Reinformcement Learning Theory Learning

Yue Wang

MSRA

Abstract

a simple note for RL theroy

Keywords:

1. Introduction

RL theory is very long history and these days

- Bullet point one
- Bullet point two

1. Numbered list item one
2. Numbered list item two

2. Notation and Formatting

some notations:

$\mathcal{S}$	the state space
S_t	the state at time t, stochastic
s_t	the state at time t, actual
s	the state, actual
R_t	the reward at time t, stochastic
r_t	the reward at time t, actual
r	the reward, actual

Treatments	**Response 1**	**Response 2**
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296

Table 1: Table caption

2.1. Subsection Two

Donec eget ligula venenatis est posuere eleifend in sit amet diam. Vestibulum sollicitudin mauris ac augue blandit ultricies. Nulla facilisi. Etiam ut turpis nunc. Praesent leo orci, tincidunt vitae feugiat eu, feugiat a massa. Duis mauris ipsum, tempor vel condimentum nec, suscipit non mi. Fusce quis urna dictum felis posuere sagittis ac sit amet erat. In in ultrices lectus. Nulla vitae ipsum lectus, a gravida erat. Etiam quam nisl, blandit ut porta in, accumsan a nibh. Phasellus sodales euismod dolor sit amet elementum. Phasellus varius placerat erat, nec gravida libero pellentesque id. Fusce nisi ante, euismod nec cursus at, suscipit a enim. Nulla facilisi.

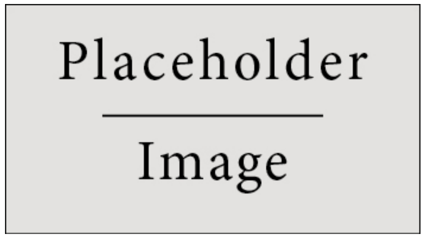

Figure 1: Figure caption

Integer risus dui, condimentum et gravida vitae, adipiscing et enim. Aliquam erat volutpat. Pellentesque diam sapien, egestas eget gravida ut, tempor eu nulla. Vestibulum mollis pretium lacus eget venenatis. Fusce gravida nisl quis est molestie eu luctus ipsum pretium. Maecenas non eros lorem, vel adipiscing odio. Etiam dolor risus, mattis in pellentesque id, pellentesque eu nibh. Mauris nec ante at orci ultricies placerat ac non massa. Aenean imperdiet, ante eu sollicitudin vestibulum, dolor felis dapibus arcu, sit amet fermentum urna nibh sit amet mauris. Suspendisse adipiscing mollis dolor quis lobortis.

$$e = mc^2 \tag{1}$$

3. RL algorithms

4. RL convergence theory

4.1. counterexample

In Sutton and Barto [1, chap 11.3] the authors give an intuitive conclusion about when these algorithms will divergence :

the danger of instability and divergence arises whenever we combine three things:

1. training on a distribution of trainsition other than that naturally generated by the process whose expectation is being estimated(e.g. off-policy learning)
2. scalable function approximation (e.g. li)

4.1.1. counterexample1

In Bertsekas [2] Dimitri [3]

$E[\theta_(t+1)|\theta_t]$

References

[1] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, 2011.

[2] D. P. Bertsekas, A counterexample to temporal differences learning, Neural Computation 7 (1995) 270–279.

[3] P. Dimitri, A Counterexample to Temporal Differences Learning, pdfs.semanticscholar.org (????).