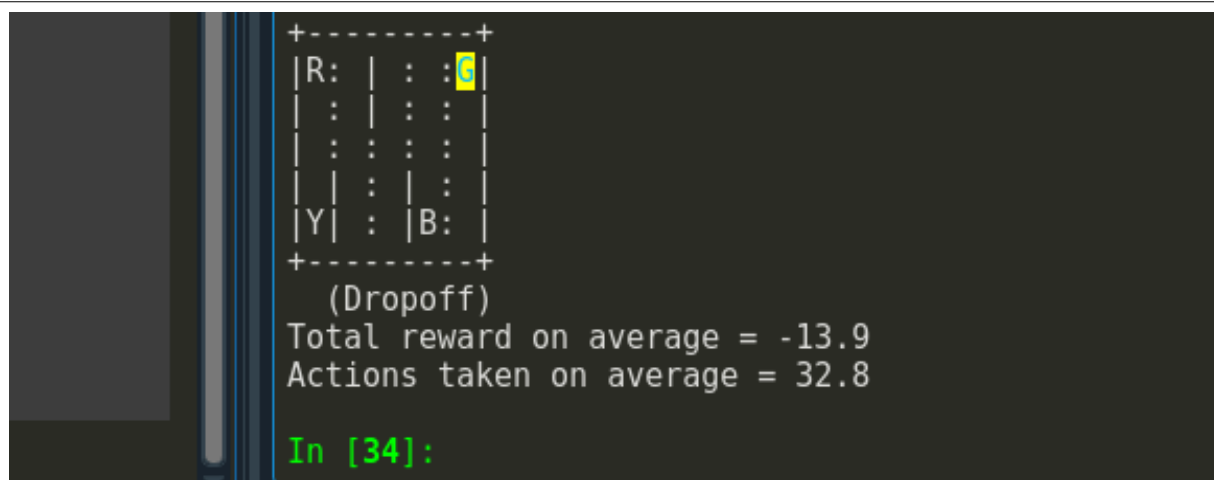


OpenAI Gym – Taxi (V3) Environment (50 points)

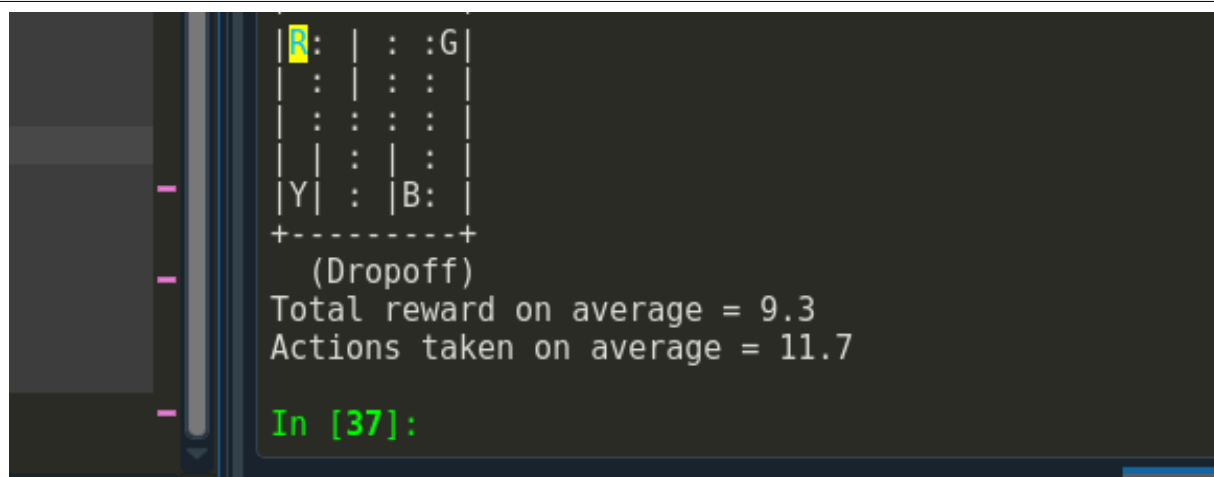
With 1000-episodes I sometimes got good results (average reward ~ 8) but sometimes the agent got stuck by making repeated actions or making same actions forever (average reward < 0). I got over that issue by increasing the number of episodes when average stabilized to ~ 8 .

I used epsilon greedy strategy when training the agent. In case of 1000 episodes I used constant value of 0.1 for epsilon. For larger count of episodes ($\sim 100\,000$) I initialized epsilon to 0.5 and decayed it by multiplying it by 0.99 after each episode, or set the minimum value for it (0.1).

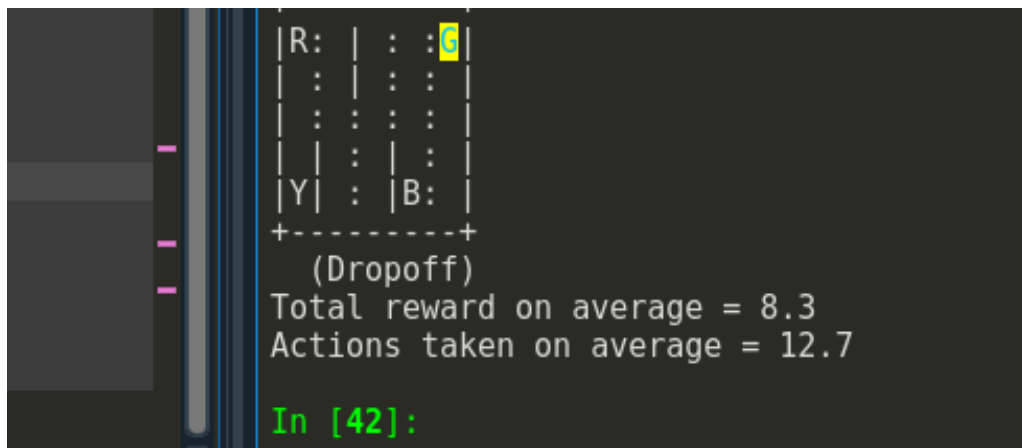
For other hyperparameters, I set alpha and gamma both to 0.9.



Example case when agent got stuck when evaluating its average performance on 10 episodes



Example case when agent performs well.



Average (and stable) performance when agent was trained on larger number of episodes (100 000).