CS-8803 RLDM Project 1: Temporal Difference Learning (sutton 88)

Nikhil Gajendrakumar

Goal : Replicate results of Sutton 88 i.e. TD methods are better.

Conventional prediction-learning methods

1. Assign credit by means of the difference between predicted and actual outcomes.

2. Learner is asked to associate pairs of items. When later presented with the first item of a pair, the learner is supposed to recall the second.

3. Supervised learning methods

4. w is updated only once for each complete observation.

Temporal Difference Learning

1. Assign credit by means of the difference between temporally successive predictions.

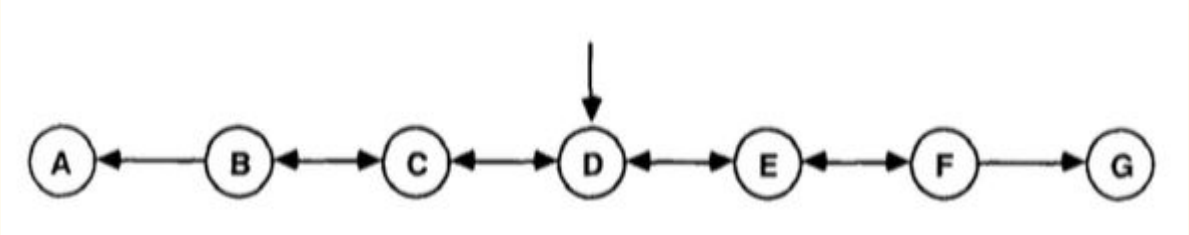
2. Can be implemented incrementally and therefore requires far less computational power.

3. Sensitive to changes in successive predictions rather than to overall error between predictions and the final outcome.

