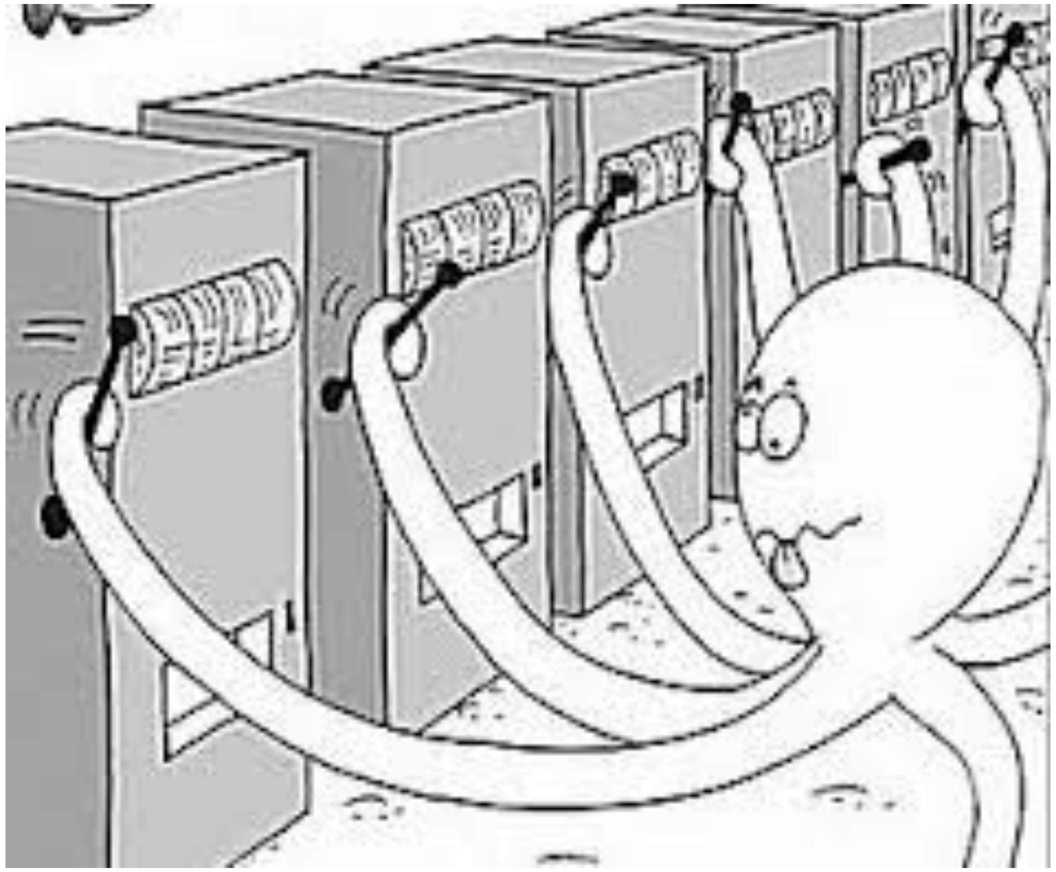
A Bandit Approach to Personalized News Article Recommendation

Qingyun Wu

News Recommendation Cycle

A K-armed Bandit Formulation

- A gambler must decide which of the K non-identical slot machines(we called them arms) to play in a sequence of trails in order to maximize total reward.



News Website <—> gambler
Candidate news articles <—> arms
User Click <—> Reward

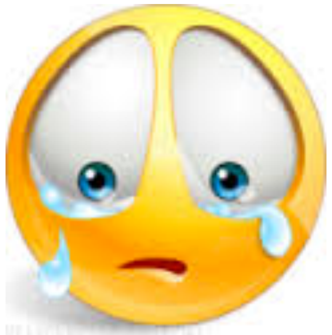
How to pull arms to maximize reward?

How to select articles to serve users to maximize user clicks

Ideal Solution

Pick $\arg\max_a \mu_a$

But we DO NOT know the mean.



