |(k_0 - k)^\top R_0(k_0 + k)| + |k^\top(R_0 - R_K)k| \\ &\le \epsilon_d + \frac{\epsilon_k(2\|k\| + \epsilon_k)}{N\lambda} + \frac{\|k\|^2\epsilon_K}{(N\lambda)^2} = \lambda D_z, \\ |s_0 - \widehat{\sigma}_N| &\le \sqrt{|s_0^2 - \widehat{\sigma}_N^2|} \le \sqrt{D_z}.\end{aligned} \tag{125}$$

By (a), b lies in the lazy ball, so $\|\phi_b(z) - \phi_{\theta_0}(z)\| \le e_g$ and

$$\|G_{\mathrm{src}} - G_0\|_{\mathrm{op}} \le \max_i \left\|\phi_b(Z_i^0) - \phi_{\theta_0}(Z_i^0)\right\|\left(\left\|\phi_b(Z_i^0)\right\| + \left\|\phi_{\theta_0}(Z_i^0)\right\|\right) \le 2Ge_g \le \rho_A\lambda, \tag{126}$$

$$G_{\mathrm{src}} \preceq G_0 + \rho_A\lambda I \preceq (1 + \rho_A)G_0, \qquad G_0 \preceq (1 + \rho_A)G_{\mathrm{src}}. \tag{127}$$

Therefore

$$\widehat{\sigma}_N(z) \le s_0(z) + \sqrt{D_z} \le \|\phi_b(z)\|_{G_0^{-1}} + \frac{e_g}{\sqrt{\lambda}} + \sqrt{D_z} \le \sqrt{1 + \rho_A}\, s_{\mathrm{src}}(z) + \frac{e_g}{\sqrt{\lambda}} + \sqrt{D_z},$$

$$s_{\mathrm{src}}(z) \le \|\phi_{\theta_0}(z)\|_{G_{\mathrm{src}}^{-1}} + \frac{e_g}{\sqrt{\lambda}} \le \sqrt{1 + \rho_A}\, s_0(z) + \frac{e_g}{\sqrt{\lambda}} \le \sqrt{1 + \rho_A}\left[\widehat{\sigma}_N(z) + \sqrt{D_z} + \frac{e_g}{\sqrt{\lambda}}\right].$$

□

E.4 Proof of Corollary D.2

On the intersection of the two events of Theorem C.1 (probability $\geq 1 - 2\delta_{\text{lazy}}$), (63) in each run gives $|f(z; b^{(1)}) - f(z; b^{(2)})| \leq \sum_{r=1,2} |f(z; b^{(r)}) - \widehat{r}_{N,\lambda}(z)| \leq 2\zeta_\mu(z)$. In each run $s_{\text{src}} \leq G/\sqrt{\lambda}$, $\widehat{\sigma}_N \leq \kappa/\sqrt{\lambda}$, and (65)–(66) give

$$\widehat{\sigma}_N - s_{\text{src}} \leq (a_\sigma - 1)s_{\text{src}} + b_{\sigma,z} \leq (a_\sigma - 1)\tfrac{G}{\sqrt{\lambda}} + b_{\sigma,z},$$
$$s_{\text{src}} - \widehat{\sigma}_N \leq (a_\sigma - 1)\widehat{\sigma}_N + a_\sigma b_{\sigma,z} \leq (a_\sigma - 1)\tfrac{\kappa}{\sqrt{\lambda}} + a_\sigma b_{\sigma,z},$$

so $|s_{\text{src}}^{(r)} - \widehat{\sigma}_N| \leq \zeta_\sigma(z)$ and (69) follows. □

F Proof of Theorem D.3 and Corollary D.4

F.1 Theorem D.3

Write $\lambda = \lambda_{\text{src}}$ and $\widehat{\Sigma} = \widehat{\Sigma}_{N,\lambda}$.

Step 1 (ghost estimator). The normal equation $\widehat{\Sigma}\,\widehat{r}_{N,\lambda} = N^{-1}\sum_i k_{Z_i^0} Y_i^0$ and $N^{-1}\sum_i k_{Z_i^0} r_{\text{src}}(Z_i^0) = (\widehat{\Sigma} - \lambda I) r_{\text{src}}$ give

$$\widehat{r}_{N,\lambda} - r_{\text{src}} = \widehat{\Sigma}^{-1}\Big[\frac{1}{N}\sum_{i=1}^{N} k_{Z_i^0}\xi_i^0 - \lambda r_{\text{src}}\Big]. \tag{128}$$

On $\mathcal{E}_{\text{sn,src}}$,

$$\|\widehat{r}_{N,\lambda} - r_{\text{src}}\|_{\widehat{\Sigma}} \leq \Big\|\frac{1}{N}\sum_i k_{Z_i^0}\xi_i^0\Big\|_{\widehat{\Sigma}^{-1}} + \lambda\Big\|\widehat{\Sigma}^{-1/2} r_{\text{src}}\Big\|_{\mathcal{H}_{\text{ntk}}} \leq \alpha_N, \tag{129}$$

using $\lambda\Big\|\widehat{\Sigma}^{-1/2} r_{\text{src}}\Big\|_{\mathcal{H}_{\text{ntk}}} \leq \sqrt{\lambda} B_{\text{src}}$. With (32) and $\widehat{\Sigma} \succeq \lambda I$,

$$\|r_{\text{tar}} - \widehat{r}_{N,\lambda}\|_{\widehat{\Sigma}} \leq \alpha_N + \epsilon, \qquad \|r_{\text{tar}} - \widehat{r}_{N,\lambda}\|_{\mathcal{H}_{\text{ntk}}} \leq \frac{\alpha_N + \epsilon}{\sqrt{\lambda}}. \tag{130}$$

Step 2 (candidate Gram). By Theorem D.1(a), b lies in the lazy ball; with (46),

$$|\phi_b(\tilde{z}_i)^\top \phi_b(\tilde{z}_j) - k_{\text{ntk}}(\tilde{z}_i, \tilde{z}_j)| \leq \epsilon_{\text{ntk}} + 2Ge_g + e_g^2, \qquad \|K_{\text{src},*} - H_*\|_{\text{op}} \leq \|K_{\text{src},*} - H_*\|_F \leq \bar{d}_{\text{src}}.$$

Since $H_* \succeq \lambda_0 I$, $\bar{d}_{\text{src}} I \preceq \bar{a} H_*$ and

$$(1 - \bar{a})H_* \preceq K_{\text{src},*} \preceq (1 + \bar{a})H_*. \tag{131}$$

Step 3 (representer). Let $h_{\text{KRR}} = (r_{\text{tar}} - \widehat{r}_{N,\lambda})|_{\mathcal{Z}_*}$ and $e = (\widehat{r}_{N,\lambda} - f_b)|_{\mathcal{Z}_*}$, so $q_* = h - g = h_{\text{KRR}} + e$. For $\varphi \in \mathcal{H}_{\text{ntk}}$, $(\varphi|_{\mathcal{Z}_*})^\top H_*^{-1} \varphi|_{\mathcal{Z}_*} \leq \|\varphi\|_{\mathcal{H}_{\text{ntk}}}^2$ (minimum-norm interpolation). With (130) and (63),

$$\Big\|H_*^{-1/2} h_{\text{KRR}}\Big\| \leq \frac{\alpha_N + \epsilon}{\sqrt{\lambda}}, \qquad \Big\|H_*^{-1/2} e\Big\| \leq \lambda_0^{-1/2}\,\|e\| \leq \sqrt{\frac{M}{\lambda_0}}\,\zeta_\mu. \tag{132}$$

Set

$$u_* = \Phi_{\text{src},*} K_{\text{src},*}^{-1} q_*. \tag{133}$$

Then $\Phi_{\text{src},*}^\top u_* = q_*$, which is (75), and by (131)–(132)

$$\|u_*\|^2 = q_*^\top K_{\text{src},*}^{-1} q_* \leq (1 - \bar{a})^{-1}\Big(\Big\|H_*^{-1/2} h_{\text{KRR}}\Big\| + \Big\|H_*^{-1/2} e\Big\|\Big)^2 \leq S_*^2. \tag{134}$$

Step 4 (measurability). b is a function of $(\theta_0, \mathcal{D}_0)$; $\mathcal{Z}_*$ is a function of the candidate list; $\Phi_{\text{src},*}$, $K_{\text{src},*}$, and $q_* = (r_{\text{tar}} - f_b)|_{\mathcal{Z}_*}$ are functions of $(b, \mathcal{Z}_*)$ and the fixed r_{tar}. By (33), $u_* \in \mathcal{S}_{\text{src}}$. □

F.2 Corollary D.4 and Remark D.5

$\langle \cdot, \cdot \rangle_*$ is an inner product since $H_* \succ 0$.

