Lent Update: Week 1

Alex Darch
Supervisor: Dr Glenn Vinnicombe

January 23, 2019

1 Results

All graphs below are produced from the same setting:

- 18MCTS recursions at each step.
- The first 15 steps of training episodes are sampled from the MCTS-Improved action probability for exploration - greedy actions are then taken (this is done by suragnair: not sure if applicable to control?).
- 20 training episodes
- 15 greedy/test episodes, the mean loss is then compared with the mean loss of the best current set of test episodes by the current best policy - if mean*0.95 ¿ best mean then update.
- Losses are trained in mini-batches of 8.

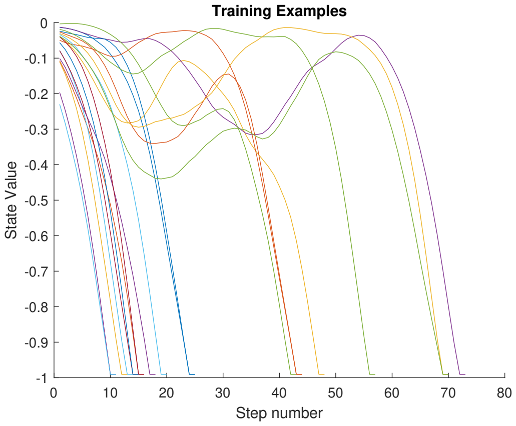

Figure 1: Training Examples

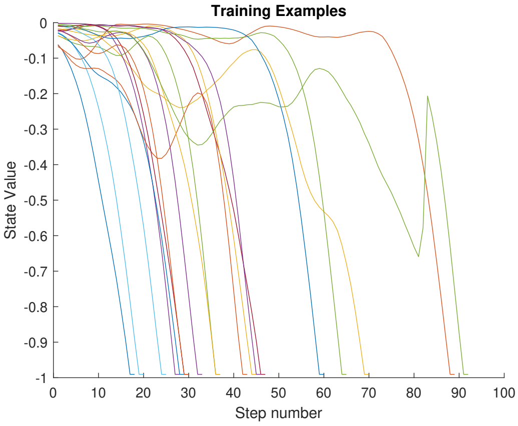

Figure 2: Training Examples

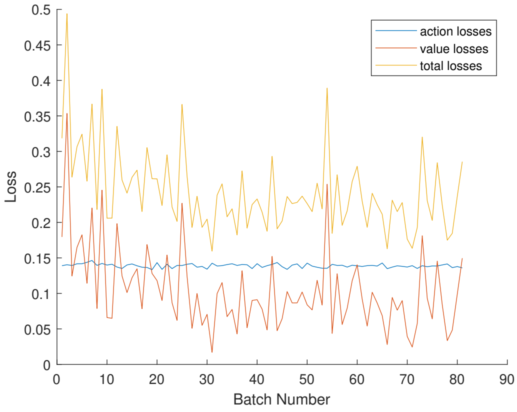

Figure 3: Loss from Neural Network Training

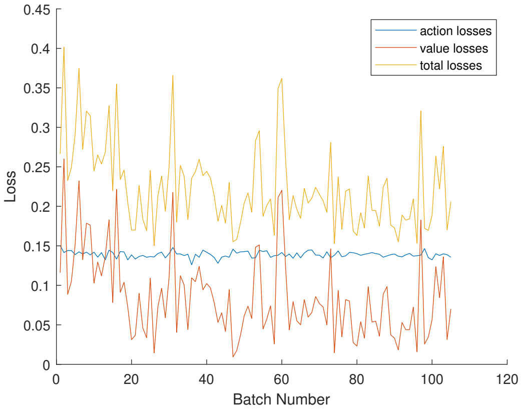

Figure 4: Loss from Neural Network Training

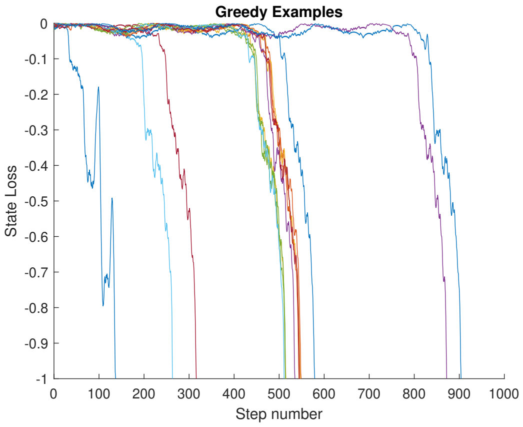

Figure 5: default

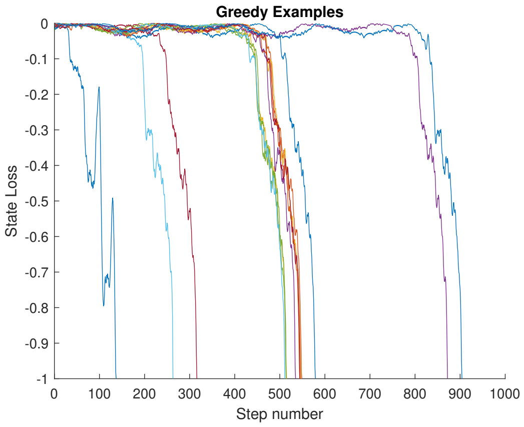

Figure 6: default