Let's estimate it

Choices	X_1	X_2	X_3	X_4	X_5	X_6	...
a_1					1	1	
a_2	0		1	0			
...							
a_k		0					

Time

Exploitation VS. Exploration

Exploitation: pull an arm for which we current have the highest estimate of mean of reward

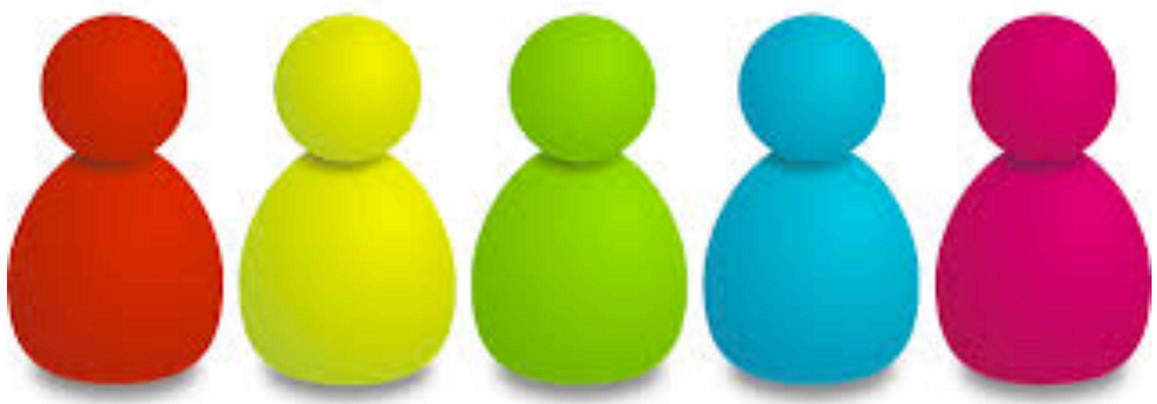
Exploration: Pull an arm we never pulled before

Not only look at the mean, but also the confidence!

Pick $\arg\max_{a}(\hat{\mu}_a + \alpha * UCB)$

UCB1

$$\arg\max_a(\hat{\mu}_a + \sqrt{\frac{2\ln T}{n_a}})$$



LinUCB (Contextual)

- Article feature:URL categories, topic categories
- User feature: demographic information, geographic features, behavioral categories

LinUCB(Contextual)

Assumption

$$E(y_{t,n} \mid x_{t,n}) = x_{t,n}^T \theta_n$$

Article Feature Vector

User preference

Parameter Estimation

$$\hat{\theta}_n = A^{-1} b$$

$$A_n = \lambda I + \sum_t x_{t,n} x_{t,n}^T \qquad b_n = \sum_t y_{t,n} x_{t,n}$$

Pick $$\arg\max_a \left(x_{t,n}^T \hat{\theta}_n + \alpha \sqrt{x_{t,n}^T (D_n^T D_n + I_d) x_{t,a}} \right)$$

From LinUCB to Collaborative-LinUCB

If user i and user j are connected by an edge, $W_{ij} > 0$

Otherwise $W_{ij} = 0$

W

$\sum_{i=1}^{N} W_{ij} = 1$ $\sum_{j=1}^{N} W_{ij} = 1$

Assumption

$$E(r_{t,n} \mid x_{t,n}) = x_{t,n}^T \theta_n \longrightarrow E(r_{t,n} \mid x_{t,n}) = x_{t,n}^T \sum_{j}^{N} W_{nj} \theta_j$$

Collaborative-LinUCB

Parameter Estimation

$$\hat{\theta}_n = A_n^{-1} b_n$$

$$A_n = \lambda I + \sum_{m=1}^{N} W_{mn}^2 \sum_t x_{tm} x_{tm}^T \qquad b_n = \sum_{m=1}^{N} W_{mn} \sum_t (x_{tm} y_{tm} - x_{tm} x_{tm}^T \sum_{j \neq n}^{N} W_{mj} \theta_j^U)$$

