TD-LAMBDA AND Q-LEARNING

Kevin Chen and Zack Khan

University of Maryland
CMSC389F: Reinforcement Learning, Spring 2018

LEARNING GOALS

- Cap off TD-Learning in context
- Understand TD-Lambda and Eligibility Traces
- Understand basic intuition of Q-Learning
- Begin coding

PREDICTION

Question: Why do we want to know the value function for a policy?
Answer: To know how good a policy is, so we can get better ones… AI

PREDICTION

Question: How do we find out the value function for a specific policy?
Answer: Policy Evaluation (DP)

Question: What if we don't know the environment dynamics?
Answer: Monte Carlo (Sampling) or
Temporal-Difference Learning (Sampling + DP)

MONTE CARLO RECAP

- Estimate the value at a state by running a full episode
  - Sum up the cumulative returns
    - First-visit vs Every-visit

TD-LEARNING RECAP

- Bootstrap the value function
- Make an initial estimate of the value at a state
- Take one step (giving us reward and value of next state)
- Update our previous estimate using our new information

TD VS MC

MC has **high va**

MC has **no bias**, TD has **bias**

verge to the real value function *(won't prove)*

roperty (faster than MC)

TD VS MC

MC has **high variance,** TD has **low variance**

MC has **no bias**, TD has **bias**

Both converge to the real value function *(won't prove)*

TD **abuses markov** property (faster than MC)

NEW STUFF

sort of

TD LAMBDA

Recall n-step TD-Learning

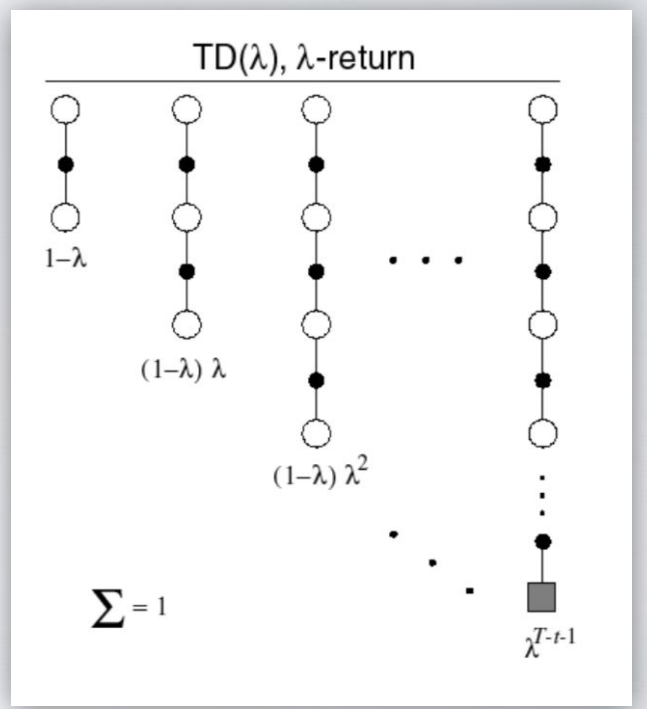
How do we harness the power of all n steps???

