University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath

SEMESTER II EXAMINATIONS

ACADEMIC YEAR 2018/2019

COMP 47590

Advanced Machine Learning

Assoc. Prof. V. Dimitrova

Prof. P. Cunningham

Dr. B. 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.

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)** Describe *three* different motivations for using **ensemble methods** in machine learning.

[9]

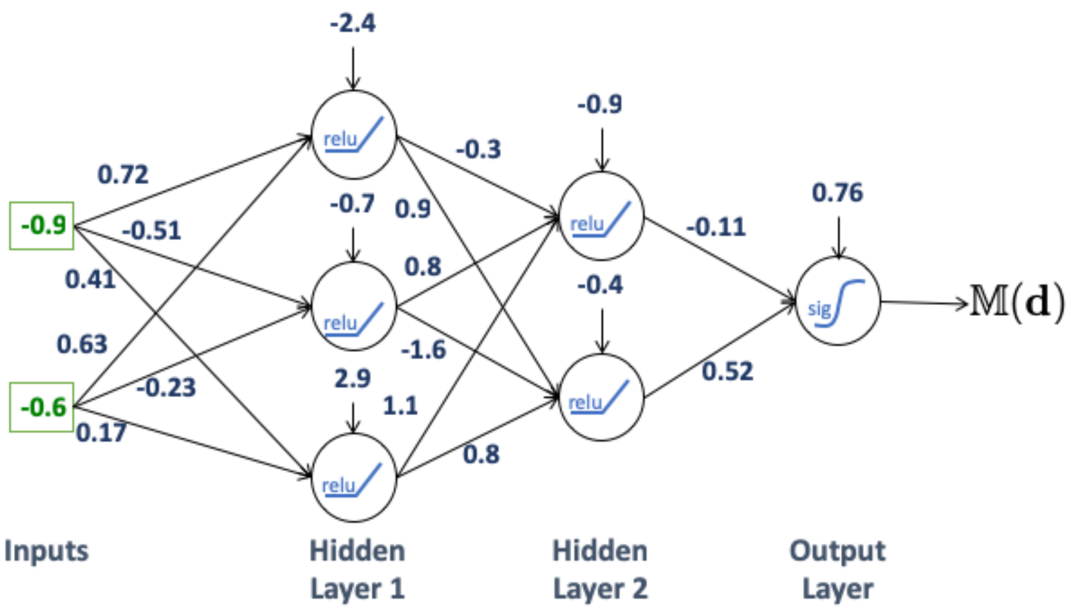
(b) Benchmark experiments have found repeatedly that ensemble models based on **bagging** are more robust to noise in the target features of a training dataset than ensemble models trained using **boosting**. Explain why this is the case. In your answer provide a short explanation of the bagging and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

[8]

(c) **Gradient boosting** has recently been shown to offer significant performance improvements over other boosted ensemble approaches. Explain what the gradient boosting algorithm trains its base models to predict.

[8]

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 *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 (-0.9, -0.6). Show your workings.

[12]

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

[3]

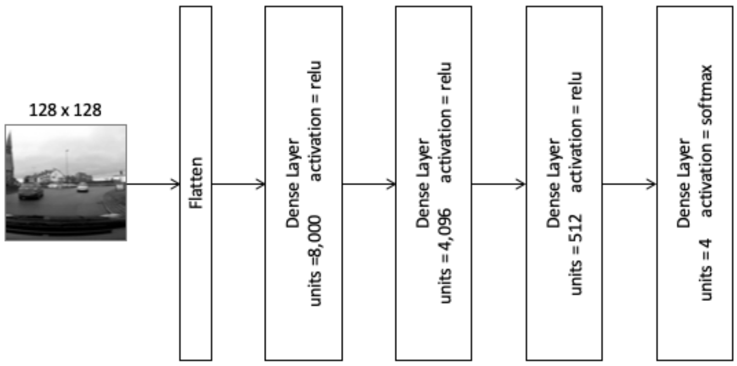
(b) Some practitioners favour replacing units using a **rectified linear** (**relu**) activation function with units using a **leaky relu** activation function. Describe each of these activation functions (include appropriate diagrams), and explain the potential advantages of using l**eaky relu**.

