Image Contextual Bandits: A Visual Transformer Approach

Ryan Wong

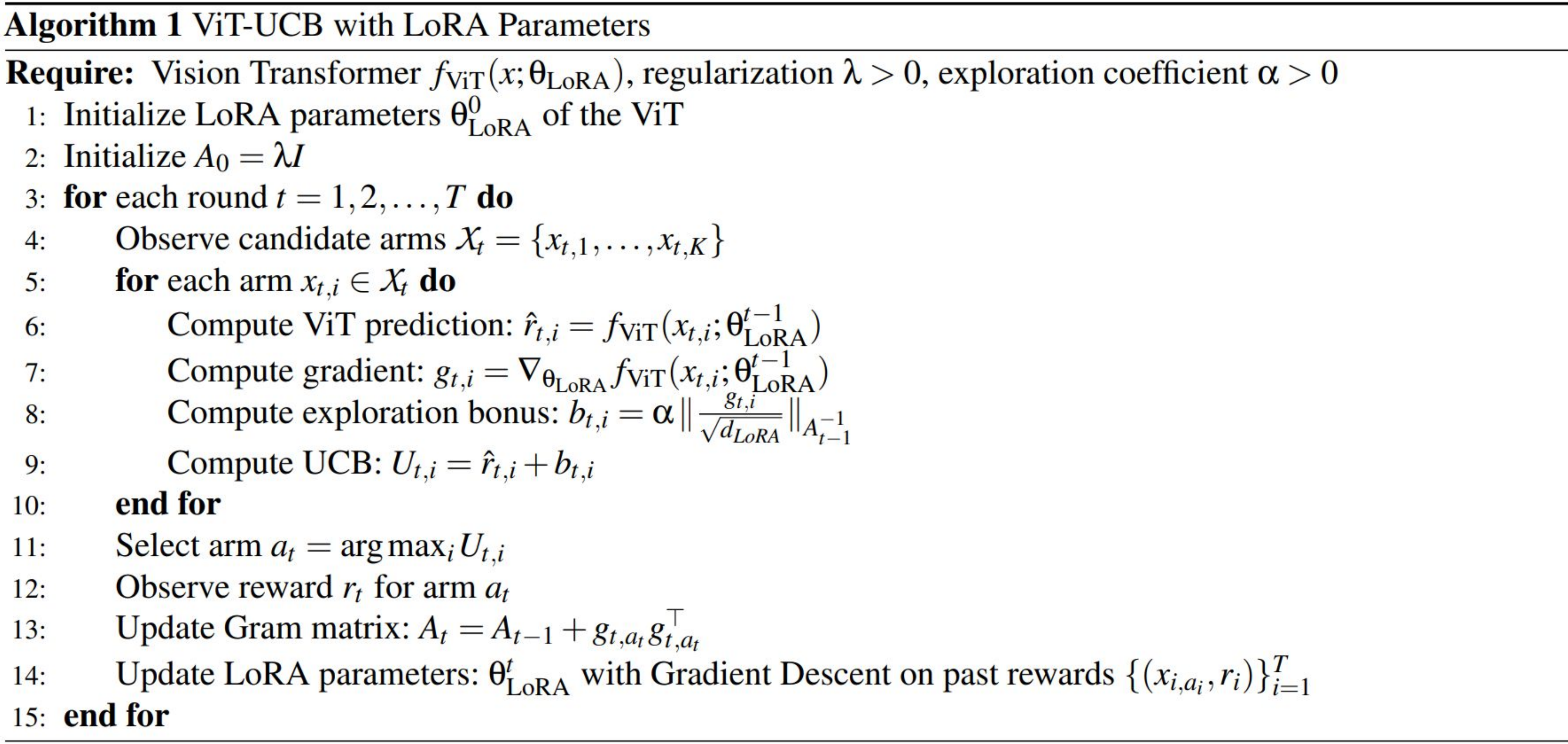
RyanW124/ViT-Contextual-Bandits

Proposal

- Goal: develop a ViT-based UCB algorithm for image contextual image
- Approach unchanged (showed later)
- Experiment:
    - Misconception that each arm has a single context vector
    - Instead, each arm’s context is different at each time step
- Deliverables:
    - GitHub
    - Paper

Problem: Image Contextual Bandits

- Agent repeatedly chooses from a set of actions (termed "arms") with initially unknown reward distributions
- Exploration vs. Exploitation
- Agent goal: maximize expected cumulative reward (minimize regret)
- At each time step, each arm has an image context
- Recommender System
    - Arm is which show to recommend
    - Context is thumbnail
    - Reward is user interaction