(i) $\|h - g\|_*^2 = \|h\|_*^2 - 2\langle h, g\rangle_* + \|g\|_*^2 = \|h\|_*^2 (1 - 2\rho\tau + \tau^2)$.

(ii) By (i): $\|q_*\|_* < \|h\|_* \iff \tau(\tau - 2\rho) < 0 \iff \tau < 2\rho$, as $\tau > 0$.

(iii) $\|h - cg\|_*^2 = \|h\|_*^2 - 2c\langle h, g\rangle_* + c^2 \|g\|_*^2$ is minimized at $c = \langle h, g\rangle_* / \|g\|_*^2 = \rho \|h\|_* / \|g\|_*$ with value $\|h\|_*^2 (1 - \rho^2)$.

(iv) By (133) and (131), $\|u_*\|^2 = q_*^\top K_{\mathrm{src},*}^{-1} q_* \le (1 - \bar{a})^{-1} \|q_*\|_*^2$. By (46), $\|K_{0,*} - H_*\|_{\mathrm{op}} \le M\epsilon_{\mathrm{ntk}} = \bar{a}_0 \lambda_0$, so

$$(1 - \bar{a}_0) H_* \preceq K_{0,*} \preceq (1 + \bar{a}_0) H_*, \qquad \frac{\|h\|_*^2}{1 + \bar{a}_0} \le \|u_{\mathrm{tar}}\|^2 = h^\top K_{0,*}^{-1} h \le \frac{\|h\|_*^2}{1 - \bar{a}_0}. \tag{135}$$

Combining with (i) gives (81). $\bar{a}_0 \le \bar{a}$ because $\bar{d}_{\mathrm{src}} \ge M\epsilon_{\mathrm{ntk}}$.

For Remark D.5: $S_* \ge (\alpha_N + \epsilon)/\sqrt{\lambda_{\mathrm{src}}} \ge \alpha_N / \sqrt{\lambda_{\mathrm{src}}} \ge B_{\mathrm{src}}$, and by (135), $S_* < \|u_{\mathrm{tar}}\|$ implies $B_{\mathrm{src}} \le S_* < (1 - \bar{a}_0)^{-1/2} \|h\|_*$. □

G Proof of Theorem D.6

Condition on $\mathcal{S}_{\mathrm{src}}$ throughout. Write $\lambda = \lambda_{\mathrm{on}}$, $\gamma = \gamma_{\mathrm{on}}$, and $B = B_y \sqrt{T/\lambda}$, so $R_1 = 7B$ and $R_{\mathrm{out}} = 4B$.

G.1 Residual labels

Every $z_s \in \mathcal{Z}_*$, so by (75)

$$r_s := Y_s - f_b(z_s) = \phi_b(z_s)^\top u_* + \eta_s. \tag{136}$$

By (34), $\Pr(|\eta_t| > x \mid \mathcal{F}_{t-1}) \le 2e^{-x^2/(2\nu^2)}$; with $x = \nu\sqrt{2\log(2T/\delta_{y,\mathrm{on}})}$ and a union bound over $t \le T$, $\Pr(\mathcal{E}_{y,\mathrm{on}}^c \mid \mathcal{S}_{\mathrm{src}}) \le \delta_{y,\mathrm{on}}$. On $\mathcal{E}_{y,\mathrm{on}}$,

$$|r_s| \le G S_* + \nu\sqrt{2\log(2T/\delta_{y,\mathrm{on}})} \le B_y. \tag{137}$$

G.2 One lazy ball for both phases

For $\|u\| \le R_1$, $b + u/\sqrt{m} = \theta_0 + (u_{J_{\mathrm{src}}} + u)/\sqrt{m}$ and, by (121) and (95),

$$\|u_{J_{\mathrm{src}}} + u\| \le \frac{2B_0}{\sqrt{\lambda_{\mathrm{src}}}} + \frac{e_g(GR + B_0) + Ge_f}{\lambda_{\mathrm{src}}} + R_1 \le R.$$

Hence (43)–(45) hold at $b + u/\sqrt{m}$ for all $\|u\| \le R_1$. (95) implies (59).

G.3 Containment and optimization error

Fix n. Let $\Phi_{b,n} = [\phi_b(z_1), \dots, \phi_b(z_n)]$, $J_n(u) = [\phi_{b+u/\sqrt{m}}(z_s)]_{s \le n}$, $F_n(u) = (f(z_s; b + u/\sqrt{m}) - f_b(z_s))_{s \le n}$, $r = (r_s)_{s \le n}$. The inner steps on (26) in the coordinate $u = \sqrt{m}(\theta - b)$ are $v_{j+1} = v_j - \gamma g_n(v_j)$ with $v_0 = u_{n-1}$ and $v_{J_{\mathrm{on}}} = u_n$, where

$$\begin{aligned} g_n(u) &= J_n(u)\big(F_n(u) - r\big) + \lambda u, \\ \bar{g}_n(u) &= \Phi_{b,n}(\Phi_{b,n}^\top u - r) + \lambda u = A_n(u - \bar{u}_n). \end{aligned} \tag{138}$$

For $\|u\| \le R_1$ on $\mathcal{E}_{y,\mathrm{on}}$,

$$\|J_n(u) - \Phi_{b,n}\|_{\mathrm{op}} \le \sqrt{n}\, e_g, \qquad \|\Phi_{b,n}\|_{\mathrm{op}} \le \sqrt{n}\, G,$$

$$\big\|F_n(u) - \Phi_{b,n}^\top u\big\| \le \sqrt{n}\, e_f, \qquad \|F_n(u) - r\| \le \sqrt{n}\,(GR_1 + e_f + B_y), \tag{139}$$

$$\|g_n(u) - \bar{g}_n(u)\| = \big\|(J_n - \Phi_{b,n})(F_n - r) + \Phi_{b,n}(F_n - \Phi_{b,n}^\top u)\big\| \le D_n. \tag{140}$$

Also

$$\lambda I \preceq A_n \preceq (\lambda + nG^2)I, \qquad \|I - \gamma A_n\|_{\mathrm{op}} \le q,$$
$$\tfrac{1}{2}\left\|\Phi_{b,n}^\top \bar{u}_n - r\right\|^2 + \tfrac{\lambda}{2}\|\bar{u}_n\|^2 \le \tfrac{1}{2}\|r\|^2 \implies \|\bar{u}_n\| \le \|r\|/\sqrt{\lambda} \le B_y\sqrt{n/\lambda} \le B. \tag{141}$$

Let $w_0 = u_{n-1}$ and $w_{j+1} = w_j - \gamma \bar{g}_n(w_j)$. Then

$$w_j - \bar{u}_n = (I - \gamma A_n)^j (u_{n-1} - \bar{u}_n), \qquad \|w_j - \bar{u}_n\| \le q^j \|u_{n-1} - \bar{u}_n\|, \tag{142}$$
$$v_j - w_j = -\gamma \sum_{\ell<j} (I - \gamma A_n)^{j-1-\ell}\big(g_n(v_\ell) - \bar{g}_n(v_\ell)\big), \quad \|v_j - w_j\| \le \gamma D_n \sum_{\ell \ge 0} q^\ell = \frac{D_n}{\lambda}, \tag{143}$$

the last bound whenever $\|v_\ell\| \le R_1$ for all $\ell < j$.

