**Contextual bandits often use contexts and rewards that are private information.**

For example, online shopping: context is user’s past purchases; actions are recommendations; and reward is whether user accepted recommendation.

We present a contextual linear bandit algorithm that balances learning with privacy preservation.

**Outputs (actions) don’t reveal too much about inputs (contexts, rewards)**

**DEFINITION: (ε, δ)-Differential Privacy**

Randomized algorithm $A$ is $(ε, δ)$-DP for $ε ≥ 0$ and $δ ∈ [0, 1]$ if for any subset of outputs $O$, $Pr(A(S) ∈ O) ≤ e^ε Pr(A(δ′) ∈ O) + δ$.

**DEFINITION: (ε, δ)-Joint Differential Privacy**

Relaxation of $(ε, δ)$-DP for sequential tasks.

- Context $c_t$ revealed by action $A_t$, but not by later actions.
- More suitable for contextual bandits (see lower bound below).

**Differential Privacy Requires Ignoring Context**

Any $(ε, δ)$-DP contextual bandit algorithm must have linear regret.

**Joint Differential Privacy Incurs Additional Regret**

Any $ε$-DP $k$-armed bandit algorithm must have $Ω(k \log(n)/ε)$ regret.

**Optimism in the Face of Uncertainty**

Choses “optimistic” action $X_t = \arg\max_{\theta \in \Theta} \langle \theta, x_t \rangle$.

**Differential Privacy**

Uses “noisy” versions of $V_t$ and $u_t$.
- Gaussian noise: variance $O(\log n \log(1/δ)/ε^2)$.
- Wishart noise: see details in paper.

**Regret Bounds**

- For both Wishart and Gaussian mechanisms, regret is $\mathbb{E}[\hat{R}_n] = O(\sqrt{n} \cdot d^{3/4}/\sqrt{ε})$.
- If suboptimal actions have a Δ reward gap, then $\mathbb{E}[\hat{R}_n] = O(Δ^{-1} \log^2 (n)/d^2/ε)$.
- Both cases: multiplicative polylog(1/Δ) dependence.
- See paper for details and high-probability bounds.

**Empirical Results on Synthetic Data**

- Small Gap (Δ = 0.0):
  - Machine Learning
  - Noisy Linear UCB
  - LinUCB (w/ DP)
  - LinUCB (w/ DP) (α=1)
- Large Gap (Δ = 0.1):
  - Machine Learning
  - Noisy Linear UCB
  - LinUCB (w/ DP)
  - LinUCB (w/ DP) (α=1)