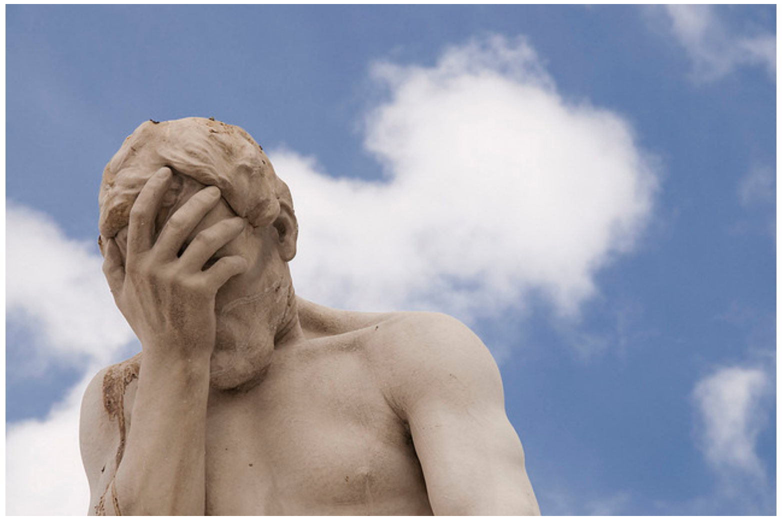
Make a choice

$$\arg\max_a (x_{tn}^T \sum_{j=1}^{N} \hat{\theta}_{nj} + \alpha \sqrt{x_{tn}^T \sum_{j=1}^{N} W_{nj} A_j^{-1} x_{tn}})$$