Double induction. $P(n)$: $\|u_n\| \le R_{\mathrm{out}}$. $P(0)$ holds since $u_0 = 0$. Assume $P(n-1)$ and let $Q(j)$: $\|v_\ell\| \le R_1$ for $\ell \le j$. $Q(0)$: $\|v_0\| \le R_{\mathrm{out}} < R_1$. $Q(j-1) \Rightarrow Q(j)$: by (141)–(143) and (97),

$$\|v_j\| \le \|\bar{u}_n\| + \|u_{n-1} - \bar{u}_n\| + \frac{D_T}{\lambda} \le B + (R_{\mathrm{out}} + B) + B = 7B = R_1. \tag{144}$$

At $j = J_{\mathrm{on}}$, with $q^{J_{\mathrm{on}}} \le 1/8$,

$$\|u_n - \bar{u}_n\| \le q^{J_{\mathrm{on}}}\big(R_{\mathrm{out}} + B_y\sqrt{n/\lambda}\big) + \frac{D_n}{\lambda} = E_n,$$
$$\|u_n\| \le B + \tfrac{1}{8}(4B + B) + B = \tfrac{21}{8}B < R_{\mathrm{out}}, \tag{145}$$

which is (98) and $P(n)$. Every inner iterate satisfies $\|v_j\| \le R_1$, hence lies in the lazy ball. □

G.4 Covariance sandwich

By $P(s-1)$, $\|u_{s-1}\| \le R_1$, so $\|\psi_s - \phi_b(z_s)\| \le e_g$ and

$$\|V_n - A_n\|_{\mathrm{op}} \le \sum_{s \le n} \|\psi_s - \phi_b(z_s)\|\left(\|\psi_s\| + \|\phi_b(z_s)\|\right) \le n\, 2Ge_g \le c_V\lambda. \tag{146}$$

With $\lambda I \preceq A_n$, $-c_V A_n \preceq V_n - A_n \preceq c_V A_n$, which is (99).

G.5 Confidence

Apply Theorem C.3 with $\mathcal{G}_s = \mathcal{F}_s$, $\mathcal{G}_0 = \mathcal{S}_{\mathrm{src}}$, $x_s = \phi_b(z_s)$ ($\mathcal{F}_{s-1}$-measurable since $b \in \mathcal{S}_{\mathrm{src}}$), and $\varepsilon_s = \eta_s$; call the event $\mathcal{E}_{\mathrm{sn,on}}$. By (136) and $\sum_{s\le n}\phi_b(z_s)\phi_b(z_s)^\top = A_n - \lambda I$,

$$\bar{u}_n - u_* = A_n^{-1}\Big[\sum_{s\le n}\phi_b(z_s)\eta_s - \lambda u_*\Big], \qquad \|\bar{u}_n - u_*\|_{A_n} \le \Big\|\sum_{s\le n}\phi_b(z_s)\eta_s\Big\|_{A_n^{-1}} + \sqrt{\lambda}\,\|u_*\| \le a_n. \tag{147}$$

With $\chi_n = \phi_{b+u_n/\sqrt{m}}$ and (45) at $(u, v) = (u_{J_{\mathrm{src}}}, u_{J_{\mathrm{src}}} + u_n)$,

$$r_{\mathrm{Tay}}(z) := f(z; b + u_n/\sqrt{m}) - f_b(z) - \chi_n(z)^\top u_n, \qquad |r_{\mathrm{Tay}}(z)| \le e_f, \tag{148}$$

$$r_{\mathrm{tar}}(z) - f(z; b + u_n/\sqrt{m}) = \chi_n(z)^\top(u_* - u_n) + \big[\phi_b(z) - \chi_n(z)\big]^\top u_* - r_{\mathrm{Tay}}(z). \tag{149}$$

By (99), (98), and (147),

$$|\chi_n(z)^\top(u_* - u_n)| \le s_n(z)\,\|u_* - u_n\|_{V_n} \le s_n(z)\sqrt{1 + c_V}\big(\|u_* - \bar{u}_n\|_{A_n} + \|A_n\|^{1/2}\,\|\bar{u}_n - u_n\|\big)$$
$$\le \sqrt{1 + c_V}\big[a_n + \sqrt{\lambda + nG^2}\,E_n\big]s_n(z) = \beta_n s_n(z), \tag{150}$$

and the last two terms of (149) are at most $e_g S_* + e_f = e_{\mathrm{on}}$. □

H Proof of Theorem D.7 and Corollary D.8

H.1 Theorem D.7

Events. Let $\mathcal{E}_{\mathrm{pre}}$ be the intersection of the event of Theorem C.1, $\mathcal{E}_{y,\mathrm{src}}$, and $\mathcal{E}_{\mathrm{sn,src}}$. Then

$$\Pr(\mathcal{E}_{\mathrm{pre}}^c) \le \delta_{\mathrm{lazy}} + \delta_y + \delta_{\mathrm{src}}, \qquad \Pr\big((\mathcal{E}_{y,\mathrm{on}} \cap \mathcal{E}_{\mathrm{sn,on}})^c\big) = \mathbb{E}\big[\Pr(\cdot \mid \mathcal{S}_{\mathrm{src}})\big] \le \delta_{y,\mathrm{on}} + \delta_{\mathrm{on}},$$

so the intersection of all events has probability at least $1 - \delta_{\mathrm{tot}}$. Work on it.

Source confidence. By (38) and (130),

$$|r_{\mathrm{tar}}(z) - \widehat{r}_{N,\lambda_{\mathrm{src}}}(z)| = |\,\langle r_{\mathrm{tar}} - \widehat{r}_{N,\lambda_{\mathrm{src}}}, k_z\rangle_{\mathcal{H}_{\mathrm{ntk}}}\,| \le \|r_{\mathrm{tar}} - \widehat{r}_{N,\lambda_{\mathrm{src}}}\|_{\widehat{\Sigma}}\, \|k_z\|_{\widehat{\Sigma}^{-1}} \le (\alpha_N + \epsilon)\widehat{\sigma}_N(z). \tag{151}$$

With (63) and (65),

$$|r_{\mathrm{tar}}(z) - f_b(z)| \le (\alpha_N + \epsilon)(a_\sigma s_{\mathrm{src}}(z) + b_\sigma) + \zeta_\mu = \rho_{\mathrm{src}} s_{\mathrm{src}}(z) + \Delta_{\mathrm{src}}. \tag{152}$$

Optimism. By (152) and (100) at $n = t-1$ (note $\widehat{\theta}_t^{\mathrm{on}} = b + u_{t-1}/\sqrt{m}$), $r_{\mathrm{tar}}(z) \le U_t^{\mathrm{src}}(z)$ and $r_{\mathrm{tar}}(z) \le U_t^{\mathrm{on}}(z)$ for $z \in \mathcal{A}_t$. Hence

$$\begin{aligned} r_{\mathrm{tar}}(z_t^*) - r_{\mathrm{tar}}(z_t) &\le U_t(z_t^*) - r_{\mathrm{tar}}(z_t) \le U_t(z_t) - r_{\mathrm{tar}}(z_t) \\ &\le 2\min\big\{\beta_{t-1} s_{t-1}(z_t) + e_{\mathrm{on}},\ \rho_{\mathrm{src}} s_{\mathrm{src}}(z_t) + \Delta_{\mathrm{src}}\big\}. \end{aligned} \tag{153}$$

Summing gives (107).