TD LAMBDA

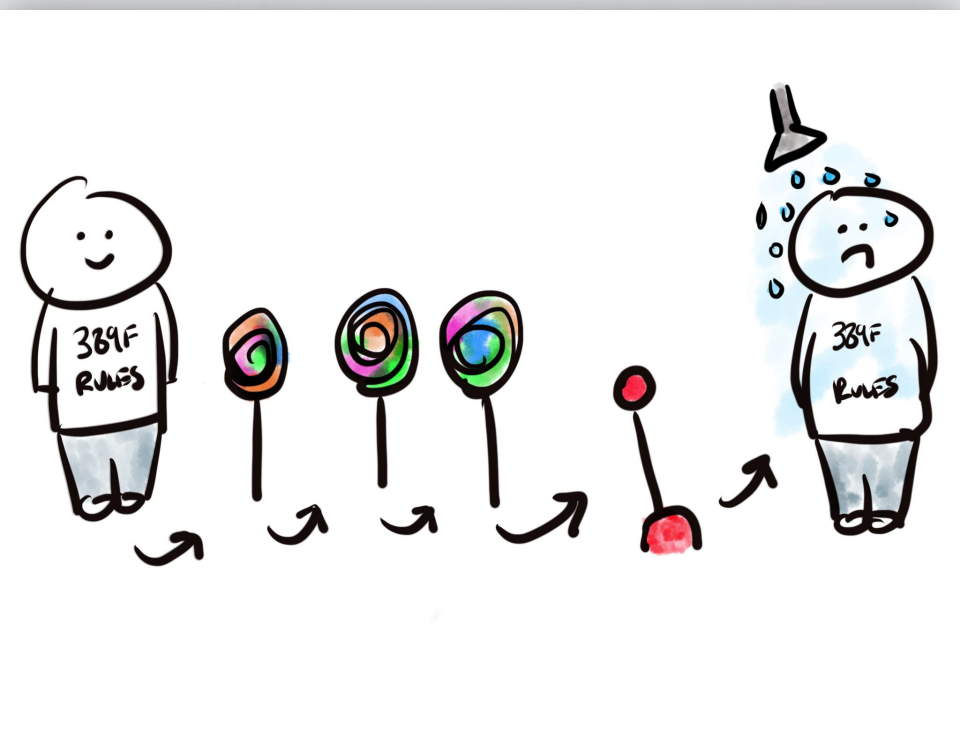
Recall n-step TD-Learning

How do we harness the power of all n steps???

AVERAGE THEM!
(discount factor style)



CREDIT ASSIGNMENT



ELIGIBILITY TRACES

Frequency Heuristic: Assign credit to most frequent states

Recency Heuristic: Assign credit to most recent states

Eligibility Traces combines both heuristics

1. Increase eligibility trace on visit
2. Exponentially decay over time

ELIGIBILITY TRACES

Frequency Heuristic: Assign credit to most frequent states

Recency Heuristic: Assign credit to most recent states

Eligibility Traces combines both heuristics

1. Increase eligibility trace on visit
2. Exponentially decay over time

Update Value Function proportionally to the Eligibility Trace! (how much was my current state the cause of the error)

ELIGIBILITY TRACES

Frequency Heuristic: Assign credit to most frequent states

Recency Heuristic: Assign credit to most recent states

Eligibility Traces combines both heuristics

1. Increase eligibility trace on visit
2. Exponentially decay over time

Update Value Function proportionally to the Eligibility Trace! (how much was my current state the cause of the error)

This is an alternative approach to TD-Lambda

NEW STUFF

briefly

Q-LEARNING 101

Off-policy Learning: "Looking over someone's shoulder". Learning about policy A by sampling from policy B

Q is a lookup table storing rewards for every state-action pair. Call these Q-values.

We learn these Q-values using Q-Learning

Q-LEARNING 101

1. Initialize Q to zero for all state-action pairs
2. Take your first action from some state s
3. You're now in state **s'**. Pick its max Q-value.
4. Set the new Q of state s as your reward + the discounted max Q value of your new state
5. Repeat
6. The optimal policy = greedily taking action with highest reward

SOME CODING
(see code from class here)

HW/GRADES/PROJECT

NEXT TIME

Q-LEARNING! **(not just 101)**

What the entire course has been leading up to...

HOMEWORK

Short coding problem set

Released on Monday

MIDTERM GRADES

Released next week!

PROJECT

Start making groups of 2-4.

Have one member of the group email cmsc389f@gmail.com with everyone's names by **next Thursday.**

Remaining students will be assigned groups.

We will talk about potential topics next week.

PROJECT

We will have several *coding lectures* where we code up RL algorithms in-class
(Monte Carlo, Q-Learning)

We will have a few research-oriented lectures including:

1. Useful plots to generate for RL
2. Quick overview of landmark RL papers
3. Quick overview of DQN

PROJECT

A 4-6 page research paper on an RL topic of your choice.

Introduction (½ page): Summary of the paper

Background (1 page): Necessary information to understand your paper (i.e lecture material)

Results (1-2 pages): Tables and plots about your results

Discussion (1 page): Discuss your results

Conclusion (½ page): Conclude everything

PROJECT

A 4-6 page research paper on an RL topic of your choice.

Introduction (½ page): Summary of the paper

Background (1 page): Necessary information to understand your paper (i.e lecture material)

Results (1-2 pages): Tables and plots about your results

Discussion (1 page): Discuss your results

Conclusion (½ page): Conclude everything

PROJECT

A 4-6 page research paper on an RL topic of your choice.

Introduction (½ page): Summary of the paper

Background (1 page): Necessary information to understand your paper (i.e lecture material)

Results (1-2 pages): Tables and plots about your results

Discussion (1 page): Discuss your results

Conclusion (½ page): Conclude everything

PROJECT

A 4-6 page research paper on an RL topic of your choice.

Introduction (½ page): Summary of the paper

Background (1 page): Necessary information to understand your paper (i.e lecture material)

Results (1-2 pages): Tables and plots about your results

Discussion (1 page): Discuss your results

Conclusion (½ page): Conclude everything

PROJECT

A 4-6 page research paper on an RL topic of your choice.

Introduction (½ page): Summary of the paper

Background (1 page): Necessary information to understand your paper (i.e lecture material)

Results (1-2 pages): Tables and plots about your results

Discussion (1 page): Discuss your results

Conclusion (½ page): Conclude everything

ENJOY THE WEEKEND!