University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath

SPRING, 22/23 TRIMESTER EXAMINATIONS

COMP47590

Advanced Machine Learning

Module Coordinator: Assoc Professor Brian Mac Namee

Student Number

Seat Number

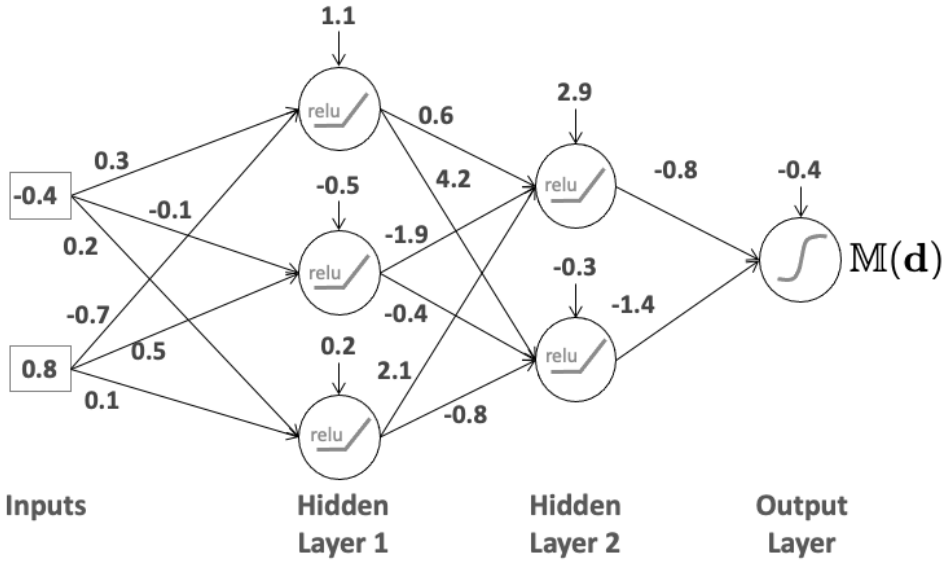
Time Allowed: 120 minutes

Materials Permitted in the Exam Venue:
Non-programmable or scientific calculator
Programmable calculator

Materials to be Supplied to Students:
8 Page Answer Booklets
New Cambridge Statistical Tables

Instructions to Students:
Answer any three out of four questions. All questions carry equal marks. Total marks available 90. The value of each part of each question is shown in brackets next to it.

1. **(a)** The image below shows a *feed forward artificial network*. The computational units in the two hidden layers use *rectified linear* (*relu*) activation functions and the output layer unit uses a *sigmoid* activation function. The *weights* and *biases* are shown along the links in the network.



(i) Perform a **forward propagation** through the network using an input feature vector of [-0.4, 0.8]. Show your workings.

[12 marks]

(ii) If the target feature value for the current input vector is 1.0, calculate the **loss** associated with this training instance using **cross entropy loss**.

[2 marks]

(b) **Gradient descent with momentum**, **RMSprop**, and **adam** are three common adaptations to the basic gradient descent algorithm used to train neural networks. Explain how these approaches improve upon basic gradient descent and how they differ from each other.

[8 marks]

(c) We can describe what artificial neural networks do as learning representations (or **embeddings**) of input data that make downstream tasks, for example classification, straight forward. Good embeddings are often said to be:

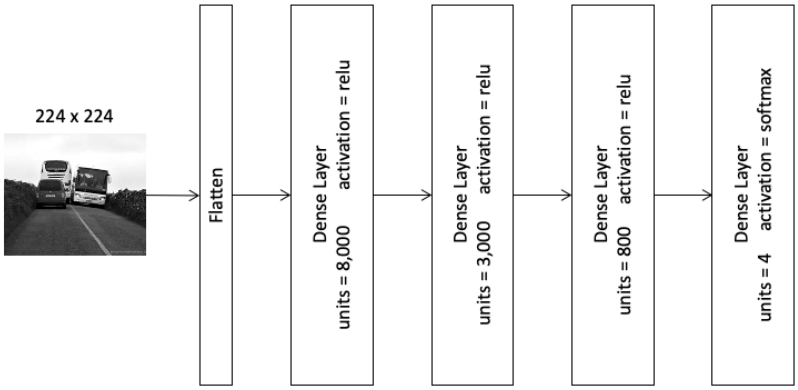
- **Distributed**
- **Abstract**
- **Invariant**
- **Disentangled**

Describe what each of these terms means in relation to embeddings generated by artificial neural networks.