Elliptical potential. $s_{t-1}(z_t) = \|\psi_t\|_{V_{t-1}^{-1}}$ and $\|\psi_t\|_{V_{t-1}^{-1}}^2 \le G^2/\lambda_{\mathrm{on}} \le 1$. By $\sum_t \min\{a_t, b_t\} \le \min\{\sum_t a_t, \sum_t b_t\}$, Cauchy–Schwarz, and Lemma C.5,

$$\sum_{t \le T} \beta_{t-1} s_{t-1}(z_t) \le \beta_T \Big(T \sum_{t \le T} \|\psi_t\|_{V_{t-1}^{-1}}^2\Big)^{1/2} \le \beta_T \sqrt{2T\Gamma_T},$$

which gives (108). □

H.2 Corollary D.8

Determinants. Let $i_t \in [Q]$ be the occurrence selected at round t (i_t lies in block t of the occurrence list, so $i_1, \dots, i_n$ are distinct), $H_{\mathrm{sel},n} = (H_{\mathrm{occ}}[i_s, i_r])_{s,r \le n}$, and

$$K_n^b = \big(\phi_b(z_s)^\top \phi_b(z_r)\big)_{s,r \le n}, \qquad K_n^{\mathrm{dyn}} = (\psi_s^\top \psi_r)_{s,r \le n}. \tag{154}$$

By Theorem C.1, both matrices are within δ_{ker} of $H_{\mathrm{sel},n}$ entrywise, so

$$\|K_n^\bullet - H_{\mathrm{sel},n}\|_{\mathrm{op}} \le n\delta_{\mathrm{ker}} \le \lambda_{\mathrm{on}}(e^{1/T} - 1). \tag{155}$$

Let μ_i and $\nu_i \ge 1$ be the ordered eigenvalues of $I_n + K_n^\bullet/\lambda_{\mathrm{on}}$ and $I_n + H_{\mathrm{sel},n}/\lambda_{\mathrm{on}}$. By Weyl,

$$\log \mu_i \le \log\big(\nu_i + e^{1/T} - 1\big) \le \log(\nu_i e^{1/T}) = \log \nu_i + \tfrac{1}{T},$$

and summing over $i \le n \le T$,

$$\log\det(I_n + K_n^\bullet/\lambda_{\mathrm{on}}) \le \log\det(I_n + H_{\mathrm{sel},n}/\lambda_{\mathrm{on}}) + 1. \tag{156}$$

$H_{\mathrm{sel},n}$ is a principal submatrix of H_{occ}; for $M = I_Q + H_{\mathrm{occ}}/\lambda_{\mathrm{on}}$ and index set S, $\det M = \det M_{SS} \det(M/M_{SS})$ with Schur complement $M/M_{SS} \succeq I$, so

$$\log\det(I_n + H_{\mathrm{sel},n}/\lambda_{\mathrm{on}}) \le I_{\mathrm{occ}}. \tag{157}$$

By Sylvester, $\det A_n / \det(\lambda_{\mathrm{on}} I) = \det(I_n + K_n^b/\lambda_{\mathrm{on}})$ and $e^{\Gamma_T} = \det(I_T + K_T^{\mathrm{dyn}}/\lambda_{\mathrm{on}})$, which proves (110).

Rate. By (111), $\sqrt{\lambda_{\mathrm{on}} + nG^2}\, E_n \le \sqrt{\lambda_{\mathrm{on}} + TG^2}\big[q^{J_{\mathrm{on}}}(R_{\mathrm{out}} + B_y\sqrt{T/\lambda_{\mathrm{on}}}) + D_T/\lambda_{\mathrm{on}}\big] \le 1$. By (110) and $c_V < 1$,

$$\begin{aligned} a_n &\le \nu\sqrt{I_{\mathrm{occ}} + 1 + 2\log(1/\delta_{\mathrm{on}})} + \sqrt{\lambda_{\mathrm{on}}} S_*, \\ \beta_T &\le \sqrt{2}\big[\nu\sqrt{I_{\mathrm{occ}} + 1 + 2\log(1/\delta_{\mathrm{on}})} + \sqrt{\lambda_{\mathrm{on}}} S_* + 1\big], \\ \mathfrak{R}_T^{\mathrm{on}} &\le 2\beta_T\sqrt{2T(I_{\mathrm{occ}} + 1)} + 2C_{\mathrm{on}}\sqrt{T}, \end{aligned}$$

which is (113). With $I_{\mathrm{occ}} \le \widetilde{d}_{\mathrm{occ}}^{+} \log(1 + Q\kappa^2/\lambda_{\mathrm{on}})$ and $\lambda_{\mathrm{on}} S_*^2 = O(\widetilde{d}_{\mathrm{occ}}^{+})$, both terms are $\widetilde{O}(\widetilde{d}_{\mathrm{occ}}^{+}\sqrt{T})$, which is (115). □

H.3 SUFFICIENT WIDTH

Fix $\gamma_{\rm src}, \gamma_{\rm on}$ ($\eta_\bullet = \gamma_\bullet/m$) and set

$$E^{(0)}_{\rm src} = (1-\gamma_{\rm src}\lambda_{\rm src})^{J_{\rm src}}\frac{B_0}{\sqrt{\lambda_{\rm src}}}, \qquad \zeta^{(0)}_\mu = GE^{(0)}_{\rm src},$$
$$S^{(0)}_* = \frac{\alpha_N+\epsilon}{\sqrt{\lambda_{\rm src}}} + \sqrt{\frac{M}{\lambda_0}}\,\zeta^{(0)}_\mu, \qquad B^{(0)}_y = GS^{(0)}_* + \nu\sqrt{2\log(2T/\delta_{y,\rm on})} + b_{\rm slack}, \tag{158}$$

$$\Delta_R = R - \Big[\frac{2B_0}{\sqrt{\lambda_{\rm src}}} + 7B^{(0)}_y\sqrt{\frac{T}{\lambda_{\rm on}}}\Big] > 0, \quad \epsilon_{\rm ntk}(m) = C_0L^{3/2}|\mathcal{U}|\big[\log(4|\mathcal{U}|L/\delta_{\rm lazy})\big]^{1/4}m^{-1/8}, \tag{159}$$

with R fixed and C_0 a large universal constant. Then, as $m\to\infty$ at fixed R:

(a) $CL^6|\mathcal{U}|^4\epsilon_{\rm ntk}(m)^{-4}\log(4|\mathcal{U}|L/\delta_{\rm lazy}) = CC_0^{-4}m^{1/2} \le m$ for $m \ge (CC_0^{-4})^2$, so (42) holds, and (41) holds with $\tau = R/\sqrt{m}$;

(b) $e_g, e_f = \widetilde{O}(m^{-1/6})$ and $\epsilon_{\rm ntk}(m) = O(m^{-1/8})$, hence $\zeta_\mu \to \zeta^{(0)}_\mu$, $S_* \to S^{(0)}_*$, $B_y \to B^{(0)}_y > 0$, $\bar a, c_g, c_V, \delta_{\rm ker}, D_T, e_{\rm on} \to 0$;

(c) by (b), the right side of (95) tends to $R - \Delta_R < R$, and (97), (109), $\bar a < 1$, $c_g < 1$, and the last two conditions of (111) hold for all

$$m \ge \mathrm{poly}\big(N, M, Q, T, L, \kappa, \lambda^{-1}_{\rm src}, \lambda^{-1}_{\rm on}, \lambda_0^{-1}, B_0, S^{(0)}_*, B^{(0)}_y, (B^{(0)}_y)^{-1}, \Delta_R, \Delta_R^{-1}, C_{\rm on}^{-1}, \log(1/\delta_{\rm lazy})\big); \tag{160}$$

(d) (96) gives $0 < q < 1$, and

$$J_{\rm on} \ge \max\Big\{\frac{\log\big(2\sqrt{\lambda_{\rm on}+TG^2}\,[R_{\rm out} + B_y\sqrt{T/\lambda_{\rm on}}]\big)}{-\log q},\ \frac{\log 8}{-\log q}\Big\} \tag{161}$$

gives the first condition of (111) and $q^{J_{\rm on}} \le 1/8$.

I EXPERIMENTS: PROTOCOL AND FULL RESULTS

This appendix gives the protocol behind Section 5 and the results of the four methods of the main text on every setting: final-regret tables and regret curves.

