Reinforcement Learning (SS18) - Exercise 4

Daniel Hennes

24.05.2018 (due 30.05.2018)

1. Give an example of an environment where you would use a Monte Carlo method to *learn* the value function rather than using dynamic programming to compute it. Explain why. (Note that in both cases a model is required.)
2. Consider the version of blackjack introduced in the lecture (Example 5.1 from Sutton and Barto). Implement first-visit Monte Carlo prediction (slide 8) and reproduce the figures on slide 11 for the given policy:
   stick if sum ≥ 20, else *hit*.
3. Implement Monte Carlo ES and obtain the optimal policy and state-value function for blackjack.

You may use the OpenAI Gym `Blackjack-v0` environment:

```python
# https://github.com/openai/gym
# https://github.com/openai/gym/blob/master/gym/envs/toy_text/blackjack.py
import gym
env = gym.make('Blackjack-v0')
obs = env.reset()
done = False
while not done:
    player_sum, dealer_card, useable_ace = obs
    obs, reward, done, _ = env.step(0 if player_sum >= 20 else 1)
```