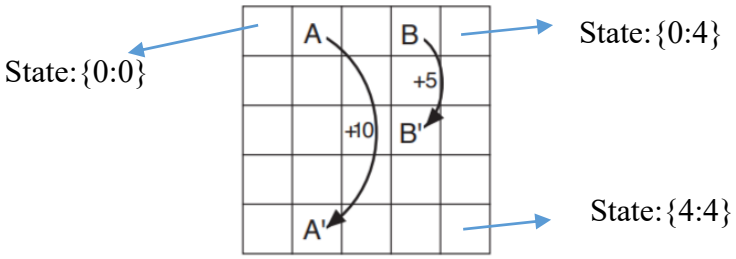
機器學習實務與應用

Homework #5
Due 2021 May 27 9:00 am

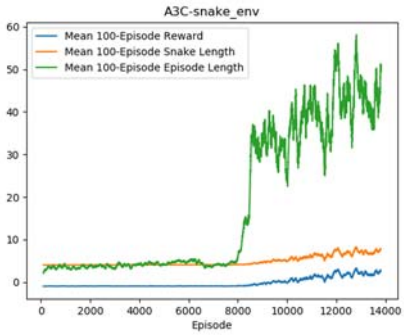
(一)針對底下的 grid world 環境，每個 cell 代表 environment 的 state。在每個 cell，可以有四個可選擇的 actions：north, south, east, west。這些 action 會將 agent 依據 action 行進方向將 agent 移到下一個 cell。在邊緣的 cell 時，若 action 會將 agent 帶出這個 grid world 時，則 agent 將不會移動，但會收到 reward= −1。另外，若 agent 處於特殊 state A 時，所有 actions 都會獲得 reward=10，同時 agent 會固定移到 A'；類似情況若 agent 處於 state B 時，所有 actions 都會獲得 reward=5，同時 agent 會固定移到 B'。除上述情況，其餘都不會獲得 reward。



1. 假定採用的 policy 為選擇四個 action 的機率是一樣的，請記算並列出每個 state 的v_π值。假設 discounted reward 的係數$\gamma = 0.9$。
2. 利用 Q-learning 使用採用最佳(reward 最大)的 policy，請記算並列出每個 state 的v_π值，及各個 state 最佳的 action。

(二)附件 ***snake_env*** 提供一個貪吃蛇的遊戲環境，針對此環境：

1. 先安裝相關套件，讓此環境可以被執行，並研究其 reward 給予的機制。
2. 請利用 RL 訓練可以自行玩這個遊戲的 agent，並請繪製如下之圖表。



3. 改變 reward 的給予方式，觀察是否會影響 agent 的訓練結果。

* 請注意上繳之作業中，請將訓練結果先行儲存，並註明清楚如何執行能得到你訓練後的 agent 操控的遊戲畫面。

(三)請尋找除了遊戲之外，Reinforcement Learning 的其他應用，請簡述該應用如何用 RL，尤其是該應用中之 reward 的設定