I.1 PROTOCOL

Common protocol. All four methods see the same instance, the same candidate sets, and the same noise draws. The three neural methods also share the initialization θ_0 of a given instance, the network, the training routine (Algorithm 2), the refit schedule, and the same $\lambda_{\rm on}$, $J_{\rm on}$, and step sizes; none of these is tuned per method. Every method gets a budget of 16 configurations, selected on three tuning instances that are disjoint from the 20 evaluation instances; the selection rule is the mean final regret over the tuning instances, ties broken by the smaller grid index. No method uses the norm of the target reward: the online radius uses $S = 1$. The only oracle input is the supplied certificate $(\epsilon, B_{\rm src})$ of the problem setting, used by the source radius of Transfer-NeuralUCB. Every method is also reported untuned, with all multipliers at their theory scale.

Network, features, training. Contexts are unit vectors with equal halves. The network is $f(z;\theta) = \sqrt{m}\,w^\top\sigma(Wz)$ with $W \in \mathbb{R}^{m\times d}$, $w \in \mathbb{R}^m$, m even, and the paired initialization of Zhou et al. (2020), so $f(z;\theta_0) = 0$; all entries of W and w are trained. The scaled tangent feature is $\phi_\theta(z) = m^{-1/2}\nabla_\theta f(z;\theta)$, and Gram matrices are formed from activations without materializing the p-dimensional features. TRAINNN$(\lambda, \eta, J, \text{data}, \text{center}, \text{start})$ runs J full-batch gradient steps on $\frac{1}{2}\sum_i(f(z_i) - y_i)^2 + \frac{m\lambda}{2}\|\theta - \text{center}\|^2$. Source training is TRAINNN$(N\lambda_{\rm src}, \gamma_{\rm src}/(Nm), J_{\rm src}, \mathcal{D}_0, \theta_0, \theta_0)$; online training after n rewards is TRAINNN$(\lambda_{\rm on}, \gamma_{\rm on}/m, J_{\rm on}, \text{data}_{1..n}, \text{center}, \text{previous iterate})$ with $\gamma_{\rm on} = 1/(\lambda_{\rm on} + 2n)$, refitted

after every round for $t \le 20$ and every 40 rounds thereafter. The online covariance uses the features recorded at selection time, $V_n = \lambda_{\text{on}} I + \sum_{s \le n} \psi_s \psi_s^\top$ with $\psi_s = \phi_{\theta_s}(z_s)$, the width is $s_n(z) = \left\|\phi_{\theta_{n+1}}(z)\right\|_{V_n^{-1}}$ computed exactly by Woodbury on the $n \times n$ Gram matrix, and $\Gamma_n = \log\det V_n / \det(\lambda_{\text{on}} I)$ is accumulated exactly. Defaults: synthetic $m = 300$, $\lambda_{\text{src}} = 3 \cdot 10^{-3}$, $\gamma_{\text{src}} = 0.45$, $J_{\text{src}} = 2500$, $\lambda_{\text{on}} = 2$, $J_{\text{on}} = 25$, $\sigma_{\text{src}} = 0.05$, $\nu = 0.10$; real data $m = 100$, $J_{\text{src}} = 1500$, $J_{\text{on}} = 20$, $\nu = 0.10$ in the radius. Everything is in double precision.

Radii. The three neural methods use one online radius,

$$\beta_n = c\left[\nu\sqrt{\Gamma_n + 2\log(1/\delta)} + \sqrt{\lambda_{\text{on}}}\, S\right], \qquad S = 1,\ \delta = 0.05, \tag{162}$$

with the multiplier c as its knob. The source width is $s_{\text{src}}(z) = \|\phi_b(z)\|_{G_{\text{src}}^{-1}}$ of (6), and the source radius of Transfer-NeuralUCB is $\rho_{\text{src}} = c_{\text{src}}(\alpha_N + \epsilon)$ with α_N the source statistical radius (70) evaluated with the exact NTK Gram matrix of the source contexts. The finite-width corrections are set to $\Delta_{\text{src}} = e_{\text{on}} = 0$, so the radii are the diagnostic instantiation of (7), not the certified constants of Appendix D.

Methods. *Transfer-NeuralUCB* is Algorithm 1 with knobs c and c_{src}. *NeuralUCB, cold start* (Zhou et al., 2020) starts at θ_0 and centers the ridge at θ_0. *NeuralUCB, warm start* starts at the source-trained b but centers the ridge at θ_0, as in the original algorithm. *LinUCB* (Abbasi-Yadkori et al., 2011) runs on the raw contexts with a per-arm intercept on real data and a constant intercept on synthetic data, with the data-dependent radius of the shape (162), with Γ_n replaced by the log-determinant of its own design matrix. The two NeuralUCB variants share all plumbing with Transfer-NeuralUCB (training routine, refit schedule, full covariance, features at selection time, radius shape), so the comparison isolates the center of the online model and the source score.

Instances. Synthetic contexts are $z = (v, v)/(\sqrt{2}\,\|v\|)$ with $v \in \mathbb{R}^6$. The source reward is the minimum-norm NTK interpolant of $\cos(6\, w^\top \alpha)$ at 60 anchors, rescaled to $\|r_{\text{src}}\|_\infty = 1$; the discrepancy d is built the same way with an independent w and frequency 6.9 and rescaled so that $\|d\|_{\widehat{\Sigma}_{N,\lambda_{\text{src}}}} = \epsilon$. Noise is Gaussian with $\sigma_{\text{src}} = 0.05$ and $\nu = 0.1$ (the noise variant uses $\nu = 0.3$). The effective frequency of $\cos(6w^\top \alpha)$ along $v/\|v\|$ is random per instance (quantiles $2.6/4.2/5.4$); it affects every method equally. Real-data instances are those of Section 5.2. Real-data features are standardized and reduced to $d_f = 10$ principal components fitted on the source instances only, then normalized, and the block one-hot contexts are duplicated as above; `shuttle` has nine raw features, so its $d_f = 9$. The online noise is zero (the rewards are deterministic) while $\nu = 0.1$ is kept in the radius, and $\sigma_{\text{src}} = 0$ in α_N. Candidate arms are offered in a random order so that the smallest-index tie rule cannot favour a class. All reward-mixture settings of a dataset share the split, the permutation π, and the arm orders for a given instance and differ only in the source labels. The covariate-shift split uses 2-means with 20 restarts on PCA features of all instances; the source is the larger cluster. The label mixture is

$$r_{\text{src}}(z_a) = (1-p)\,\mathbf{1}\{a = y\} + p\,\mathbf{1}\{a = \pi(y)\}. \tag{163}$$

Real-data certificate. $\epsilon^2 = \lambda_{\text{src}}\|d\|_{\mathcal{H}_{\text{ntk}}}^2 + \frac{1}{N}\sum_i d(Z_i^0)^2$ with the empirical part exact and $\|d\|_{\mathcal{H}_{\text{ntk}}}^2$ replaced by the minimum-norm NTK interpolant on a random subsample of 300 source and 300 online contexts; B_{src} is computed on the same subsample. Both are lower bounds that do not converge as the subsample grows (the $0/1$ labels are not in the RKHS): they roughly double from 300 to 1000 points and are 1.2 to 4 times larger on the full sample. Real-data Transfer-NeuralUCB therefore uses a plug-in source radius, not a certified one.

Tuning grid and statistics. $C_{16} = \{10^{-3+4j/15} : j = 0, \dots, 15\}$. The two NeuralUCB variants and LinUCB take $c \in C_{16}$; Transfer-NeuralUCB takes $c \in C_8 = C_{16}[0 :: 2]$ and $c_{\text{src}} \in \{1, 3\}$, also 16 configurations. The metric is the final cumulative pseudo-regret R_T, mean and standard error (sample standard deviation over $\sqrt{20}$) over the 20 evaluation instances at the selected configuration. Each baseline is compared with Transfer-NeuralUCB by a two-sided Wilcoxon signed-rank test on the paired instances, with Holm correction over the three comparisons within a setting at $\alpha = 0.05$.