Performance Evaluation

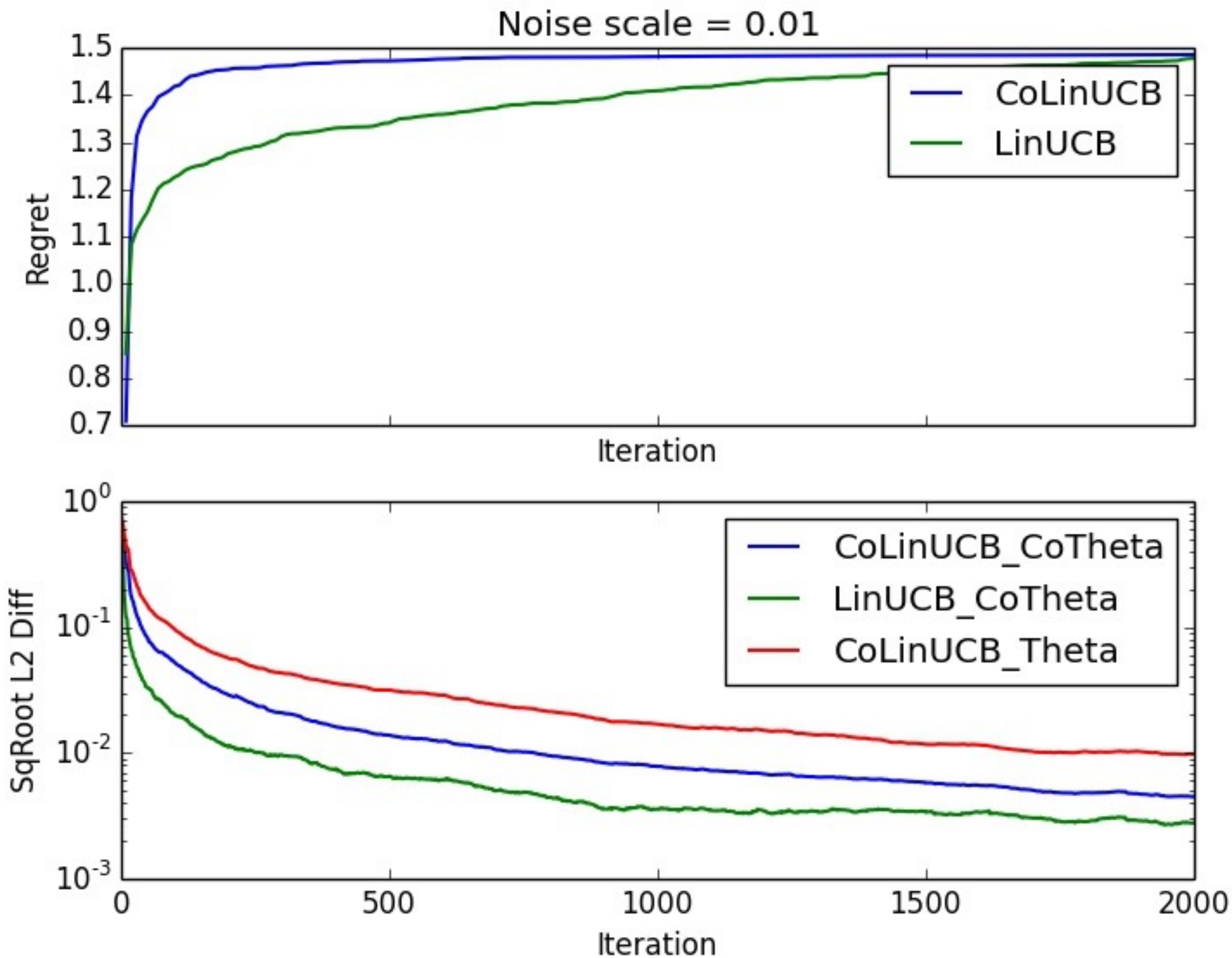
Measurement criteria

Regret

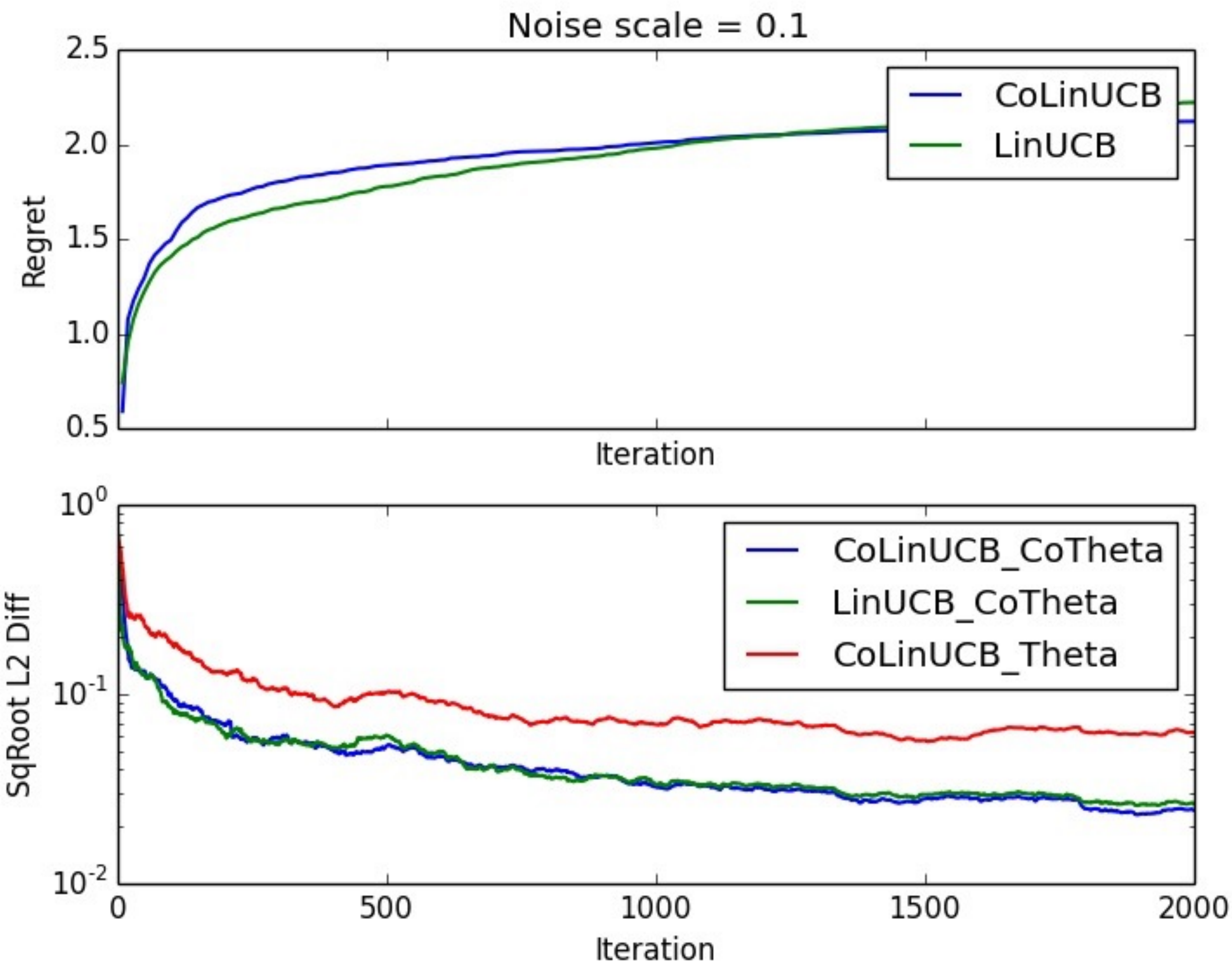


$$R_A(T) = E[\sum_t r_{t,a_t^*}] - E[\sum_t r_{t,a_t}]$$