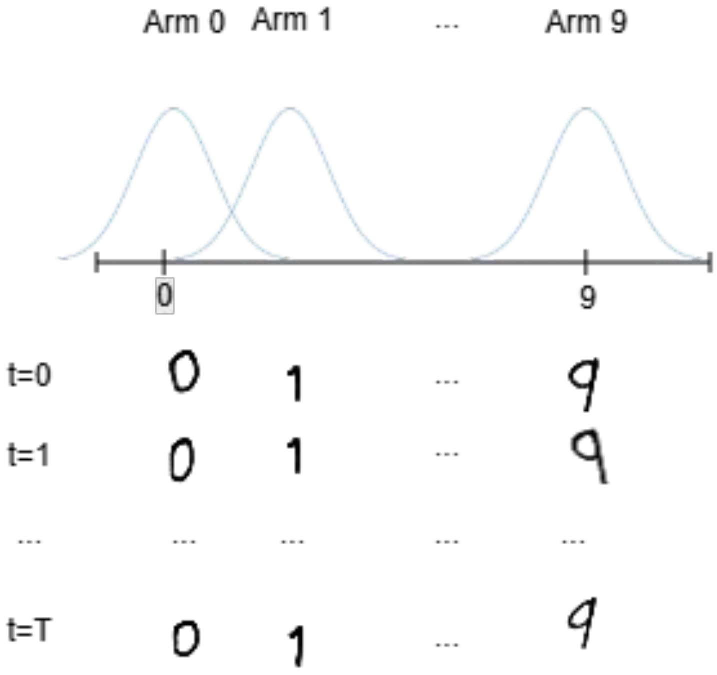


Algorithm 1 ViT-UCB with LoRA Parameters

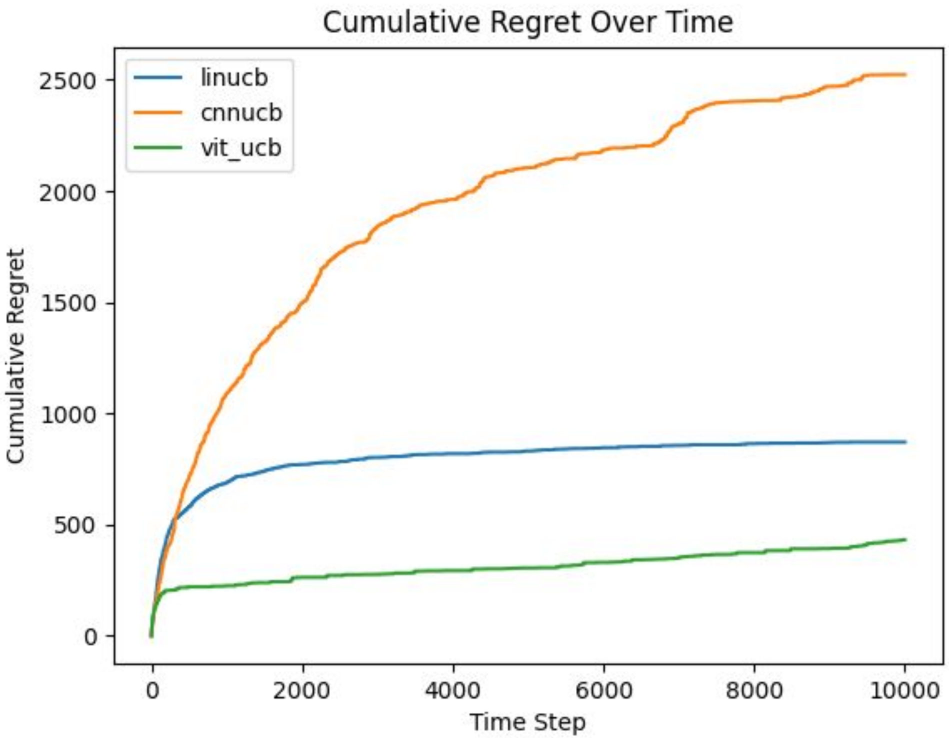
```
Require: Vision Transformer $f_{\text{ViT}}(x;\theta_{\text{LoRA}})$, regularization $\lambda > 0$, exploration coefficient $\alpha > 0$
Initialize LoRA parameters $\theta^0_{\text{LoRA}}$ of the ViT
Initialize $A_0 = \lambda I$
for each round $t = 1, 2, \ldots, T$ do
    Observe candidate arms $\mathcal{X}_t = \{x_{t,1}, \ldots, x_{t,K}\}$
    for each arm $x_{t,i} \in \mathcal{X}_t$ do
        Compute ViT prediction: $\hat{r}_{t,i} = f_{\text{ViT}}(x_{t,i};\theta^{t-1}_{\text{LoRA}})$
        Compute gradient: $g_{t,i} = \nabla_{\theta_{\text{LoRA}}} f_{\text{ViT}}(x_{t,i};\theta^{t-1}_{\text{LoRA}})$
        Compute exploration bonus: $b_{t,i} = \alpha \| \frac{g_{t,i}}{\sqrt{d_{LoRA}}} \|_{A^{-1}_{t-1}}$
        Compute UCB: $U_{t,i} = \hat{r}_{t,i} + b_{t,i}$
    end for
    Select arm $a_t = \arg\max_i U_{t,i}$
    Observe reward $r_t$ for arm $a_t$
    Update Gram matrix: $A_t = A_{t-1} + g_{t,a_t} g^{\top}_{t,a_t}$
    Update LoRA parameters: $\theta^t_{\text{LoRA}}$ with Gradient Descent on past rewards $\{(x_{i,a_i}, r_i)\}^T_{i=1}$
end for
```

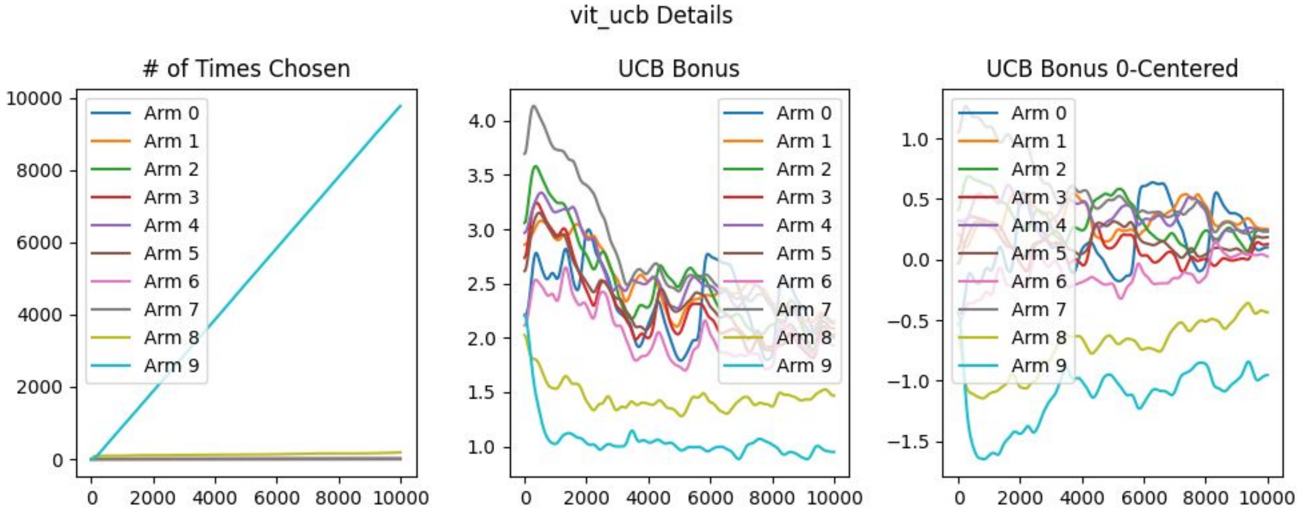
Experiment 1 (handwritten digits)

