HW3 Report

Y.S.S.V Sasi Kiran

November 12, 2018

1 Plot

The plot showing the mean squared TD error for gridWorld with tabular policy and cartPole using 3rd and 5th order fourier basis is shown below.

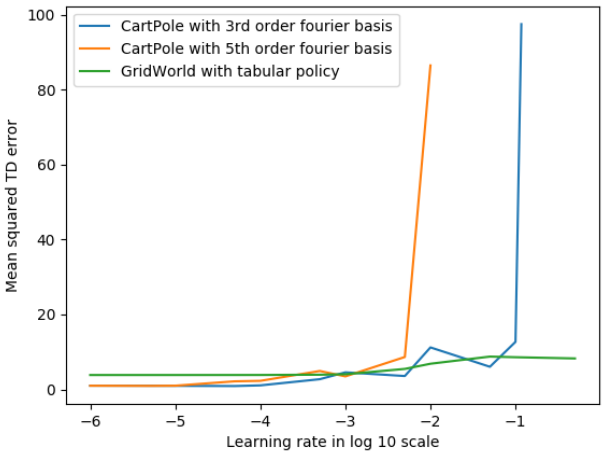

Figure 1.1: The above figure shows the variation in mean squared TD error (Y-axis) with variation in learning rate (X-axis). The learning rate is plotted in a log10 scale.

2 Observations

We observed that cartPole with 3rd order fourier basis diverges at $\alpha > 0.15$. Similarly, cartPole with 5th order fourier basis diverges at $\alpha > 0.012$. This divergence results in Nan being obtained as mean squared TD error due to overflow for learning rates greater than the specified. This is consistent with the theory because we know that TD converges for linear function approximation

with decaying step sizes or small enough step sizes. When step size is constant, then for higher step sizes TD may diverge, which is the result observed.

Additionally, we also observed an inverse relation between the number of basis expanded features and minimum step size beyond which TD diverges i.e. as the number of features increase, TD starts to diverge for smaller step sizes. Let L be the minimum step size beyond which TD diverges. Then the following table shows the observations for different orders of fourier series.

Fourier order	Value of L
1	0.999
3	0.15
5	0.012
7	0.002
9	0.0006

Table 1: The above table shows the variation in minimum step size beyond which TD diverges with varying fourier order for cart pole domain.

I thought that 0.001 is a small enough step size for TD to converge, but never expected that TD's convergence changes with the number of features in basis expansion for cart pole domain. However, gridWorld doesn't diverge even for large step sizes (like 0.999) but the convergence point changes with different random seeds. This is also according to theory which says that with constant step size, TD for tabular converges only **in mean** to v^{π}.

Furthermore, we observe convergence of cart pole with decaying step size. Specifically, the step size is decayed as $\alpha \leftarrow \alpha/\sqrt{k}$ for the kth update. This decay follows square summable but not summable criteria. This is also consistent with the theory for convergence of linear function approximation. One observation here was that k is the total number of updates and not the time step in the episode (so that it doesn't reset after each episode).