Table 3: Paired comparisons of Transfer-NeuralUCB with each baseline over the 12 synthetic and 20 real-data settings at $T = 2000$ (Wilcoxon, Holm over the three comparisons within a setting): number of settings in which our regret is significantly lower ($<$), not significantly different ($\approx$), or significantly higher ($>$).

	synthetic			real data			all		
ours vs.	$<$	$\approx$	$>$	$<$	$\approx$	$>$	$<$	$\approx$	$>$
NeuralUCB, warm start	12	0	0	15	3	2	27	3	2
NeuralUCB, cold start	12	0	0	14	4	2	26	4	2
LinUCB	12	0	0	16	1	3	28	1	3

Two disclosures. First, at $c_{\mathrm{src}} \in \{1, 3\}$ the source score of Transfer-NeuralUCB is vacuous in every setting: at the selected c it never attains the minimum at the chosen action, so the reported regret of our method is that of its source-centered online model. Second, the warm start keeps almost no source information: at $n = 1$ its ridge contracts $\theta - \theta_0$ by 0.5 per gradient step, so after its first refit it is close to the cold start. Its role is to show that initializing at b alone does not transfer.

Compute. One run takes 5 to 10 seconds at $T = 2000$ on one CPU core, and at $T = 20000$ about 70 minutes for `magic` and about two hours for `mnist` (the exact $n \times n$ Woodbury update grows with the number of rounds); the long-horizon runs use 12 GB of memory.

I.2 Summary

Table 3 counts, over the 32 settings, how often Transfer-NeuralUCB is significantly better than, indistinguishable from, or significantly worse than each baseline. It has the lowest mean regret of the four methods in 26 settings, and in each of them the difference to all three baselines is significant. It is better than all three baselines on every synthetic setting. The four settings in which some baseline is significantly better are all real-data settings with strong reward mismatch ($p = 0.75$ on `magic`, `covertype`, and `mnist`) or covariate shift (`covertype`).

I.3 Synthetic results

Table 4 gives the final regret on the 12 synthetic settings and Figure 4 the regret curves. Transfer-NeuralUCB starts lower and grows more slowly than both NeuralUCB variants in every panel; its regret rises with ϵ from 9.3 to 77, while the cold start stays at 56–112. The warm start coincides with the cold start everywhere. More source data helps little beyond $N = 500$ at $\epsilon = 0.1$, because the ridge floor $\sqrt{\lambda_{\mathrm{src}}} B_{\mathrm{src}}$ of α_N does not shrink with N; a larger universe ($M = 1000$) and larger noise ($\nu = 0.3$) raise every method's regret. LinUCB grows linearly on these nonlinear rewards.

I.4 Real-data results

Table 5 gives the final regret on the 20 real-data settings and Figure 5 the regret curves. At $p \leq 0.5$ our method is below the three baselines on every dataset, with the largest gain on `shuttle`. At $p = 0.75$ the source labels are mostly permuted and the curves cross: our method starts higher and stays higher on `magic` and `covertype`. Under covariate shift it is best on `magic` and `mnist`, significantly worse than LinUCB on `covertype`, and not significantly different on `shuttle`, where its mean is higher. In the failure cases the source predictor is confidently wrong on the online instances, and the plug-in certificate, which is computed on the source design, does not flag it.

I.5 Untuned radii

Table 6 reports the four methods with all multipliers at their theory scale. Untuned, the neural UCB methods are over-conservative, by up to a factor of 17 relative to their tuned regret on the synthetic settings. The ordering between Transfer-NeuralUCB and the NeuralUCB variants is kept: its mean regret is lower than the cold start's in 31 of the 32 settings. LinUCB, whose untuned radius is less conservative, has the lowest untuned regret on `shuttle`, `covertype`, and `mnist` (15 settings in

Table 4: Synthetic settings. Top: mismatch sweep at $N = 1000$, $M = 100$, $\nu = 0.1$. Bottom: source size N, universe size M, and noise level ν at $\epsilon = 0.1$. Mean $\pm$ one standard error of the final cumulative regret R_T ($T = 2000$) over 20 instances, each method at the configuration selected on three separate tuning instances from a budget of 16. Markers against Transfer-NeuralUCB (paired two-sided Wilcoxon signed-rank test, Holm correction over the three comparisons, $\alpha = 0.05$): $\uparrow$ significantly higher regret, $\downarrow$ significantly lower, $\approx$ not significant. Bold: lowest mean.

method	$\epsilon = 0$	0.02	0.05	0.1	0.2	0.4	0.8
Transfer-NeuralUCB (ours)	**9.34 ± 0.70**	**9.66 ± 0.71**	**9.97 ± 0.77**	**12.4 ± 0.8**	**18.8 ± 1.3**	**37.2 ± 1.9**	**77.3 ± 4.3**
NeuralUCB, warm start	56.9 ± 5.2$^{\uparrow}$	60.7 ± 4.0$^{\uparrow}$	70.9 ± 2.9$^{\uparrow}$	63.1 ± 2.2$^{\uparrow}$	68.2 ± 5.7$^{\uparrow}$	75.2 ± 8.7$^{\uparrow}$	108 ± 6$^{\uparrow}$
NeuralUCB, cold start	55.7 ± 4.4$^{\uparrow}$	63.0 ± 3.8$^{\uparrow}$	64.0 ± 5.9$^{\uparrow}$	57.5 ± 4.5$^{\uparrow}$	64.7 ± 3.5$^{\uparrow}$	74.4 ± 6.9$^{\uparrow}$	112 ± 6$^{\uparrow}$
LinUCB	291 ± 21$^{\uparrow}$	305 ± 21$^{\uparrow}$	322 ± 26$^{\uparrow}$	363 ± 28$^{\uparrow}$	423 ± 36$^{\uparrow}$	472 ± 39$^{\uparrow}$	739 ± 66$^{\uparrow}$

method	$N = 125$	$N = 500$	$N = 4000$	$M = 1000$	$\nu = 0.3$
Transfer-NeuralUCB (ours)	**14.4 ± 1.2**	**11.9 ± 0.9**	**12.7 ± 1.1**	**22.3 ± 1.3**	**39.3 ± 3.4**
NeuralUCB, warm start	84.2 ± 9.3$^{\uparrow}$	54.5 ± 4.1$^{\uparrow}$	64.3 ± 2.1$^{\uparrow}$	73.9 ± 5.1$^{\uparrow}$	110 ± 9$^{\uparrow}$
NeuralUCB, cold start	78.6 ± 7.1$^{\uparrow}$	57.7 ± 4.7$^{\uparrow}$	60.0 ± 3.2$^{\uparrow}$	77.8 ± 4.9$^{\uparrow}$	82.2 ± 4.5$^{\uparrow}$
LinUCB	336 ± 22$^{\uparrow}$	303 ± 18$^{\uparrow}$	290 ± 13$^{\uparrow}$	390 ± 17$^{\uparrow}$	340 ± 23$^{\uparrow}$

