IMPERIAL COLLEGE LONDON

MSc EXAMINATION 2019

For internal students of Imperial College London

Taken by students of the Masters of Applied Computational Science and Engineering

ACSE-8 Machine Learning

Friday 31 May 2019, 10:00-12:00

Total number of marks = 100.

Each question must be answered.

DO NOT OPEN THIS EXAMINATION PAPER UNTIL INSTRUCTED TO DO SO BY THE INVIGILATOR

Module 8 Exam

Question 1 (13 points)

Suppose we perform logistic regression to predict, from a number of measurements on a device, whether this device is in good or poor working condition. We code "good working condition" as 1 and "poor working condition" as 0. We assume that we have a training set of m devices $\left(\left(x^{(i)}\right),\left(y^{(i)}\right)\right)$ where, for each device i the vector $x^{(i)}$ contains the measurements and $y^{(i)}$ is the known working condition (equal to 0 or 1). We call $h_\theta\left(x^{(i)}\right)$ the hypothesis produced by logistic regression for device i. A sigmoid activation function is used.

The expression of the cross-entropy loss function over the m devices of the training set is:

$$-\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(h_\theta\left(x^{(i)}\right)\right)+\left(1-y^{(i)}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)\right]$$

Questions :

- Explain how $h_\theta\left(x^{(i)}\right)$ is calculated and what is its range of variation. How can $h_\theta\left(x^{(i)}\right)$ be interpreted? (5 points)
- Is the cross-entropy negative or positive? (2 points)
- Explain how the cross-entropy treats high and low values of $h_\theta\left(x^{(i)}\right)$ in the case of a device i in good working condition. (3 points)
- Explain how the cross-entropy treats high and low values of $h_\theta\left(x^{(i)}\right)$ in the case of a device i in poor working condition. (3 points)

Question 2 (12 points)

Suppose that the last layer of a neural network includes 10 output neurons, and that the values of the 10 neurons are:

(-3, 2, 4, 1, 7, -5, -1, -3, 3, 5).

For each coordinate a_i of a vector $(a_1, a_2, .., a_9, a_{10})$, the Softmax function is defined as:

$$\text{Softmax}(a_i) = \frac{e^{a_i}}{\sum_{j=1}^{j=10} e^{a_j}}$$

Questions:

- What are the values obtained by Softmax for the 10 output neurons (calculations up to two decimal places)? (4 points)
- Interpret the results of the Softmax function and explain how it is used in practice. (8 points)

Question 3 (18 points)

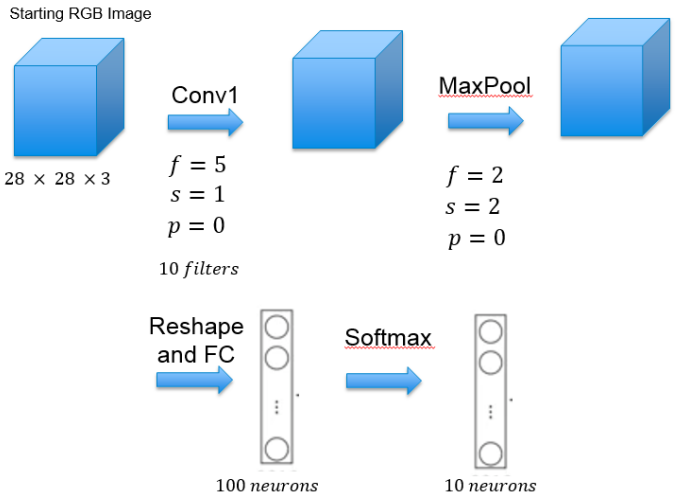
You have a training set of 1000 labelled images that you want to use to build a binary classification neural network. Each image is a color image of 128x128 pixels. The first class is that of images containing a car and the second class is that of images containing a plane. 500 images contain a car and 500 contain a plane.

In addition, you have a test set of 200 labelled images which will be used to evaluate the generalization performance of your trained model. 100 images of the test set contain cars and 100 contain planes.
Questions:

- What is the benefit of using transfer learning to train your neural network? (5 points)
- In this example, would it be more appropriate to start from a network trained on MNIST or trained on Imagenet? Explain your answer. (5 points)
- Give an example of three hyperparameters to optimize. (3 points)
- Which approach do you recommend to choose the optimal values of your hyperparameters? (5 points)

Question 4 (30 points)
The figure below describes a simple convolutional network. We assume that each convolution operation is performed with a bias term.

- Calculate the number of parameters associated with each of the three layers (Conv1, Reshape and FC, and Softmax). (20 points)
- Calculate the number of output neurons for the Conv1 and MaxPool layers. (10 points)



	Size of input image n	Number of input channels	convolution filter	padding p	stride s	Size of output image (n+2p-f)/s + 1	Number of output channels or filters	Number of output neurons	Size of Filter + 1	Number of Parameters
Conv1	28	3	5	0	1		10			
MaxPool			2	0	2					
	Size of input							Number of output neurons		
Reshape and FC								100		
Softmax	100							10		

Question 5 (12 points)

- Suppose the random variable X follows a uniform distribution between 0 and 1. λ is a positive parameter. What is the cumulative density function (CDF) of the random variable Y defined as $Y = -\frac{log(1-X)}{\lambda}$? (7 points)
- What is the probability density function (PDF) of Y? (5 points)

Question 6 (15 points)

This question is about the differences and similarities between Variational AutoEncoders (VAEs) and Generative Adversarial Networks (GANs).

- Explain how each technique treats the latent vector to obtain a sample in the model space. (9 points)
- Define what is an encoder and a decoder, and explain how VAEs and GANs address encoding and decoding. (6 points)