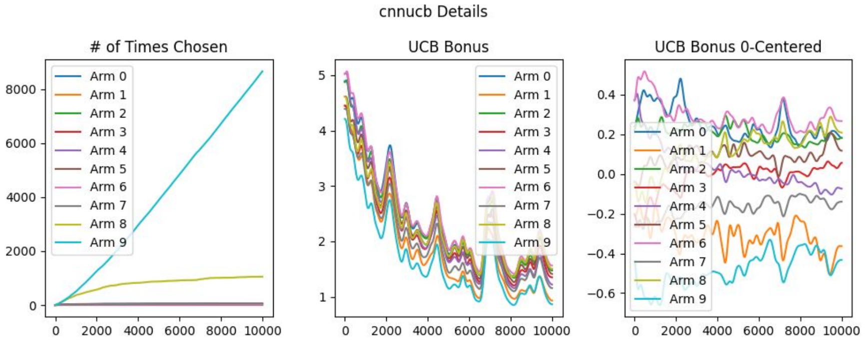
- Each arm is a digit (10 arms total from 0-9)
- Arm i has contexts of handwritten digit i
  - Sampled uniformly random at each t
- Arm i samples reward from N(i, 9)
- T=10000
- Baselines
  - CNN UCB
  - LinUCB (On embedding from a pretrained model)
- ViT UCB
  - Model: WinKawaks/vit-tiny-patch16-224 (5.7M param)
  - LoRA rank = 20, LoRA alpha = 24, alpha=85
  - Tuned using 50 trials of Bayesian Optimization

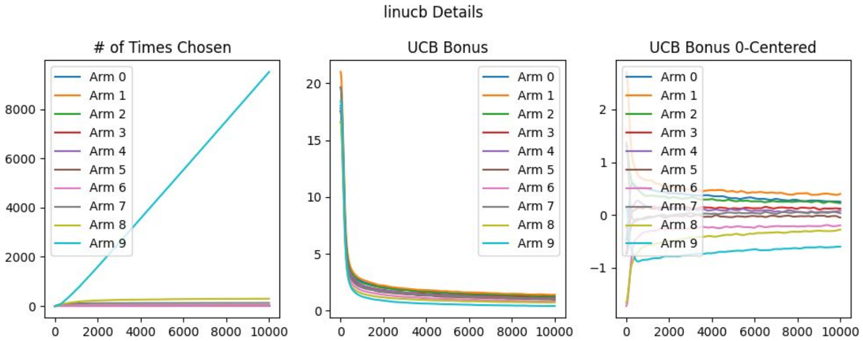


Experiment 1 Results




vit_ucb Details
# of Times Chosen
UCB Bonus
UCB Bonus 0-Centered
Arm 0
Arm 1
Arm 2
Arm 3
Arm 4
Arm 5
Arm 6
Arm 7
Arm 8
Arm 9
0
2000
4000
6000
8000
10000
1.0
1.5
2.0
2.5
3.0
3.5
4.0
−1.5
−1.0
−0.5
0.0
0.5
1.0

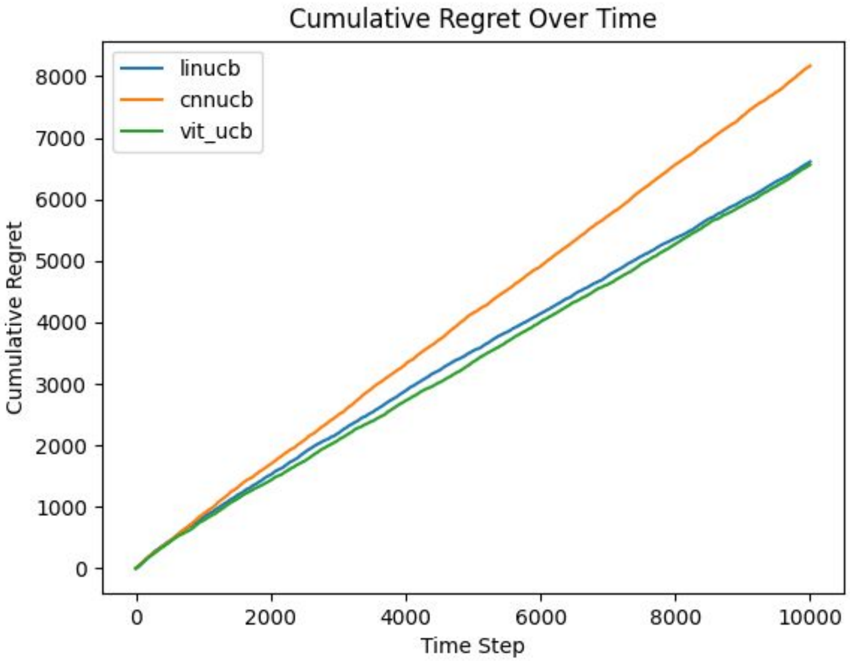

cnnucb Details
# of Times Chosen
UCB Bonus
UCB Bonus 0-Centered
Arm 0
Arm 1
Arm 2
Arm 3
Arm 4
Arm 5
Arm 6
Arm 7
Arm 8
Arm 9
0
2000
4000
6000
8000
10000
1
2
3
4
5
−0.6
−0.4
−0.2
0.0
0.2
0.4

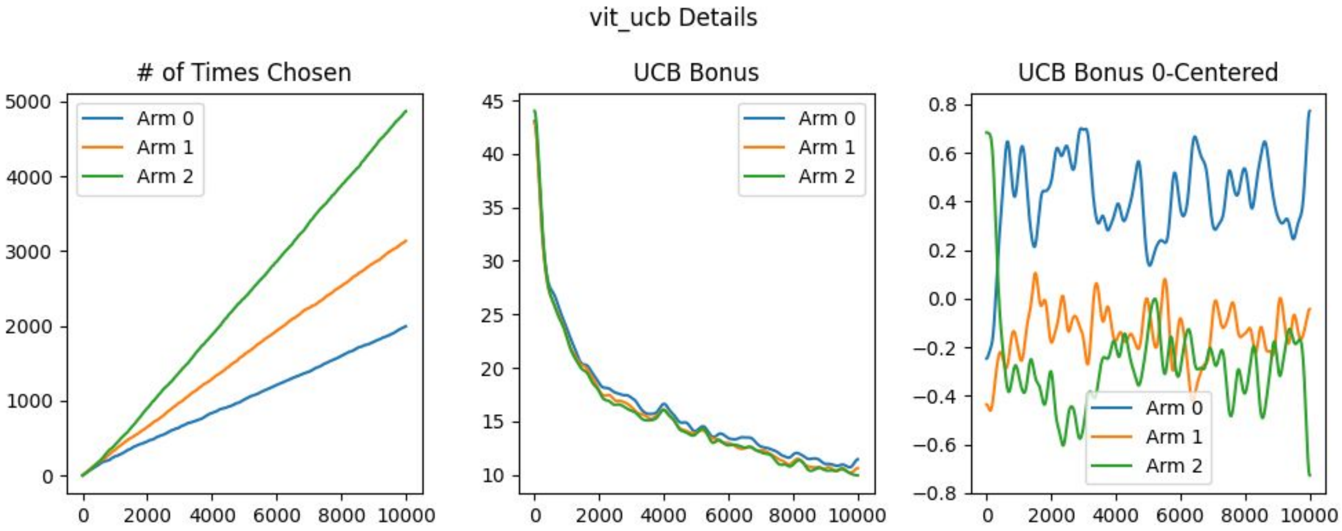

linucb Details
# of Times Chosen
UCB Bonus
UCB Bonus 0-Centered
Arm 0
Arm 1
Arm 2
Arm 3
Arm 4
Arm 5
Arm 6
Arm 7
Arm 8
Arm 9
0
2000
4000
6000
8000
10000
0
5
10
15
20
−1
0
1
2

