University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath

SPRING TRIMESTER EXAMINATIONS

ACADEMIC YEAR 2019/2020

COMP 47590

Advanced Machine Learning

Prof. Steve Counsell

Assoc. Prof. Chris Bleakley

Assoc. Prof. Brian Mac Namee*

Time Allowed: 2 Hours

Instructions for Candidates
Answer any **four** out of five questions. All questions carry equal marks. Total marks available **100**. The value of each part of each question is shown in brackets next to it.

Student Number

Seat Number

Instructions for Invigilators
This is a Closed Book/Notes exam.
Students are **not** permitted to bring materials to the Exam Hall.
Non-programmable calculators allowed.

1. **(a)** Thomas Deittrich describes three motivations for using ensemble models: **statistical**, **computational**, and **representational**. Describe each of these motivations and how they explain the performance of ensemble models.

[9 marks]

(b) Benchmark experiments have found repeatedly that ensemble models based on **bagging** are more robust to noise in the target features of a training dataset than ensemble models trained using **boosting**. Explain why this is the case. In your answer provide a short explanation of the bagging and boosting techniques.

[8 marks]

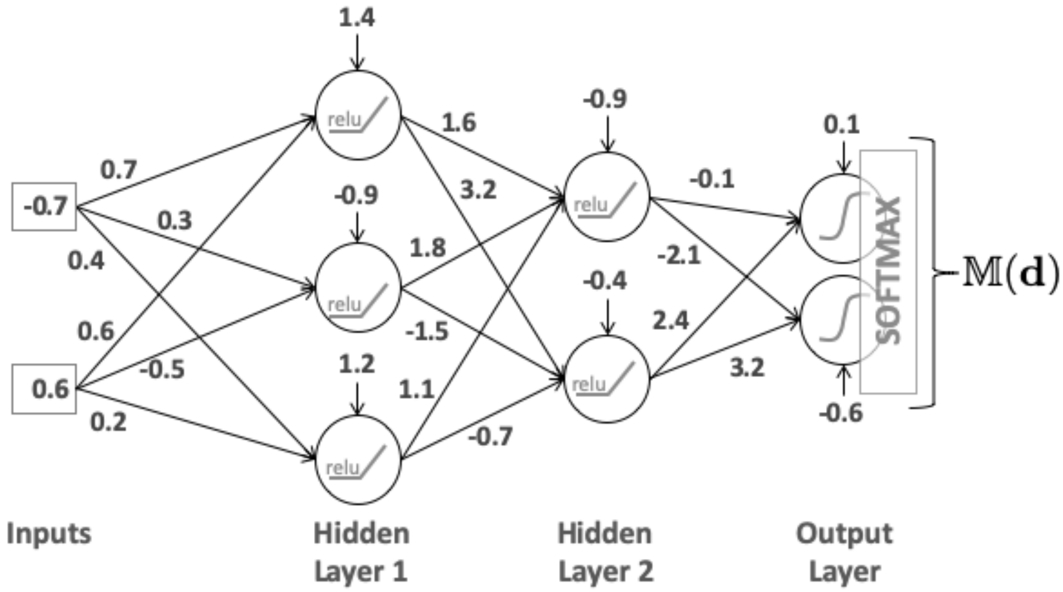
(c) The table below shows a simple dataset that describes the average temperature on a given day and the number of bicycles rented by a rental shop on that day. This dataset is being used to train a gradient boosting model to predict the number of rentals from the daily temperature. The column labelled $\mathbb{M}_0(\mathbf{d})$ shows the output of the first model added to the gradient boosting model for each instance in the training dataset.

ID	Temperature	Rentals	$\mathbb{M}_0(\mathbf{d})$
1	4	602	1,287.1
2	5	750	1,287.1
3	7	913	1,287.1
4	12	1229	1,287.1
5	18	1827	1,287.1
6	23	2246	1,287.1
7	27	2127	1,287.1
8	28	1714	1,287.1
9	32	838	1,287.1
10	35	625	1,287.1

Calculate the target values that the next base model added to the gradient boosting ensemble will be trained to predict.

[8 marks]

2. **(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 *softmax* 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 [-0.7, 0.6]. Show your workings.

[14 marks]

(ii) If the target feature vector for the current input vector is [1.0, 0.0], calculate the **loss** associated with this training instance using **cross entropy loss**.

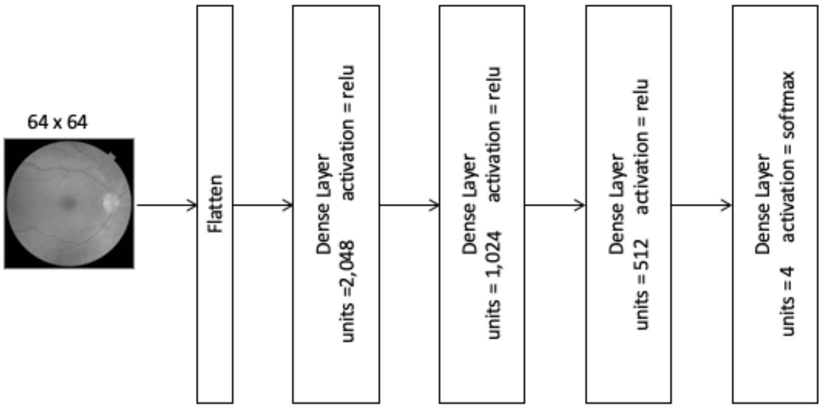
[3 marks]

(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 relate to each other.

