Good Morning Colleagues

- Are there any questions?
Logistics

- Next step: literature surveys
Logistics

- Next step: literature surveys
  - Build on proposal
Logistics

- Next step: literature surveys
  - Build on proposal
- Next week’s readings
Logistics

- Next step: literature surveys
  - Build on proposal
- Next week’s readings
  - Learning from human input
Exploration

- Q-learning: converges to optimal policy \textit{in the limit}
Exploration

- Q-learning: converges to optimal policy *in the limit*
- RMax: converges in polynomial time
Exploration

- Q-learning: converges to optimal policy \textit{in the limit}
- RMax: converges in polynomial time
  - Polynomial in size of state/action space and mixing time
Exploration

- Q-learning: converges to optimal policy \textit{in the limit}
- RMax: converges in polynomial time
  - Polynomial in size of state/action space and mixing time
  - Converges \textit{in theory}
**Exploration**

- **Q-learning**: converges to optimal policy *in the limit*
- **RMax**: converges in polynomial time
  - Polynomial in size of state/action space and mixing time
  - Converges *in theory*
  - Not particularly useful (scalable) in practice
Exploration

- Q-learning: converges to optimal policy *in the limit*

- RMax: converges in polynomial time
  - Polynomial in size of state/action space and mixing time
  - Converges *in theory*
  - Not particularly useful (scalable) in practice
  - Drives agent to explore *everywhere*
Exploration

• Q-learning: converges to optimal policy \textit{in the limit}

• RMax: converges in polynomial time
  • Polynomial in size of state/action space and mixing time
  • Converges \textit{in theory}
  • Not particularly useful (scalable) in practice
  • Drives agent to explore \textit{everywhere}

• TEXPLORE: avoids visiting all states
**Exploration**

- Q-learning: converges to optimal policy *in the limit*

- RMax: converges in polynomial time
  - Polynomial in size of state/action space and mixing time
  - Converges *in theory*
  - Not particularly useful (scalable) in practice
  - Drives agent to explore *everywhere*

- TEXPLORE: avoids visiting all states
  - No theoretical guarantees
Exploration

- Q-learning: converges to optimal policy *in the limit*
- RMax: converges in polynomial time
  - Polynomial in size of state/action space and mixing time
  - Converges *in theory*
  - Not particularly useful (scalable) in practice
  - Drives agent to explore everywhere

- TEXPLORE: avoids visiting all states
  - No theoretical guarantees
  - Can work well *in practice*
Discussion

- Which is more interesting? An algorithm with theoretical guarantees? Or one that can work in practice?