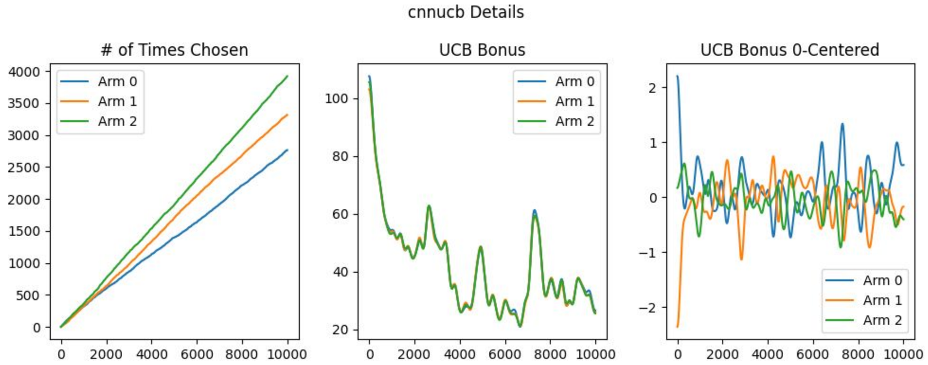
Experiment 2 (Anime)

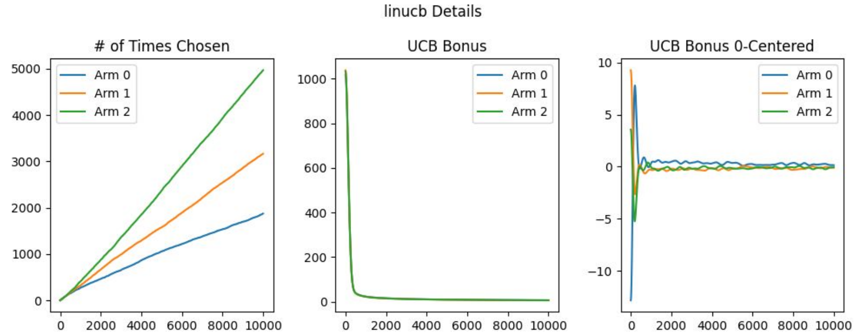
- Contexts are 10000 anime thumbnails pulled from MyAnimeList
- Sort animes by rating and divide into 3 groups of equal size
    - E.g. if animes have ratings {7.2, 2.6, 4.4, 5.3, 9.1, 8.3}
    - Group 0 = {2.6, 4.4}
    - Group 1 = {5.3, 7.2}
    - Group 2 = {8.3, 9.1}
- Arm i has contexts from group i
    - Sampled uniformly random at each t
- Arm i samples reward from N(mean rating of group i, 1)
- In the experiment
    - Group 0 mean ≈ 5.5
    - Group 1 mean ≈ 6.5
    - Group 2 mean ≈ 7.4

Experiment 2 Results




vit_ucb Details
# of Times Chosen
UCB Bonus
UCB Bonus 0-Centered
Arm 0
Arm 1
Arm 2
5000
4000
3000
2000
1000
0
45
40
35
30
25
20
15
10
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0
2000
4000
6000
8000
10000


cnnucb Details
# of Times Chosen
UCB Bonus
UCB Bonus 0-Centered
Arm 0
Arm 1
Arm 2
4000
3500
3000
2500
2000
1500
1000
500
0
100
80
60
40
20
2
1
0
−1
−2
0
2000
4000
6000
8000
10000

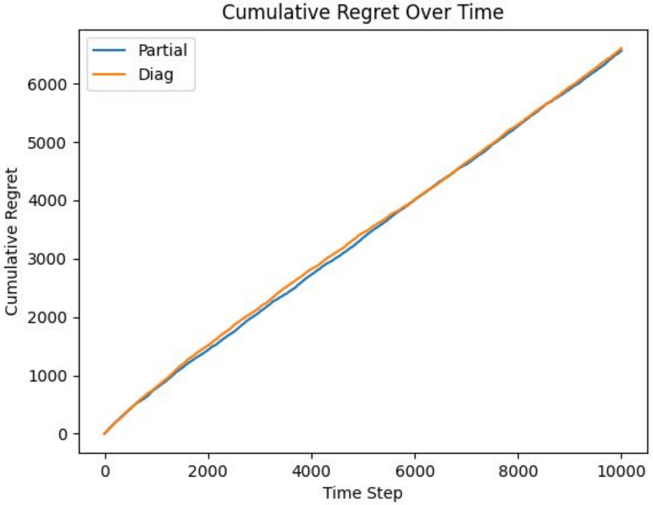

linucb Details
# of Times Chosen
UCB Bonus
UCB Bonus 0-Centered
Arm 0
Arm 1
Arm 2
5000
4000
3000
2000
1000
0
1000
800
600
400
200
0
10
5
0
−5
−10
0
2000
4000
6000
8000
10000

