Aston University
Machine Learning

Portfolio Task 3: Reinforcement Learning

Released: 23/11/20
Due: 04/12/20, before 23:59

Instructions:
In this assessed task, you will be answering questions related to the third family of machine learning techniques covered in this module: reinforcement learning. The aim of these tasks is to test your understanding of the foundations of these techniques and of how to apply them to well-specified tasks.

Details:
Follow the instructions below to complete the portfolio task. The task requires you to carry out some calculation and to provide short written justifications of your choices, of maximum 250 words. You are expected to submit your answers as either a word document or a PDF. You are not expected to write any code to support your answers below but, if you do, please include it in your submitted document as an appendix.

Marking:
This portfolio task is worth 15% of the overall module mark.

The mark scheme for the task is as follows:

- **50-59** The solution to the first sub-task shows understanding of the relationship between utility and policy, but there are some errors in calculation resulting in the calculated policy being incorrect. In the second sub-task, a reasonable solution methodology has been proposed, but the discussion and justification shows some errors or gaps in understanding.
- **60-69** The calculations for sub-task 1 are correct and are used to infer the correct policy. A well-reasoned solution methodology has been proposed in sub-task 2 and the justification and discussion show clear understanding of the problem and of relevant algorithms.
- **70-79** As above, but the methodology proposed for sub-task 2 is carefully designed. All important considerations receive coverage and all show clear understanding.
- **80+** As above, but with additional evidence (for the second sub-tasks) of some or all of: attention to quality, thorough understanding of algorithm design, excellent justification.

No specific descriptors are provided for marks below the threshold of 50. Marks in the range **0-49** are allocated where the submitted work has not reached the expectation for the threshold descriptor.

Sub-task 3.1:
The below grid is a variant of the grid world problem studied in unit 8. It has three inaccessible states (the black shaded squares) and four terminal states (the grey shaded squares).

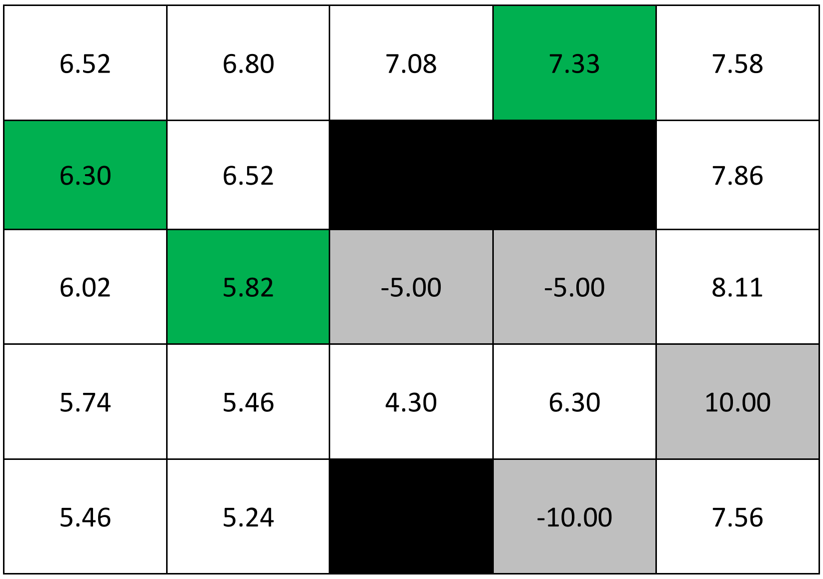
In each state, the agent:

- receives a reward of -0.2 in a non-terminal state or of the value indicated below if in a terminal state,
- ends the game if it is in a terminal state,
- otherwise, it must choose to try and move to one of the neighbouring states (the horizontally or vertically adjacent states. Diagonal movement is not permitted).

After attempting to move, the agent:

- reaches the state it was attempting to move to with probability 0.8,
- fails and makes a perpendicular move with probability 0.2 (each direction is equally likely),
- if, as a result of this, the agent attempts to move outside of the grid or to an inaccessible state, it instead remains where it is.

In the diagram below, the number in each state shows the (expected) utility of that state under some policy, rounded to two decimal places. Using this information, draw a diagram to show the policy when to deriving these utility values. For each of the three states highlighted in green, show how you determined the policy action for that state.



6.52	6.80	7.08	7.33	7.58
6.30	6.52			7.86
6.02	5.82	-5.00	-5.00	8.11
5.74	5.46	4.30	6.30	10.00
5.46	5.24		-10.00	7.56

Sub-task 3.2:
You are asked to work on a different variant of the grid world problem. In this variant, you do not have access to a diagram of the form given above – showing which states are adjacent – and nor do you know the rules governing how agents probabilistically transition between states after choosing an action. You are asked to propose a method for finding the optimal policy in this setting.

One of your colleagues proposes using temporal difference learning (TDL) to solve the problem. Explain to them why TDL would not be well suited to solving this problem.

Drawing on the techniques discussed in units 8 and 9, propose, with justification, a method for solving this problem. In your description, highlight any factors you consider important in solving the problem successfully.