[4]

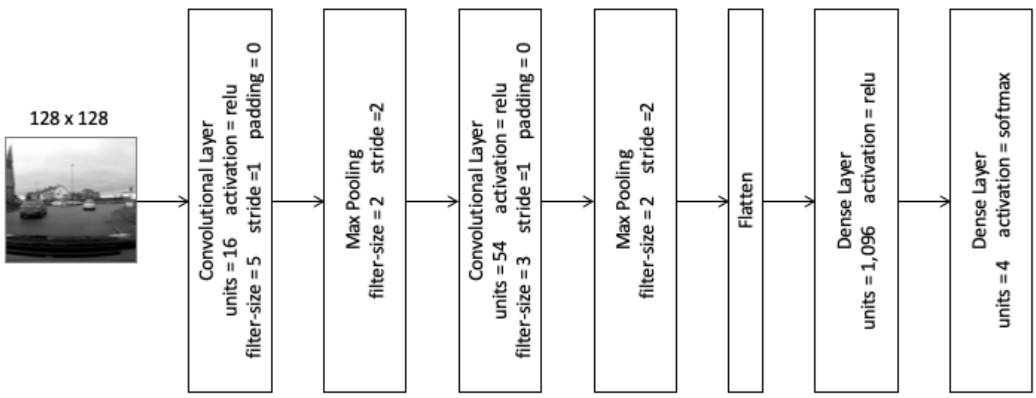
(c) There are three common variants of the gradient descent algorithm when training deep neural networks: **batch gradient descent**, **mini-batch gradient descent**, and **stochastic gradient descent**. Describe each of these approaches and discuss the advantages and disadvantages of each.

[6]

3. (a) You have been tasked with building a neural network system for controlling a self-driving car. The car has just four controls: accelerate, brake, turn left, and turn right. The only input is a 128 pixel by 128 pixel greyscale image from a dashboard camera within the car. Image (a) shows a multi-layer perceptron neural network architecture designed for this problem. Image (b) shows a convolutional neural network architecture designed for this problem. Both architectures are composed of four layers.



(a) A multi-layer perceptron network for self-driving car control



(b) A convolutional neural network for self-driving car control

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

[12]

(b) The convolutional neural network in part (a) has much fewer weights than the multi-layer perceptron network. Explain how the convolutional neural network is able to learn accurate models with so many fewer learned parameters.

[5]

(c) It is often said that a **recurrent neural network** (**RNN**) can be "*unrolled through time*". Explain what this means.

[8]

4. (a) Describe the *four* key components of a **reinforcement learning** system.

[8]

(b) The differences between the **SARSA** (**State-Action-Reward-State-Action**) and **Q-learning** algorithms for reinforcement learning are often summarised by saying that SARSA is *on-policy* while Q-learning is *off-policy*. Explain what this means.

[7]

(c) OutRunLite is a simplified version of OutRun, the popular 1980s arcade game. In OutRunLite:

- the car is constantly moving forward at a fixed speed
- the player is aware only of obstacles (e.g. other cars) within a certain distance
- possible actions are moving left, moving right, or going straight
- crashing into an obstacle (e.g. another car) leads to the end of the game
- the goal of the game is to maximise the distance covered by the car in a single run

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

[10]

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

[8]

(b) The table below shows the results of a benchmark experiment to compare the performance of a number of variants of a new learning algorithm, *YALA*, against each other and two baseline methods (random forests and multi-layer perceptrons). The performance of these algorithms has been measured across five different classification datasets using *10-fold cross validation*. Performance is measured in all cases using *micro-average accuracy*.

	YALA-1	YALA-2	YALA-3	Random Forest	MLP
MNIST	0.492	0.291	0.431	0.595	0.279
CiFAR-10	0.826	0.809	0.557	0.851	0.765
Flowers	0.728	0.456	0.635	0.467	0.733
Animals	0.296	0.641	0.522	0.495	0.362
Cars	0.941	0.701	1.000	0.976	0.993

(i) Describe concerns that you would have with the use of micro-average accuracy in this experiment.

[4]

(ii) Based on these performance scores determine which algorithm is performing *best* in this experiment.

[5]

(c) The new EU General Data Protection Regulation (GDPR) came into effect May 25^{th} 2018. In the run up to this Prof. Pedro Domingez, author of The Master Algorithm, stated that:

> "*Starting May 25, the European Union will require algorithms to explain their output,* ***making deep learning illegal****.*"

Discuss this claim.

[8]