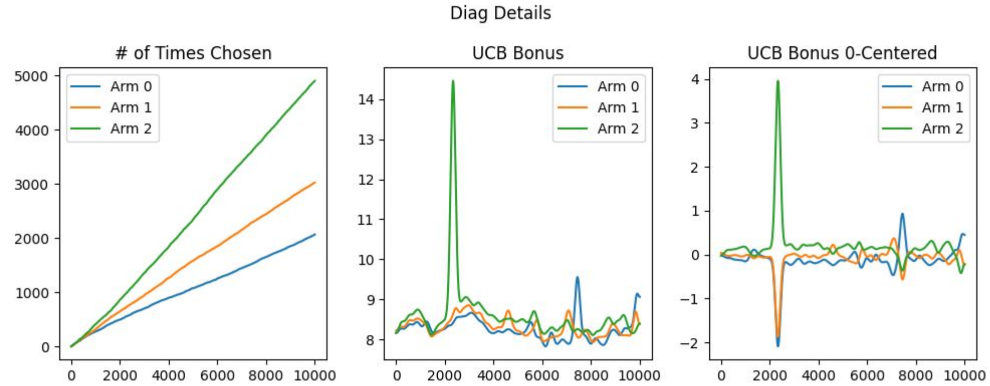
Feedback from Midterm

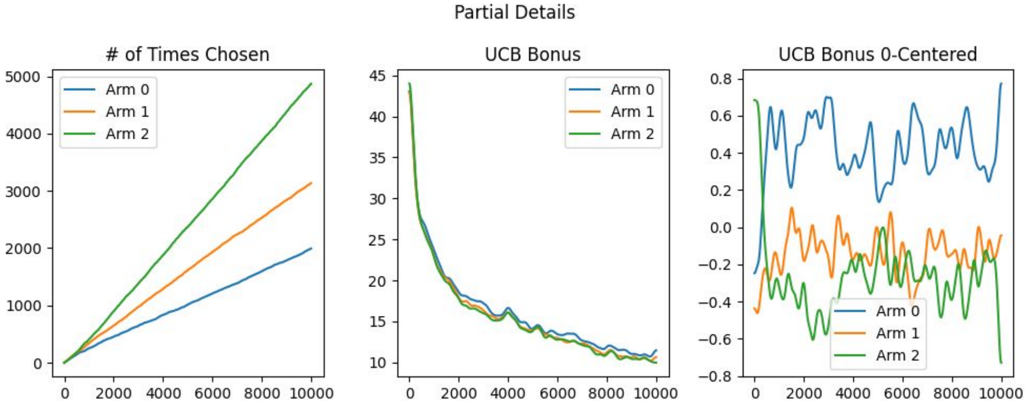
- Recall $A = \lambda I + \sum_{t=1}^{T} gg^\intercal$
- So if model has *n* trainable parameters, need to store n^2
- Storing the matrix *A* was the major bottleneck from using bigger models
  - Any way to approximate
- Using a portion of parameters vs. only storing diagonal

Experiment: Portion vs. Diagonal

- Ran Anime experiment
- Portion: WinKawaks/vit-tiny-patch16-224
    - 5.6M total params
    - 24041 params tracked
- Diagonal: google/vit-base-patch16-224
    - 87M total params
    - 1106921 params tracked


Cumulative Regret Over Time
Partial
Diag
Cumulative Regret
0
1000
2000
3000
4000
5000
6000
Time Step
0
2000
4000
6000
8000
10000

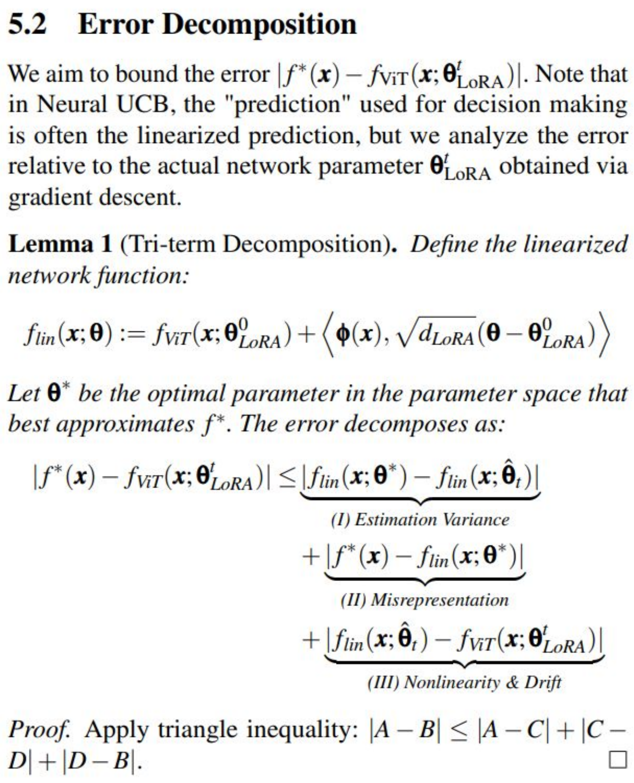

Diag Details
# of Times Chosen
Arm 0
Arm 1
Arm 2
0
1000
2000
3000
4000
5000
UCB Bonus
8
9
10
11
12
13
14
UCB Bonus 0-Centered
−2
−1
0
1
2
3
4
0
2000
4000
6000
8000
10000


Partial Details
# of Times Chosen
Arm 0
Arm 1
Arm 2
0
1000
2000
3000
4000
5000
UCB Bonus
10
15
20
25
30
35
40
45
UCB Bonus 0-Centered
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
0
2000
4000
6000
8000
10000

Theoretical Analysis

- Goal: Justify our choice of upper confidence bound
- Recall, exploration bonus is $\alpha \left\| \frac{g_{t,i}}{\sqrt{d_{LoRA}}} \right\|_{A_{t-1}^{-1}}$
- Show that the bonus bounds the error between estimate and actual reward

Theoretical Analysis: Intermediate Steps 1



5.2 Error Decomposition

We aim to bound the error $|f^*(\boldsymbol{x}) - f_{\text{ViT}}(\boldsymbol{x};\boldsymbol{\theta}^t_{\text{LoRA}})|$. Note that in Neural UCB, the "prediction" used for decision making is often the linearized prediction, but we analyze the error relative to the actual network parameter $\boldsymbol{\theta}^t_{\text{LoRA}}$ obtained via gradient descent.

