A Smoothed Analysis of the Greedy Algorithm for Linear Contextual Bandits

Sampath Kannan
Jamie Morgenstern
Aaron Roth
Bo Waggoner
Z. Steven Wu

University of Pennsylvania
Georgia Tech
University of Pennsylvania
Microsoft Research, NYC
University of Minnesota

Neural Information Processing Systems, December 2018
Linear contextual bandits

Model for repeated decisionmaking:

options

\( x_1 \)

\( x_2 \)

contextual information about this option

\( x_3 \)
Linear contextual bandits

Model for repeated decisionmaking:

<table>
<thead>
<tr>
<th>options</th>
<th>linear functions</th>
<th>expected reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>$x_2$</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>$x_3$</td>
<td></td>
<td>?</td>
</tr>
</tbody>
</table>

Initially unknown
Linear contextual bandits

Model for repeated decisionmaking:

<table>
<thead>
<tr>
<th>options</th>
<th>linear functions</th>
<th>expected reward</th>
<th>observed reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td></td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>$x_2$</td>
<td></td>
<td>?</td>
<td>57</td>
</tr>
<tr>
<td>$x_3$</td>
<td></td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

noise
Greedy algorithm

Each step: max estimated reward (pure exploitation)

In the worst case: arbitrarily bad performance!

⇒ Exploration seems necessary...
Smoothed Analysis

Suppose there is some randomness in the world...

e.g. Normal(0, $\sigma^2$)
Results

**Theorem.** With a small amount of training data, the Greedy algorithm achieves good performance.

Builds on Bastani, Bayati, Khosravi (2017).
**Theorem.** With a small amount of training data, 
\[ n = \text{poly}(1/\sigma, 1/\min_i \| \beta_i \|), \]
the Greedy algorithm achieves good performance.
\[ \text{Regret} \leq O(\sqrt{T}) \]
Results

Theorem. With a small amount of training data,
\[ n = \text{poly}\left(1/\sigma, 1/\min_i \|\beta_i\|\right), \]
the Greedy algorithm achieves good performance.

Regret \( \leq O(\sqrt{T}) \)

Theorem. In the single parameter setting \((\beta_i = \beta)\),
with no initial training data, Greedy achieves

Regret \( \leq O(\sqrt{T}) \)
Motivation and future work

(1) Understand *when exploration is necessary*

(2) Understand *myopic decisionmaking*:
   - Incentives
   - Fairness/ethics (medical treatments)

Thanks!