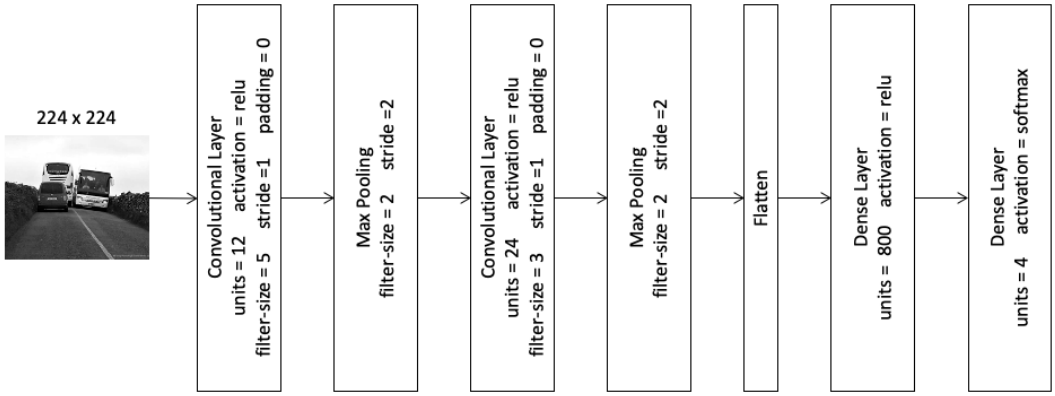
[8 marks]

2. **(a)** You have been tasked with training a neural network to control a self-driving car from image input. The model should output one of four control signals - *left, right, brake, or accelerate* - from each input image frame. The only input to the model is a 224 pixel by 224 pixel greyscale image from the front of the car.

Image (a) shows the architecture of a multi-layer perceptron neural network designed for this problem. Image (b) shows the architecture of a convolutional neural network designed for this problem. Both architectures are composed of four layers.



(a) Multi-layer perceptron network architecture

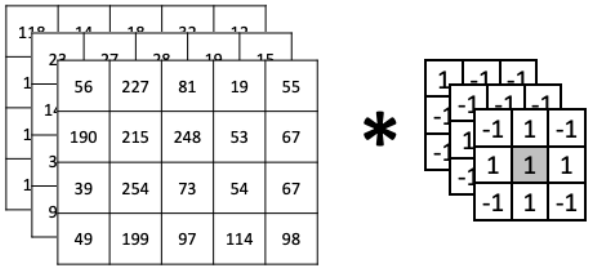


(b) Convolutional neural network architecture

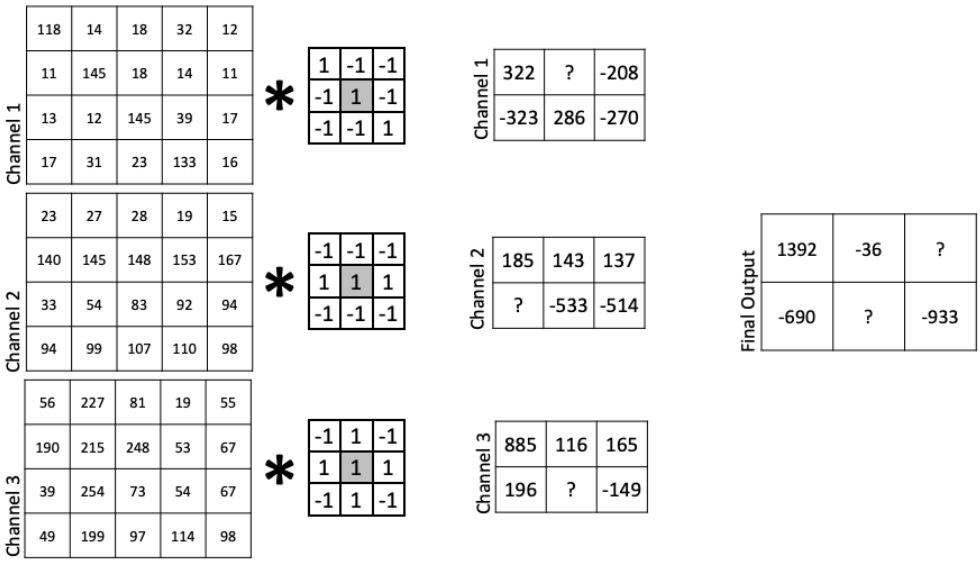
Calculate the number of parameters (weights and biases) that need to be learned for each network architecture.

[12 marks]

(b) The image below shows a 3-channel input that is being convolved (cross correlated) with a 3-channel 3 x 3 kernel.