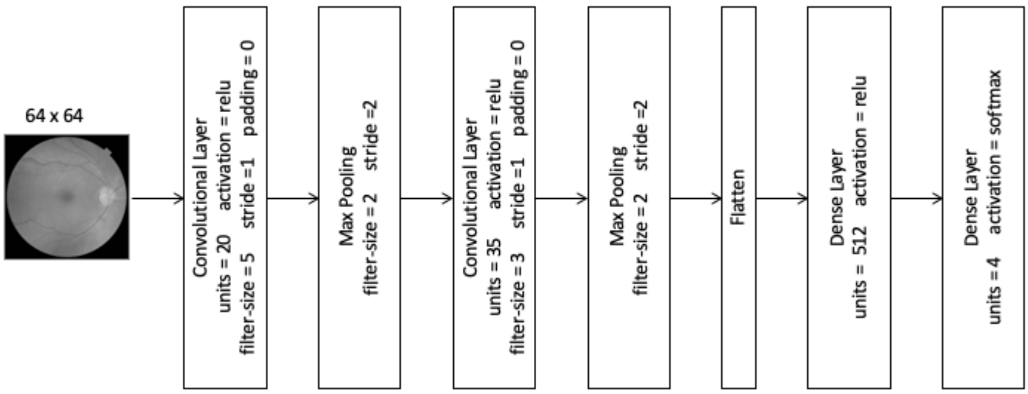
[8 marks]

3. (a) You have been tasked with training a neural network to diagnose diabetes from images of a person's retina. The model should produce diagnoses on a four-point scale: *grade-0*= no diabetes, *grade-1*= early stage diabetes, *grade-2*= diabetes, *grade-3*= advanced diabetes. The only input to the model is a 64 pixel by 64 pixel greyscale image of the retina of a person's right eye.

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) A multi-layer perceptron network for diabetes diagnosis



(b) A convolutional neural network for diabetes diagnosis

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

[12 marks]

(b) The ability of convolutional networks to learn accurate models with many fewer weights than multi-layer perceptrons of similar depth is often attributed to **shared weights** and **sparse connections**. Explain the meaning of these two terms.

[5 marks]

(c) R**ecurrent neural networks** can be built to model *many-to-one*, *one-to-many,* or *many-to-many* relationships. For each of these three types of relationships briefly describe the network architecture that could be used to model it, and give an example application for which it would be appropriate to model that type of relationship.

[8 marks]

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

[3 marks]

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

$\{r_0 = -33, r_1 = -11, r_2 = -12, r_3 = 27, r_4 = 87, r_5 = 156\}$

Compare the discounted returns calculated at time $t = 0$ based on this reward sequence when discounting factors of 0.72 and 0.22 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 the course of a day recording her activity at 20 minute intervals. The activity stream looked like this (with time flowing down through the columns):

SLEEPING	SLEEPING	SLEEPING	CRYING	SLEEPING	SLEEPING
CRYING	SLEEPING	HAPPY	HAPPY	CRYING	HAPPY
SLEEPING	SLEEPING	CRYING	HAPPY	SLEEPING	HAPPY
SLEEPING	CRYING	SLEEPING	HAPPY	SLEEPING	HAPPY
SLEEPING	CRYING	SLEEPING	HAPPY	SLEEPING	HAPPY
HAPPY	SLEEPING	HAPPY	HAPPY	SLEEPING	HAPPY
HAPPY	SLEEPING	HAPPY	SLEEPING	HAPPY	HAPPY
HAPPY	HAPPY	HAPPY	SLEEPING	HAPPY	SLEEPING
SLEEPING	SLEEPING	HAPPY	SLEEPING	HAPPY	SLEEPING
SLEEPING	HAPPY	HAPPY	SLEEPING	SLEEPING	SLEEPING
SLEEPING	HAPPY	HAPPY	SLEEPING	HAPPY	SLEEPING
SLEEPING	CRYING	CRYING	SLEEPING	SLEEPING	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.

[6 marks]

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

[3 marks]

(c) TempleRunLite is a simplified version of Temple Run, the popular endless runner game. In TempleRunLite:

- the player is constantly moving forward at a fixed speed
- the player is aware only of obstacles (e.g. fallen columns) within a certain distance of its position
- possible actions are moving left, moving right, or going straight
- crashing into an obstacle (e.g. a fallen column) leads to the end of the game
- the player can collect coins as they run
- the goal of the game is to maximise the number of coins collected by the player in a single run

Describe how you could use reinforcement learning to build an automated TempleRunLite player. In your answer describe how you would model **states**, **actions**, **rewards**, and any other important elements of the solution**.**

[8 marks]

5. (a) Machine learning algorithms face a constant struggle between **over-fitting** and **under-fitting**. Explain what this means.

[8 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.

[8 marks]

(c) The EU General Data Protection Regulation (GDPR) came into effect on May 25th, 2018. What effect does this regulation have on the use of machine learning?

[8 marks]