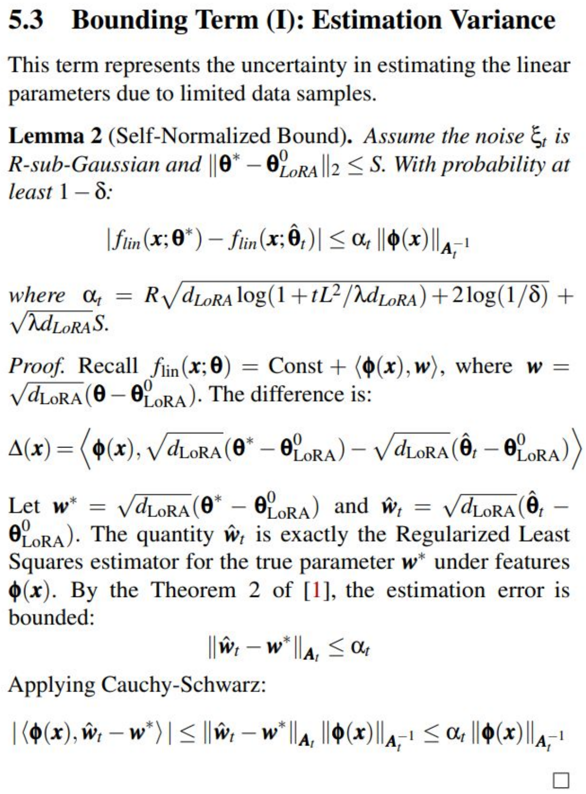
Lemma 1 (Tri-term Decomposition)**.** *Define the linearized network function:*

$$f_{lin}(\boldsymbol{x};\boldsymbol{\theta}) := f_{ViT}(\boldsymbol{x};\boldsymbol{\theta}^0_{LoRA}) + \left\langle \boldsymbol{\phi}(\boldsymbol{x}), \sqrt{d_{LoRA}}(\boldsymbol{\theta}-\boldsymbol{\theta}^0_{LoRA})\right\rangle$$

Let $\boldsymbol{\theta}^*$ be the optimal parameter in the parameter space that best approximates f^*. The error decomposes as:

$$\begin{aligned} |f^*(\boldsymbol{x}) - f_{ViT}(\boldsymbol{x};\boldsymbol{\theta}^t_{LoRA})| \leq & \underbrace{|f_{lin}(\boldsymbol{x};\boldsymbol{\theta}^*) - f_{lin}(\boldsymbol{x};\hat{\boldsymbol{\theta}}_t)|}_{\textit{(I) Estimation Variance}} \\ & + \underbrace{|f^*(\boldsymbol{x}) - f_{lin}(\boldsymbol{x};\boldsymbol{\theta}^*)|}_{\textit{(II) Misrepresentation}} \\ & + \underbrace{|f_{lin}(\boldsymbol{x};\hat{\boldsymbol{\theta}}_t) - f_{ViT}(\boldsymbol{x};\boldsymbol{\theta}^t_{LoRA})|}_{\textit{(III) Nonlinearity \& Drift}} \end{aligned}$$

Proof. Apply triangle inequality: $|A-B| \leq |A-C| + |C-D| + |D-B|$. □



5.3 Bounding Term (I): Estimation Variance

This term represents the uncertainty in estimating the linear parameters due to limited data samples.

Lemma 2 (Self-Normalized Bound)**.** *Assume the noise ξ_t is R-sub-Gaussian and $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}^0_{LoRA}\|_2 \leq S$. With probability at least $1-\delta$:*

$$|f_{lin}(\boldsymbol{x};\boldsymbol{\theta}^*) - f_{lin}(\boldsymbol{x};\hat{\boldsymbol{\theta}}_t)| \leq \alpha_t \, \|\boldsymbol{\phi}(\boldsymbol{x})\|_{\boldsymbol{A}_t^{-1}}$$

where $\alpha_t = R\sqrt{d_{LoRA}\log(1+tL^2/\lambda d_{LoRA}) + 2\log(1/\delta)} + \sqrt{\lambda d_{LoRA}}S$.

Proof. Recall $f_{\text{lin}}(\boldsymbol{x};\boldsymbol{\theta}) = \text{Const} + \langle\boldsymbol{\phi}(\boldsymbol{x}),\boldsymbol{w}\rangle$, where $\boldsymbol{w} = \sqrt{d_{\text{LoRA}}}(\boldsymbol{\theta}-\boldsymbol{\theta}^0_{\text{LoRA}})$. The difference is:

$$\Delta(\boldsymbol{x}) = \left\langle \boldsymbol{\phi}(\boldsymbol{x}), \sqrt{d_{\text{LoRA}}}(\boldsymbol{\theta}^* - \boldsymbol{\theta}^0_{\text{LoRA}}) - \sqrt{d_{\text{LoRA}}}(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^0_{\text{LoRA}})\right\rangle$$

Let $\boldsymbol{w}^* = \sqrt{d_{\text{LoRA}}}(\boldsymbol{\theta}^* - \boldsymbol{\theta}^0_{\text{LoRA}})$ and $\hat{\boldsymbol{w}}_t = \sqrt{d_{\text{LoRA}}}(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^0_{\text{LoRA}})$. The quantity $\hat{\boldsymbol{w}}_t$ is exactly the Regularized Least Squares estimator for the true parameter $\boldsymbol{w}^*$ under features $\boldsymbol{\phi}(\boldsymbol{x})$. By the Theorem 2 of [1], the estimation error is bounded:

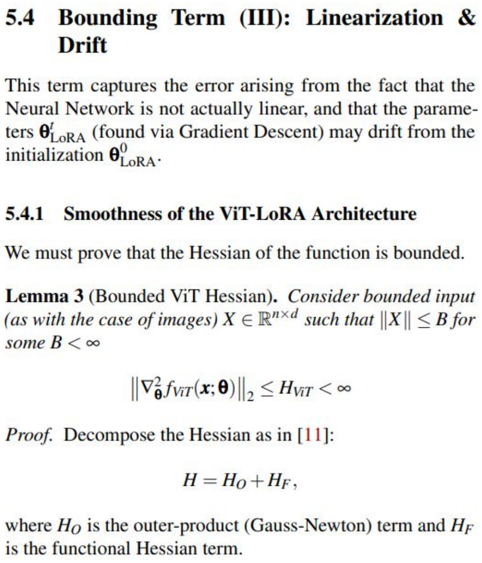
$$\|\hat{\boldsymbol{w}}_t - \boldsymbol{w}^*\|_{\boldsymbol{A}_t} \leq \alpha_t$$

Applying Cauchy-Schwarz:

$$|\langle\boldsymbol{\phi}(\boldsymbol{x}), \hat{\boldsymbol{w}}_t - \boldsymbol{w}^*\rangle| \leq \|\hat{\boldsymbol{w}}_t - \boldsymbol{w}^*\|_{\boldsymbol{A}_t} \|\boldsymbol{\phi}(\boldsymbol{x})\|_{\boldsymbol{A}_t^{-1}} \leq \alpha_t \|\boldsymbol{\phi}(\boldsymbol{x})\|_{\boldsymbol{A}_t^{-1}}$$

□

Theoretical Analysis: Intermediate Steps 2



5.4 Bounding Term (III): Linearization & Drift

This term captures the error arising from the fact that the Neural Network is not actually linear, and that the parameters $\boldsymbol{\theta}^t_{\text{LoRA}}$ (found via Gradient Descent) may drift from the initialization $\boldsymbol{\theta}^0_{\text{LoRA}}$.

