UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

FACULTY *of* SCIENCE *and* ENGINEERING

Department of Computer Science
and Information Systems

End-of-Semester Assessment Paper

Academic Year:	2022-2023	Semester:	Spring
Module Title:	Deep Reinforcement Learning	Module Code:	CS6482
Duration of Exam:	2 Hours	Percent of Total Marks:	30
Lecturer(s):	J.J. Collins	Paper marked out of:	30

Instructions to Candidates:

- **Answer Question 1 and any two others.**
- Please do not use red ink

Q1 Answer ALL parts. Total marks awarded for this question: 10 marks.

a) How is experience captured, processed, stored, and sampled for training a Deep Q Networks (DQN) on Atari? Illustrate the discussion with coding fragments or pseudocode.

3 marks.

b) Describe the target used to train a DQN.
Illustrate the discussion with coding fragments or pseudocode.

3 marks.

c) What is the cause of maximisation bias in DQNs?
Describe an approach that can be used to reduce maximisation bias.

4 marks.

Q2 Answer ALL parts. Total marks awarded for this question: 10 marks.

a) Explain why the number of parameters in GoogleLeNet using Inception modules is significantly less than AlexNet - 6 million as opposed to 60 million. The answer should focus on the Inception module. Illustrate the answer with a diagram and/or rough calculations.

3 marks.

b) Ioffe and Szegedy (2015) proposed Batch Normalisation as a mechanism to reduce the impact of vanishing gradients. How many parameters in the three Batch Normalisation layers in Figure Q2? Of these, how many are trainable? Please show the calculations.

```
model = keras.models.Sequential([
keras.layers.Flatten(input_shape=[28, 28]),
keras.layers.BatchNormalization(),
keras.layers.Dense(150,activation="relu", ernel_initializer=“he_normal"),
keras.layers.BatchNormalization(),
keras.layers.Dense(100, activation="relu", kernel_initializer=“he_normal"),
keras.layers.BatchNormalization(),
 keras.layers.Dense(10, activation="softmax")])
```

Figure Q2

3 marks.

c) Describe the key concept(s) in ResNet. Include a discussion on the purpose of a kernel of size 1x1 with stride 2. Illustrate the discussion with a diagram.

4 marks.

Q3 Answer ALL parts. Total marks awarded for this question: 10 marks.

a) List the steps in the REINFORCE Algorithm.

3 marks.

b) Describe three advantages of Policy Gradient approaches.

3 marks.

c) The coding fragment is Figure Q3 is an excerpt of an implementation of the REINFORCE method for the cartpole. Provide a detailed explanation for lines 2 to 11.

4 marks.

```
L1   for iterations in range(n_Iterations):
L2       all_rewards, all_grads = play_multiple_episodes(env, n_episodes_per_update,
             n_max_steps, model, loss_fn)
L3       total_rewards = sum(map(sum, all_rewards))
L4       all_final_rewards = discount_and_normalise_rewards(all_rewards, discount_rate)
L5       all_mean_grads = []
L6       for var_index in range(len(model.trainable_variables)):
L7           mean_grads = tf.reduce.mean([final_reward * all_grads[episode_index][step][var_index]
L8               for episode_index, final_rewards in enumerate(all_final_rewards)
L9                   for step, final_reward in enumerate(final_rewards)], axis=0)
L10          all_mean_grads.append(mean_grads)
L11      optimizer.apply_gradients(zip(all_mean_grads, models.trainable_variables))
```

Figure Q3. Adapted from Maxim Lapan. Deep Reinforcement Learning. 2018. Packt Publications.

Q4 Answer ALL parts. Total marks awarded for this question: 10 marks.

a) Briefly describe the Physical Symbol System Hypothesis (PSSH).
Compare and contrast the PSSH with Machine Learning.

3 marks.

b) Compare and contrast Evolutionary Computation with Reinforcement Learning

3 marks.

c) Describe in detail one learning paradigm discussed in tutorials. For example, Inductive Decision Trees, Generative Adversarial Networks (GANs), Long Short Term Memory (LSTM), Autoencoders, etc. For the paradigm selected, describe the training algorithm, network structure where relevant, training data, and sample applications.

4 marks.

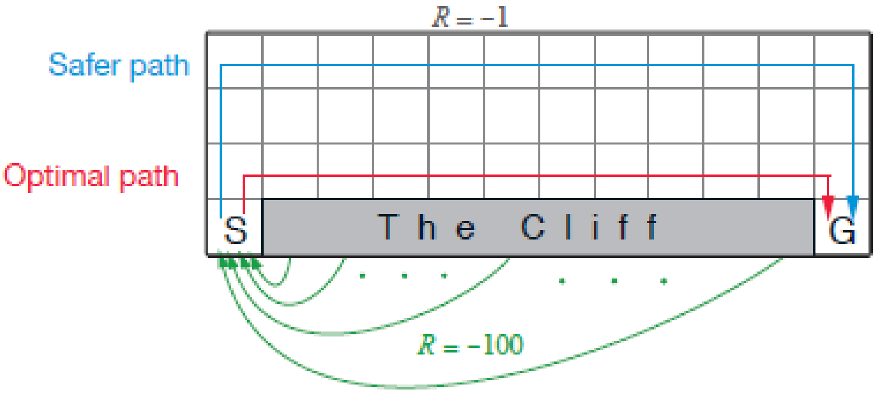

Figure Q5. Adapted from Sutton and Barto, Reinforcement Learning, 2nd Ed. 2018. MIT Press.

Q5 Answer ALL parts. Total marks awarded for this question: 10 marks.

a) What is Reinforcement Learning?
What are the key issues that an RL agent must address?

3 marks.

b) The path computed by the Q learning algorithm in Figure-Q5 is adjacent to the cliff edge, with the Sarsa path being further back. Draw a plot of the expected average rewards for Q and Sarsa. Explain why Sarsa computes a “safer” path when compared to Q learning.

3 marks.

c) What is meant by the terms on-policy and off policy in the context of TD methods? The discussion should include the equations for an on policy and off policy update.

4 marks.