3. Exercise Sheet

Assignment 14  Monte Carlo method

In Monte Carlo method, if we start with a deterministic $\pi$, some/many $(s, a)$-pairs will never be visited! How can we make sure that (almost) all pairs are visited?

Assignment 15  Race track (MC)

- States: grid squares, velocity horizontal and vertical
- Rewards: $-1$ on track, $-5$ off track
- Only the right turns allowed
- Actions: $+1$, $-1$, $0$ to velocity
- $0 < \text{Velocity} < 2$ in each direction
- Stochastic: 50% of the time it moves 1 extra square up or right
- Goal: reach the finish line as fast as possible without leaving the track
- No discounting ($\gamma = 1$)
- Return for each state is the negative number of steps to go from that state
- $V(s)$: predicted negative number of steps

You start from cell 2 with no velocity

Estimated return of $-3$

You add $+1$ to horizontal velocity.

State = $(2, 0, 0)$

You stay in cell 4

Estimated return of $-3 \rightarrow 5$ steps in total

You add $+1$ to vertical velocity.

Reward $-1$

State = $(4, 0, 0)$

A wind from south moves you to cell 13 (instead of 9)

Estimated return of $-1 \rightarrow 4$ steps in total

You add $+1$ to horizontal velocity.

Reward $-1$

State = $(13, 0, 1)$

Final State = $(17, 1, 1)$
a) Complete the table below for the Race track example:

<table>
<thead>
<tr>
<th>State $s(cell, h, v)$</th>
<th>Rewards so far</th>
<th>$G_t$</th>
<th>$V(s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2, 0, 0)</td>
<td>0</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>(4, 1, 0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4, 0, 0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(13, 0, 1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(17, 1, 1)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b) Compute an iteration of Monte Carlo with $\alpha = 0.5$ for the Race track example

Iteration                               | $K = 0$ | $K = 1$, ($\alpha = 0.5$) |
----------------------------------------|---------|---------------------------|
$G_t(2, 0, 0)$                          |         |                           |
$V(2, 0, 0)$                            |         |                           |
$G_t(4, 1, 0)$                          |         |                           |
$V(4, 1, 0)$                            |         |                           |
$G_t(4, 0, 0)$                          |         |                           |
$V(4, 0, 0)$                            |         |                           |
$G_t(13, 0, 1)$                         |         |                           |
$V(13, 0, 1)$                           |         |                           |

Assignment 16  Race track (TD)

Compute an iteration of $TD(0)$ with $\alpha = 0.5$ for the Race track example

<table>
<thead>
<tr>
<th>Iteration</th>
<th>$K = 0$</th>
<th>$K = 1$, ($\alpha = 0.5$)</th>
<th>Error $\delta$</th>
</tr>
</thead>
</table>
$V(2, 0, 0), R_1 = -1$|         |                           |                 |
$V(4, 1, 0), R_2 = -1$|         |                           |                 |
$V(4, 0, 0), R_3 = -1$|         |                           |                 |
$V(13, 0, 1), R_4 = -1$|         |                           |                 |