RL Tutorial 1

Scott Brownlie

Problem 1

In supervised learning we have a training set of examples (x, y). Each x is an input and the corresponding y is its label. The label can be thought of as the correct answer for the input. The training set is generally fixed before learning and we wish to learn a mapping from inputs to labels which generalises to new unseen data from the same distribution as the training data.

In contrast, RL is goal-directed learning which is achieved through interaction with an environment. In RL we do not have a fixed training set. Instead, an agent generates data by interacting with an environment over discrete time steps. On each time step the agent is presented with a representation of its current state, x say, and it has to decided how to act. Unlike in supervised learning, there is no label y which tells the agent how to act. Instead, the agent chooses and executes an action and receives feedback from the environment in the form of a scalar reward. The reward signal should be such that maximising the long term reward results in the agent achieving its goal. In the absence of supervised labels the agent must find the best action for each state in a trial and error fashion.

The answer given by Shaikh is flawed because it suggests that only in supervised learning do we have knowledge of the environment. In fact, in many reinforcement learning tasks we have complete knowledge of the environment, however, for all but the smallest problems learning without interaction is intractable.