The image below expands the three-channel input and three-channel kernel so that all values can be seen and shows the intermediate convolution result for each channel as well as the final output. Calculate the values marked with a **?** in the intermediate convolution results and the final output.



[8 marks]

(c) The 2017 paper "*Attention is all you need*" by Vaswani et al is now one of the most cited papers in machine learning research. Explain what **attention** is and how the **transformer** architecture utilises it.

[10 marks]

3. **(a)** Describe the concept of **discounted return** that is frequently used in reinforcement learning.

[5 marks]

(b) An intelligent agent trained to play a video game completes an episode and receives the following sequence of rewards over six timesteps:

$\{r_0 = 84, r_1 = 98, r_2 = -57, r_3 = -104, r_4 = -96\}$

Compare the **discounted returns** calculated at time $t = 0$ based on this reward sequence when discounting factors of 0.9 and 0.1 are used.

[5 marks]

(c) To try to better understand the slightly baffling behaviour of her new baby girl, Maria - a scientifically minded new mother - monitored her baby over a period of time, recording her activity at 20 minute intervals. The activity stream looked like this (with time flowing down through the columns):

0	SLEEPING	12	SLEEPING	24	SLEEPING	36	CRYING
1	CRYING	13	SLEEPING	25	HAPPY	37	HAPPY
2	SLEEPING	14	SLEEPING	26	CRYING	38	HAPPY
3	SLEEPING	15	CRYING	27	SLEEPING	39	HAPPY
4	SLEEPING	16	CRYING	28	SLEEPING	40	HAPPY
5	HAPPY	17	SLEEPING	29	HAPPY	41	HAPPY
6	HAPPY	18	SLEEPING	30	HAPPY	42	SLEEPING
7	HAPPY	19	HAPPY	31	HAPPY	43	SLEEPING
8	SLEEPING	20	SLEEPING	32	HAPPY	44	SLEEPING
9	SLEEPING	21	HAPPY	33	HAPPY	45	SLEEPING
10	SLEEPING	22	HAPPY	34	HAPPY	46	SLEEPING
11	SLEEPING	23	CRYING	35	CRYING	47	SLEEPING

Maria noticed that her baby could occupy one of three states - HAPPY, CRYING, or SLEEPING - and moved quite freely between them.

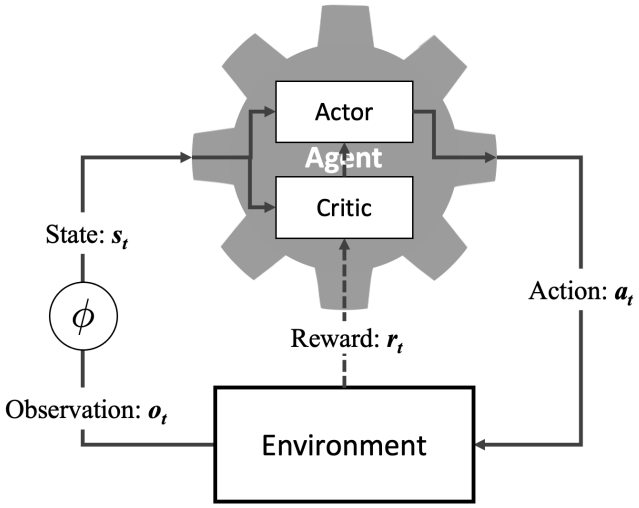
(i) Based on the sequence of states given above calculate a transition matrix that gives the probability of moving between each of the three states.

[10 marks]

(ii) Draw a Markov process diagram to capture the behaviour of a small baby as described above.

[5 marks]

(d) The image below shows an illustration of reinforcement learning using **actor-critic** method.



Describe the role of the actor and critic models in this approach.

[5 marks]

4. **(a)** Some benchmark experiments have found **ensemble models based on gradient boosting** can be prone to overfitting to incorrectly labelled instances in the training dataset (for example a positive instance mislabelled as a negative instance). Explain why this is the case and describe how a learning rate can be introduced to the gradient boosting algorithm to mitigate this.

[10 marks]

(b) When developing a machine learning model that will be deployed to perform a task for a user, we can describe three different goals of evaluation:

1. to determine which model is the most suitable for a task
2. to estimate how the model will perform after deployment
3. to convince users that the model will meet their needs

Describe the differences between these goals, and how the evaluation methods used to achieve each of them can be different.

[10 marks]

(c) The following is a definition of machine learning:

The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience.
- Tom Mitchell

Do you believe that this definition accurately defines machine learning? In your answer discuss the appropriateness of the definition, the scope of the definition, and any recommendations for improvements you would suggest.

[10 marks]