4. TD (lambda).

5. Lambda $= 1 \Rightarrow$ Supervised Learning procedure.

A random-walk example



1. Can be represented as a Markov decision process.
2. Used numpy.np.rand() to simulate random walk

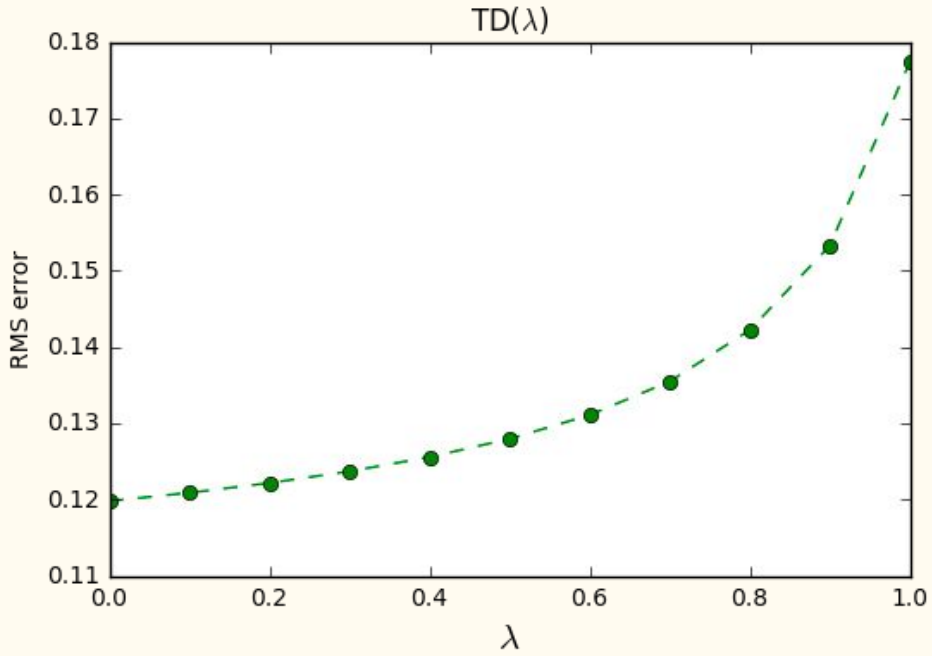
Two computational experiments

1. Repeated presentations
2. One presentation
3. Ideal predictions
4. RMS error

Repeated presentations

Procedure

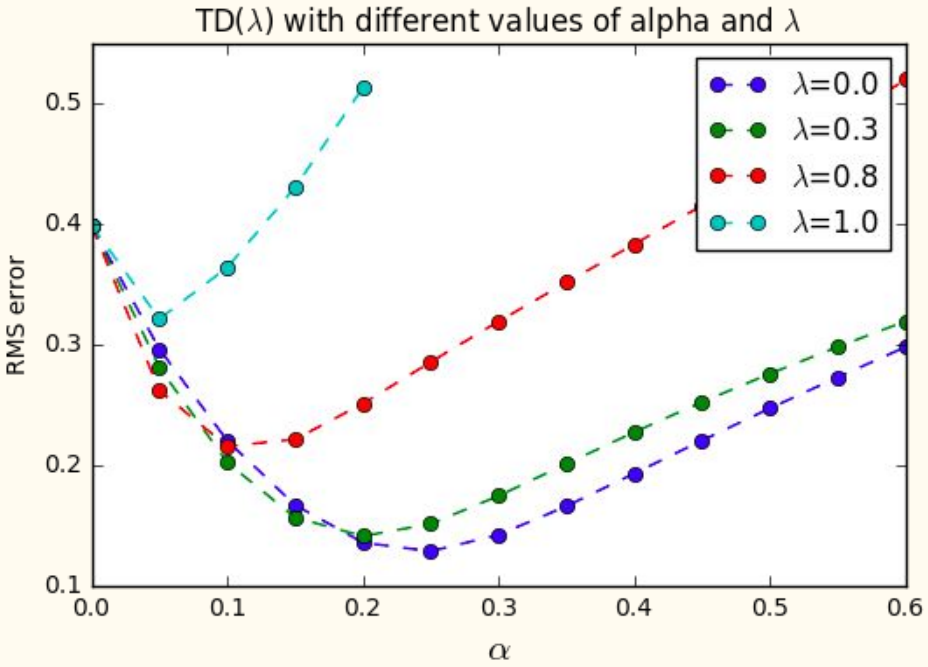
1. 100 training sets, with 10 sequences per set.
2. Find RMS error for different lambda values



One presentations

Procedure

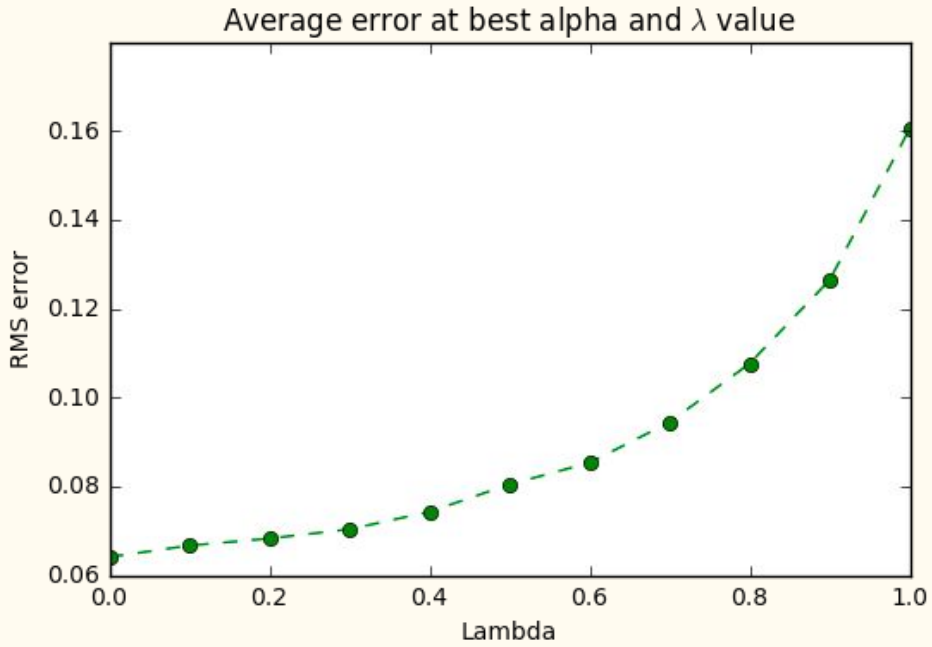
1. 100 training sets, with 10 sequences per set.
2. Effect of alpha
3. Deviation from Sutton output



Best error level

Procedure

1. 100 training sets, with 10 sequences per set.
2. Optimal alpha for each lambda
3. Deviation from Sutton output



Conclusion

From the results of above 3 experiments, we can clearly observe that TD methods provide better prediction accuracy than conventional-prediction learning methods (example: TD(1)). The results of experiment 2 and 3 deviate a little compared to results given in Sutton, but even with little deviation in result, we prove the TD methods perform better on multi-step prediction problems than other supervised learning procedures.

Thank you