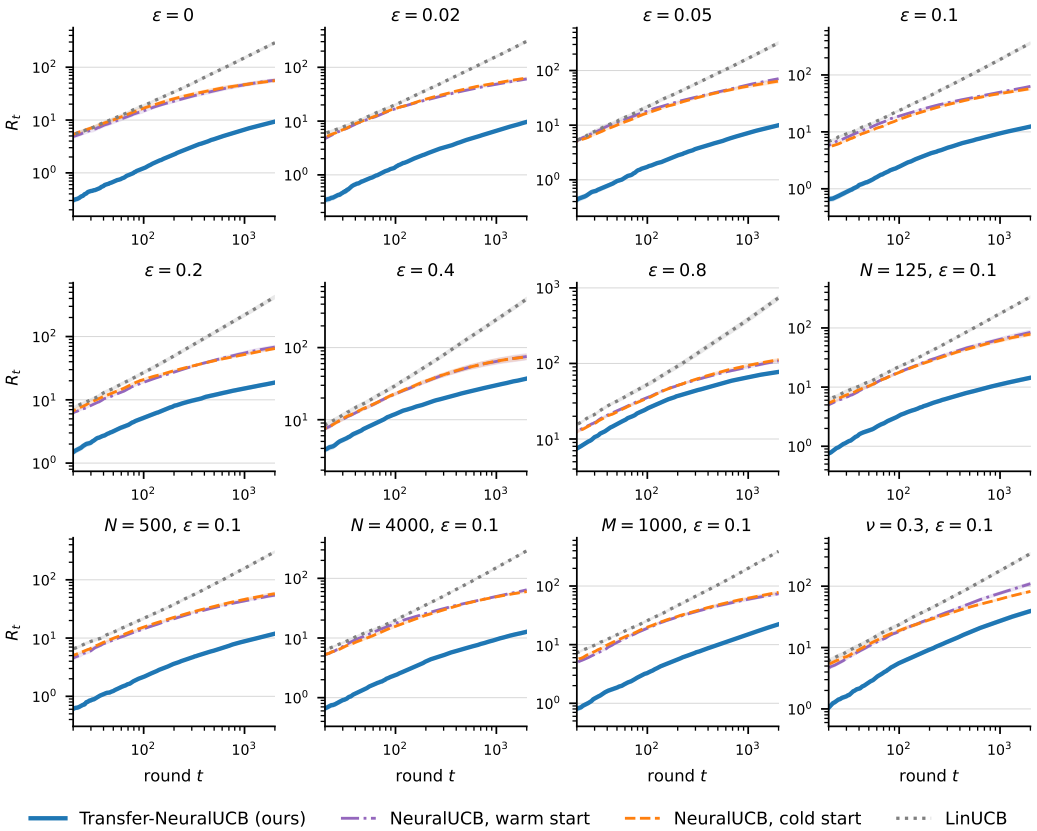


Figure 4: Synthetic cumulative regret against the round, log–log, on all 12 synthetic settings. Mean and one standard error over 20 instances, each method at its selected configuration.

all), and Transfer-NeuralUCB on all synthetic settings (16 in all). At $c = 1$ its source score attains the minimum in at most 1.4% of the rounds.

Table 5: Classification-derived bandits: reward-mixture shift p and covariate shift. Mean $\pm$ one standard error of the final cumulative regret R_T ($T = 2000$) over 20 instances, each method at the configuration selected on three separate tuning instances from a budget of 16. Markers against Transfer-NeuralUCB (paired two-sided Wilcoxon signed-rank test, Holm correction over the three comparisons, $\alpha = 0.05$): $\uparrow$ significantly higher regret, $\downarrow$ significantly lower, $\approx$ not significant. Bold: lowest mean. On real data the certificate is a plug-in value from a $300 + 300$ subsample, so the source radius of Transfer-NeuralUCB is not certified.

method	$p = 0$	$p = 0.25$	$p = 0.5$	$p = 0.75$	cov. shift
`magic`					
Transfer-NeuralUCB (ours)	**316 ± 4**	**325 ± 4**	**351 ± 4**	407 ± 4	**183 ± 2**
NeuralUCB, warm start	363 ± 4$^{\uparrow}$	369 ± 5$^{\uparrow}$	358 ± 4$^{\uparrow}$	359 ± 4$^{\downarrow}$	211 ± 3$^{\uparrow}$
NeuralUCB, cold start	366 ± 4$^{\uparrow}$	368 ± 5$^{\uparrow}$	361 ± 4$^{\uparrow}$	**359 ± 4**$^{\downarrow}$	213 ± 3$^{\uparrow}$
LinUCB	430 ± 4$^{\uparrow}$	430 ± 4$^{\uparrow}$	430 ± 4$^{\uparrow}$	430 ± 4$^{\uparrow}$	203 ± 2$^{\uparrow}$
`shuttle`					
Transfer-NeuralUCB (ours)	**87.5 ± 4.2**	**94.8 ± 4.4**	**136 ± 9**	201 ± 14	234 ± 37
NeuralUCB, warm start	214 ± 10$^{\uparrow}$	188 ± 8$^{\uparrow}$	192 ± 7$^{\uparrow}$	199 ± 9$^{\approx}$	**176 ± 8**$^{\approx}$
NeuralUCB, cold start	193 ± 6$^{\uparrow}$	191 ± 5$^{\uparrow}$	195 ± 8$^{\uparrow}$	**186 ± 9**$^{\approx}$	180 ± 8$^{\approx}$
LinUCB	195 ± 3$^{\uparrow}$	195 ± 3$^{\uparrow}$	195 ± 3$^{\uparrow}$	195 ± 3$^{\approx}$	310 ± 13$^{\uparrow}$
`covertype`					
Transfer-NeuralUCB (ours)	**716 ± 5**	**731 ± 5**	**769 ± 7**	1004 ± 28	710 ± 11
NeuralUCB, warm start	874 ± 10$^{\uparrow}$	883 ± 19$^{\uparrow}$	881 ± 9$^{\uparrow}$	874 ± 11$^{\downarrow}$	700 ± 9$^{\approx}$
NeuralUCB, cold start	876 ± 12$^{\uparrow}$	877 ± 11$^{\uparrow}$	877 ± 9$^{\uparrow}$	884 ± 9$^{\downarrow}$	708 ± 12$^{\approx}$
LinUCB	809 ± 8$^{\uparrow}$	809 ± 8$^{\uparrow}$	809 ± 8$^{\uparrow}$	**809 ± 8**$^{\downarrow}$	**670 ± 11**$^{\downarrow}$
`mnist`					
Transfer-NeuralUCB (ours)	**554 ± 11**	**583 ± 11**	**641 ± 12**	784 ± 7	**659 ± 17**
NeuralUCB, warm start	795 ± 16$^{\uparrow}$	778 ± 12$^{\uparrow}$	853 ± 19$^{\uparrow}$	833 ± 8$^{\uparrow}$	752 ± 21$^{\uparrow}$
NeuralUCB, cold start	800 ± 16$^{\uparrow}$	792 ± 11$^{\uparrow}$	839 ± 17$^{\uparrow}$	786 ± 15$^{\approx}$	753 ± 19$^{\uparrow}$
LinUCB	759 ± 10$^{\uparrow}$	759 ± 10$^{\uparrow}$	759 ± 10$^{\uparrow}$	**759 ± 10**$^{\downarrow}$	725 ± 13$^{\uparrow}$

I.6 Long horizon

Table 7 gives the final regret of the $T = 20000$ runs behind Figure 3, together with the ratio of Transfer-NeuralUCB to the cold start after 2000 and after 20000 rounds. The ratio moves toward one on every dataset at $p = 0$, and the lead at $p = 0.5$ is significant only on `magic` and `mnist`. After the source split `magic` has fewer than 20000 instances left, so its online stream continues with a with-replacement resample of the same instances; `shuttle` and `mnist` are sampled without replacement.