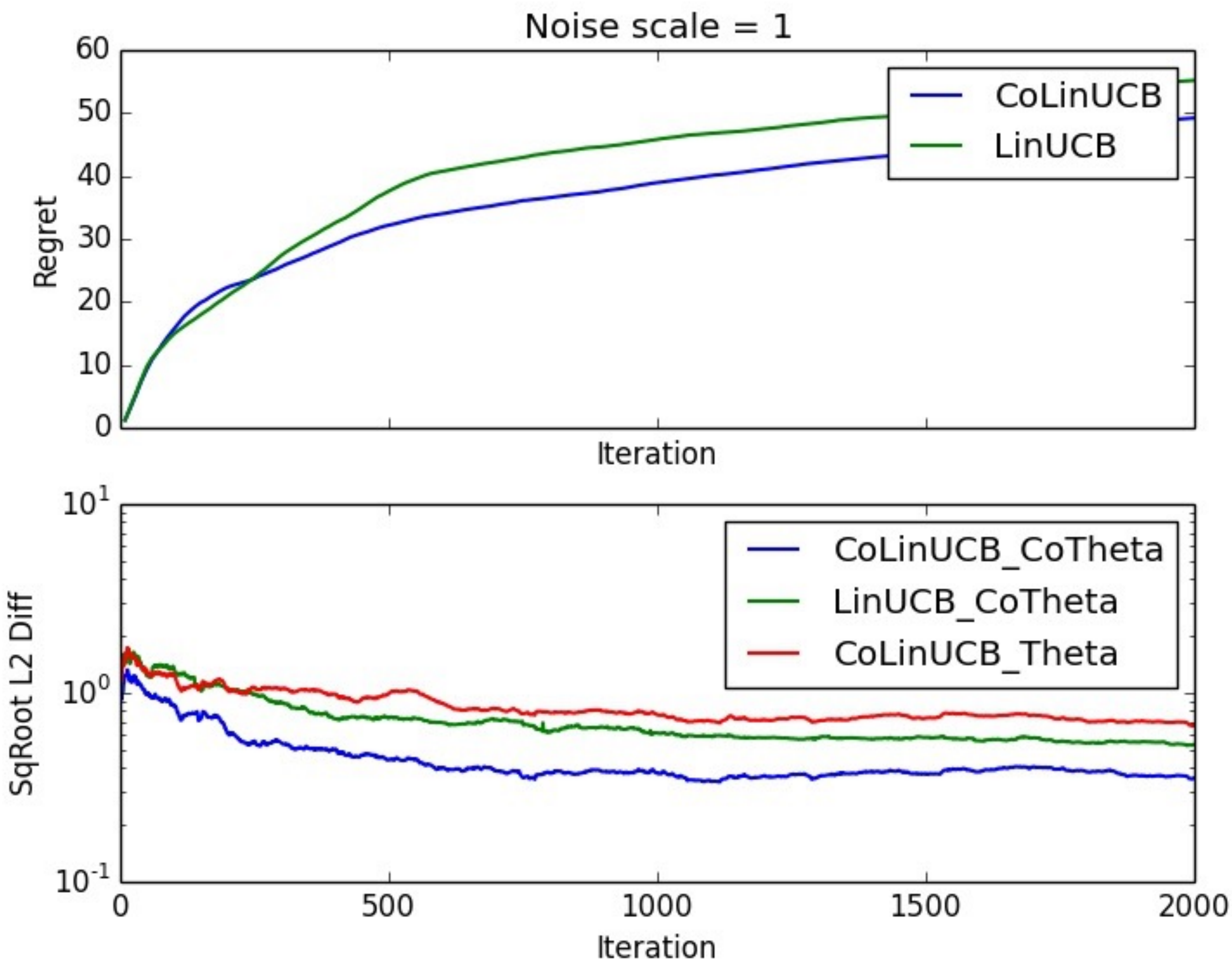
Performance Evaluation



Performance Evaluation



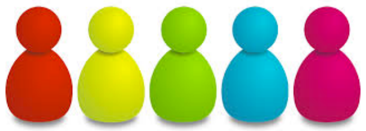
Performance Evaluation



Summary

CoLinUCB

UCB1 

LinUCB 

Q&A