5.4.1 Smoothness of the ViT-LoRA Architecture

We must prove that the Hessian of the function is bounded.

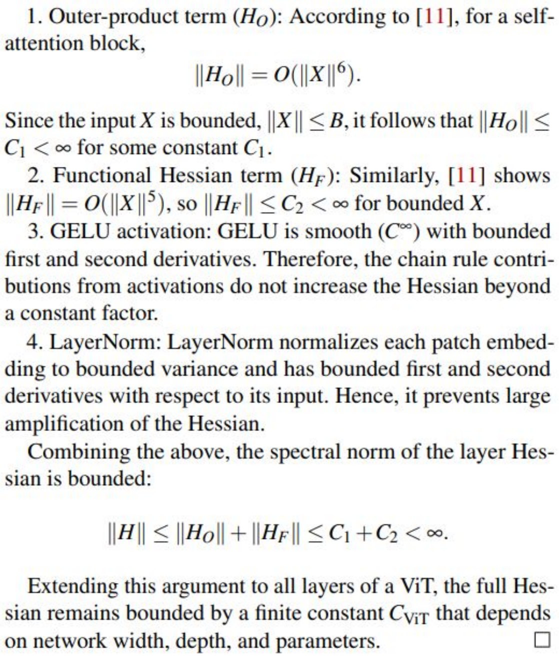
Lemma 3 (Bounded ViT Hessian). *Consider bounded input (as with the case of images) $X \in \mathbb{R}^{n \times d}$ such that $\|X\| \leq B$ for some $B < \infty$*

$$\left\|\nabla^2_{\boldsymbol{\theta}} f_{ViT}(\boldsymbol{x};\boldsymbol{\theta})\right\|_2 \leq H_{ViT} < \infty$$

Proof. Decompose the Hessian as in [11]:

$$H = H_O + H_F,$$

where H_O is the outer-product (Gauss-Newton) term and H_F is the functional Hessian term.



1. Outer-product term (H_O): According to [11], for a self-attention block,

$$\|H_O\| = O(\|X\|^6).$$

Since the input X is bounded, $\|X\| \leq B$, it follows that $\|H_O\| \leq C_1 < \infty$ for some constant C_1.

2. Functional Hessian term (H_F): Similarly, [11] shows $\|H_F\| = O(\|X\|^5)$, so $\|H_F\| \leq C_2 < \infty$ for bounded X.

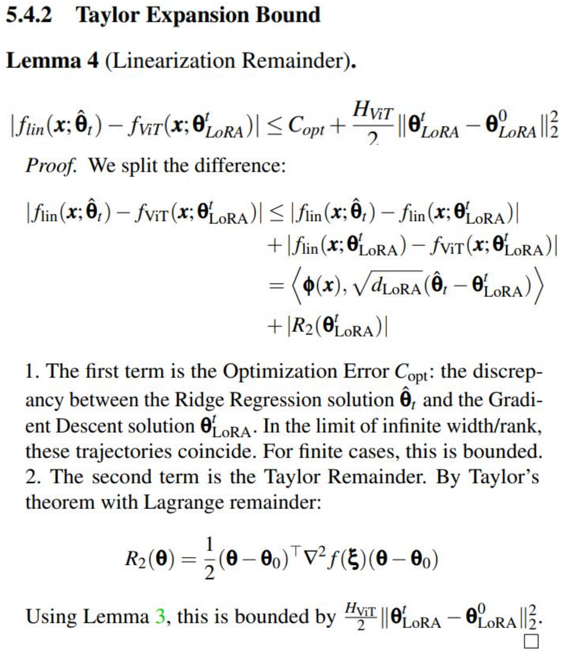
3. GELU activation: GELU is smooth (C^∞) with bounded first and second derivatives. Therefore, the chain rule contributions from activations do not increase the Hessian beyond a constant factor.

4. LayerNorm: LayerNorm normalizes each patch embedding to bounded variance and has bounded first and second derivatives with respect to its input. Hence, it prevents large amplification of the Hessian.

Combining the above, the spectral norm of the layer Hessian is bounded:

$$\|H\| \leq \|H_O\| + \|H_F\| \leq C_1 + C_2 < \infty.$$

Extending this argument to all layers of a ViT, the full Hessian remains bounded by a finite constant C_{ViT} that depends on network width, depth, and parameters. □



5.4.2 Taylor Expansion Bound

Lemma 4 (Linearization Remainder).

$$|f_{lin}(\boldsymbol{x};\hat{\boldsymbol{\theta}}_t) - f_{ViT}(\boldsymbol{x};\boldsymbol{\theta}^t_{LoRA})| \leq C_{opt} + \frac{H_{ViT}}{2}\|\boldsymbol{\theta}^t_{LoRA} - \boldsymbol{\theta}^0_{LoRA}\|_2^2$$

Proof. We split the difference:

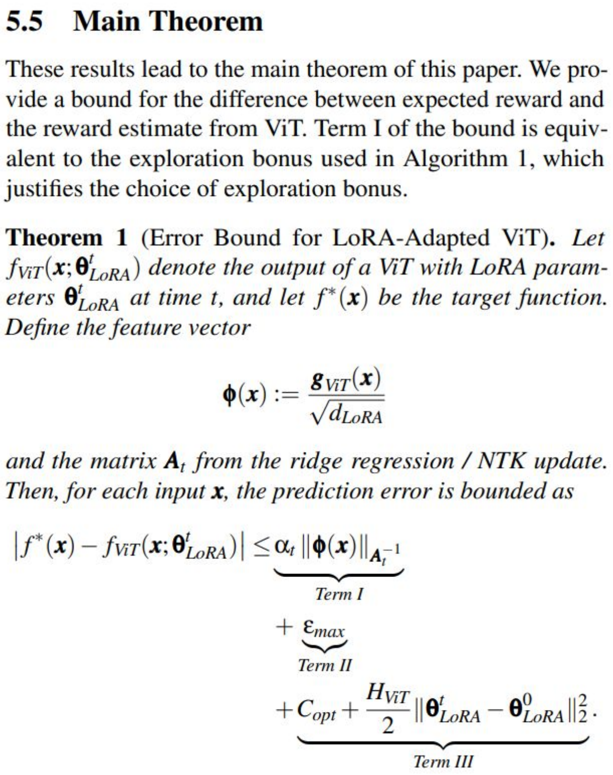
$$\begin{aligned}|f_{\text{lin}}(\boldsymbol{x};\hat{\boldsymbol{\theta}}_t) - f_{\text{ViT}}(\boldsymbol{x};\boldsymbol{\theta}^t_{\text{LoRA}})| &\leq |f_{\text{lin}}(\boldsymbol{x};\hat{\boldsymbol{\theta}}_t) - f_{\text{lin}}(\boldsymbol{x};\boldsymbol{\theta}^t_{\text{LoRA}})| \\ &+ |f_{\text{lin}}(\boldsymbol{x};\boldsymbol{\theta}^t_{\text{LoRA}}) - f_{\text{ViT}}(\boldsymbol{x};\boldsymbol{\theta}^t_{\text{LoRA}})| \\ &= \left\langle \boldsymbol{\phi}(\boldsymbol{x}), \sqrt{d_{\text{LoRA}}}(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^t_{\text{LoRA}})\right\rangle \\ &+ |R_2(\boldsymbol{\theta}^t_{\text{LoRA}})|\end{aligned}$$