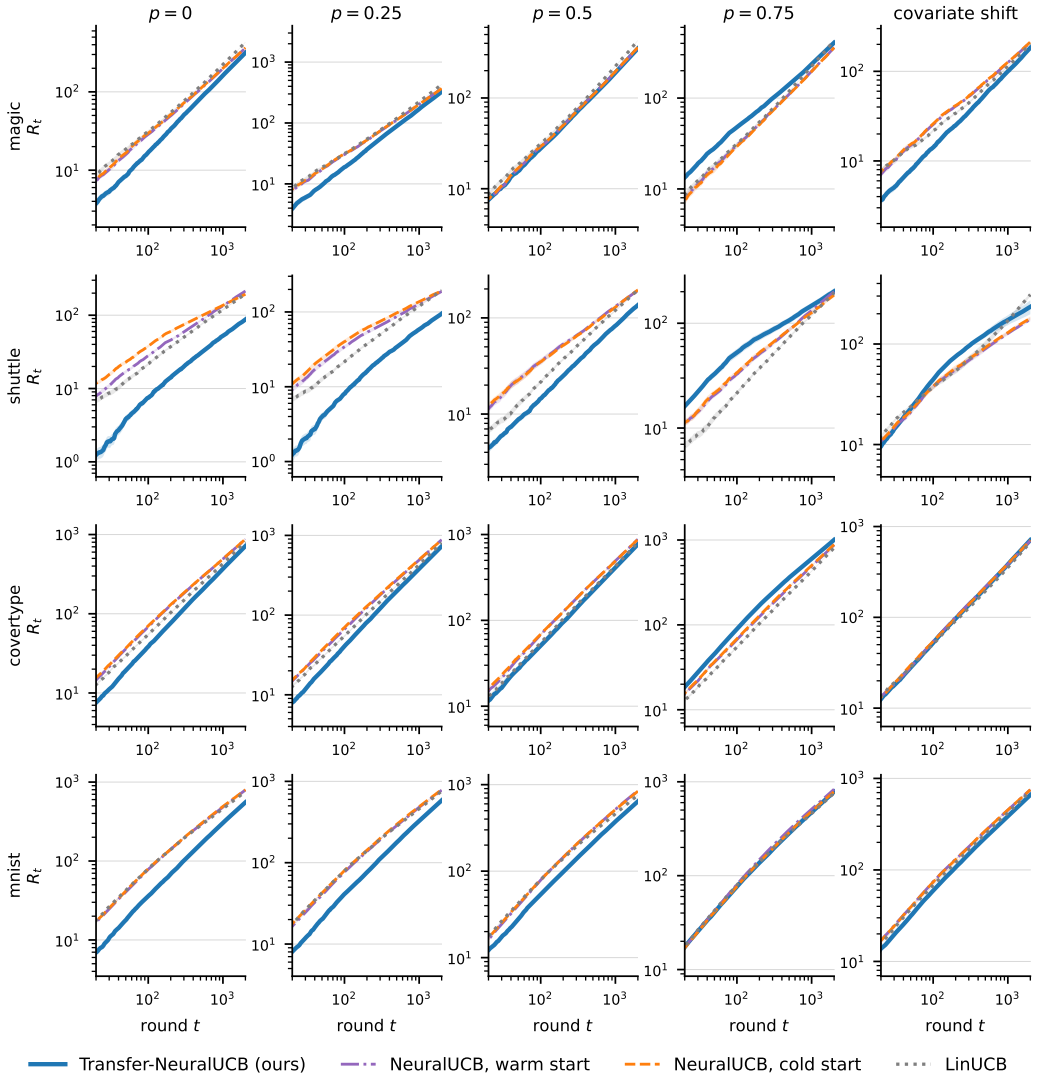


Figure 5: Classification-derived bandits: cumulative regret against the round, log–log, for the four datasets (rows) under the reward-mixture source with $p \in \{0, 0.25, 0.5, 0.75\}$ and under covariate shift (columns). Mean and one standard error over 20 instances, $T = 2000$, each method at its selected configuration.

Table 6: Untuned radii: every multiplier at its theory scale ($c = c_{\mathrm{src}} = 1$). Final regret at $T = 2000$, mean $\pm$ one standard error over 20 instances; bold: lowest mean. Top: synthetic mismatch sweep. Bottom: classification-derived bandits.

method	$\epsilon = 0$	0.02	0.05	0.1	0.2	0.4	0.8
Transfer-NeuralUCB (ours)	**160 ± 6**	**159 ± 5**	**161 ± 4**	**157 ± 7**	**142 ± 6**	**115 ± 6**	**91.6 ± 2.1**
NeuralUCB, warm start	186 ± 6	184 ± 5	189 ± 3	184 ± 6	166 ± 6	134 ± 5	113 ± 5
NeuralUCB, cold start	187 ± 6	187 ± 5	188 ± 4	182 ± 6	167 ± 6	134 ± 5	116 ± 4
LinUCB	309 ± 23	310 ± 14	322 ± 22	347 ± 14	398 ± 34	469 ± 33	689 ± 52

method	$p = 0$	$p = 0.25$	$p = 0.5$	$p = 0.75$	cov. shift
`magic`					
Transfer-NeuralUCB (ours)	**347 ± 4**	**369 ± 4**	**388 ± 3**	427 ± 5	214 ± 2
NeuralUCB, warm start	392 ± 4	392 ± 4	**388 ± 4**	**396 ± 4**	239 ± 2
NeuralUCB, cold start	392 ± 4	393 ± 4	**388 ± 4**	**396 ± 4**	239 ± 2
LinUCB	439 ± 4	439 ± 4	439 ± 4	439 ± 4	**208 ± 2**
`shuttle`					
Transfer-NeuralUCB (ours)	407 ± 13	414 ± 11	438 ± 13	466 ± 13	**220 ± 7**
NeuralUCB, warm start	474 ± 13	461 ± 11	468 ± 12	473 ± 13	242 ± 7
NeuralUCB, cold start	472 ± 12	462 ± 12	468 ± 12	473 ± 13	246 ± 6
LinUCB	**311 ± 5**	**311 ± 5**	**311 ± 5**	**311 ± 5**	318 ± 5
`covertype`					
Transfer-NeuralUCB (ours)	1160 ± 11	1174 ± 12	1174 ± 9	1223 ± 9	856 ± 23
NeuralUCB, warm start	1236 ± 9	1230 ± 11	1212 ± 9	1237 ± 9	870 ± 23
NeuralUCB, cold start	1229 ± 10	1228 ± 11	1211 ± 9	1234 ± 9	869 ± 22
LinUCB	**1017 ± 7**	**1017 ± 7**	**1017 ± 7**	**1017 ± 7**	**732 ± 10**
`mnist`					
Transfer-NeuralUCB (ours)	1031 ± 19	1065 ± 9	1095 ± 13	1144 ± 13	967 ± 17
NeuralUCB, warm start	1158 ± 17	1168 ± 10	1169 ± 12	1167 ± 13	1019 ± 17
NeuralUCB, cold start	1159 ± 18	1168 ± 10	1170 ± 13	1172 ± 12	1021 ± 17
LinUCB	**858 ± 8**	**858 ± 8**	**858 ± 8**	**858 ± 8**	**827 ± 11**

Table 7: Long horizon: final regret at $T = 20000$ on the classification-derived bandits under the reward-mixture source, mean $\pm$ one standard error over 10 instances, each method at the multiplier selected for the same problem at $T = 2000$. $\uparrow$: significantly higher regret than Transfer-NeuralUCB (paired Wilcoxon, Holm over the three comparisons). The last two columns give the ratio of the mean regret of Transfer-NeuralUCB to that of the cold start after 2000 and after 20000 rounds.

dataset	p	ours	NeuralUCB warm	NeuralUCB cold	LinUCB	ratio @2000	@20000
`magic`	0	2830 ± 10	2892 ± 12$^{\uparrow}$	2893 ± 13$^{\uparrow}$	4047 ± 12$^{\uparrow}$	0.88	0.98
`shuttle`	0	458 ± 15	639 ± 40$^{\uparrow}$	603 ± 17$^{\uparrow}$	1377 ± 86$^{\uparrow}$	0.49	0.76
`mnist`	0	3695 ± 91	4152 ± 86$^{\uparrow}$	4143 ± 89$^{\uparrow}$	5434 ± 116$^{\uparrow}$	0.70	0.89
`magic`	0.5	2879 ± 9	2899 ± 9$^{\uparrow}$	2903 ± 9$^{\uparrow}$	4047 ± 12$^{\uparrow}$	0.96	0.99
`shuttle`	0.5	591 ± 46	666 ± 39	624 ± 24	1377 ± 86$^{\uparrow}$	0.70	0.95
`mnist`	0.5	3793 ± 98	4224 ± 101$^{\uparrow}$	4061 ± 87$^{\uparrow}$	5434 ± 116$^{\uparrow}$	0.79	0.93