1. The first term is the Optimization Error C_{opt}: the discrepancy between the Ridge Regression solution $\hat{\boldsymbol{\theta}}_t$ and the Gradient Descent solution $\boldsymbol{\theta}^t_{\text{LoRA}}$. In the limit of infinite width/rank, these trajectories coincide. For finite cases, this is bounded.
2. The second term is the Taylor Remainder. By Taylor's theorem with Lagrange remainder:

$$R_2(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \nabla^2 f(\boldsymbol{\xi})(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

Using Lemma 3, this is bounded by $\frac{H_{\text{ViT}}}{2}\|\boldsymbol{\theta}^t_{\text{LoRA}} - \boldsymbol{\theta}^0_{\text{LoRA}}\|_2^2$. □

Theoretical Analysis: Main Bound



5.5 Main Theorem

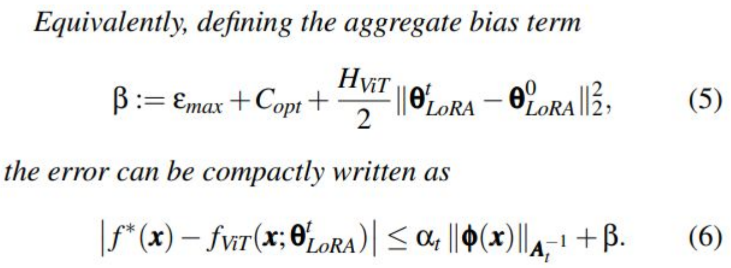
These results lead to the main theorem of this paper. We provide a bound for the difference between expected reward and the reward estimate from ViT. Term I of the bound is equivalent to the exploration bonus used in Algorithm 1, which justifies the choice of exploration bonus.

Theorem 1 (Error Bound for LoRA-Adapted ViT). *Let $f_{ViT}(\boldsymbol{x};\boldsymbol{\theta}^t_{LoRA})$ denote the output of a ViT with LoRA parameters $\boldsymbol{\theta}^t_{LoRA}$ at time t, and let $f^*(\boldsymbol{x})$ be the target function. Define the feature vector*

$$\boldsymbol{\phi}(\boldsymbol{x}) := \frac{\boldsymbol{g}_{ViT}(\boldsymbol{x})}{\sqrt{d_{LoRA}}}$$

and the matrix $\boldsymbol{A}_t$ from the ridge regression / NTK update. Then, for each input $\boldsymbol{x}$, the prediction error is bounded as

$$\begin{aligned}\left|f^*(\boldsymbol{x}) - f_{ViT}(\boldsymbol{x};\boldsymbol{\theta}^t_{LoRA})\right| \le & \underbrace{\alpha_t \,\|\boldsymbol{\phi}(\boldsymbol{x})\|_{\boldsymbol{A}_t^{-1}}}_{Term\ I} \\ & + \underbrace{\varepsilon_{max}}_{Term\ II} \\ & \underbrace{+C_{opt} + \frac{H_{ViT}}{2}\|\boldsymbol{\theta}^t_{LoRA} - \boldsymbol{\theta}^0_{LoRA}\|_2^2}_{Term\ III}.\end{aligned}$$

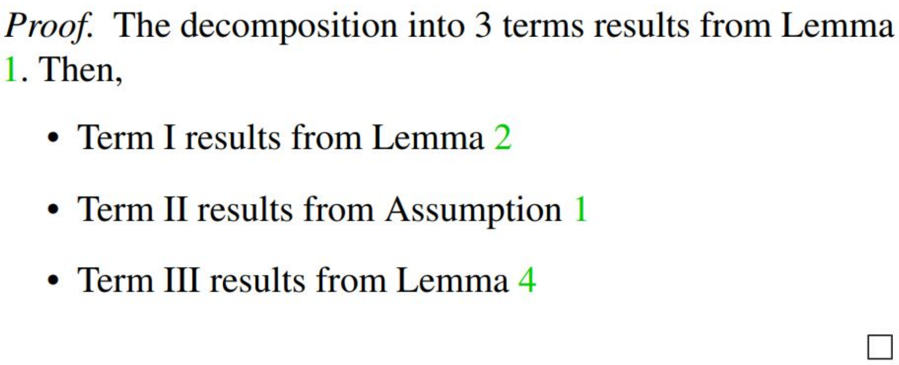


Equivalently, defining the aggregate bias term

$$\beta := \varepsilon_{max} + C_{opt} + \frac{H_{ViT}}{2}\|\boldsymbol{\theta}^t_{LoRA} - \boldsymbol{\theta}^0_{LoRA}\|_2^2, \qquad (5)$$

the error can be compactly written as

$$\left|f^*(\boldsymbol{x}) - f_{ViT}(\boldsymbol{x};\boldsymbol{\theta}^t_{LoRA})\right| \le \alpha_t \,\|\boldsymbol{\phi}(\boldsymbol{x})\|_{\boldsymbol{A}_t^{-1}} + \beta. \qquad (6)$$



Proof. The decomposition into 3 terms results from Lemma 1. Then,

- Term I results from Lemma 2
- Term II results from Assumption 1
- Term III results from Lemma 4

□

Future Work

- Elaborate on bound for UCB
    - Currently depends on constants like C_opt and H_ViT
- Regret bound Analysis
- Rerun experiments with bigger models if given more compute

Thank You