Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

FACULTY OF ENGINEERING, MATHEMATICS & SCIENCE

SCHOOL OF ENGINEERING

Electronic & Electrical Engineering

Sample Exam

Machine Learning with Applications in Media Engineering (EE4C16/EE5M16)

Dr. F. Pitié

Instructions to candidates:

- Answer FOUR (4) questions.
- Please answer questions from each section in separate answer books.

Materials Permitted for this Examination:

- New Formulae & Statistic Tables
- Graph Paper
- Non-programmable calculators

This is sample exam. Each Question is worth 25 marks. You need to answer 4 of the 5 Questions at your choice.
In this sample not all sub-questions have been fleshed out or answered.

Question 1

1. Explain what an epoch is in DNN training.

   [3 marks]

2. Explain what over fitting and under fitting are in the context of Machine Learning. Name a few techniques used in DNNs to mitigate overfitting.

   [6 marks]

3. Explain the steps involved in Gradient Descent. How it is used to train DNNs?

   [6 marks]

4. What is the historical importance of AlexNet?

   [5 marks]

   AlexNet is the name of a convolutional neural network which won the ILSVR challenge in 2012. It is THE paper that sparked the Deep Learning revolution. It outperformed all other non-DNN techniques by a 10% margin. All subsequent winning entries used DNN. It was the first time that a DNN won a major challenge and this caught the attention of most researchers in the field and beyond.

5. What is the problem of vanishing gradients? Why is it a problem for DNN training?

   [5 marks]

Question 2

1. A big retail company contacts your team to design a system that can recognise the make and type of each car on their supermarket car parks.

   You are the Tech Lead on this project. Make a project plan discussing the technical challenges, your proposed solutions. Also discuss any non-technical issue that mi

   [12 marks]

   The answer should mention how your company can get hold of training data/validation and test data.

   One technical issue when sourcing training data is to get data that is representative of the application use cases. For instance in this application, you will need to source existing images from existing CCTV systems, taken at different locations, times of the day, times of the year and weather.

   Other difficulties are around the legality of getting the data, possible ethical considerations and finally scale of the labelling task. You might need to hand label a million pictures!

   Then you need to suggest some possible DNN architecture. How deep do you expect the network to be? Are you going to use CNN, RNN, Dense layers? Can you avail of Transfer Learning and re-use existing off-the-shelf networks such as VGG or ResNet?

   Depending on your architecture and the expected recognition rate, very roughly indicate how much data may be required.

   Say in this case, you probably require tens of thousands of images at least to cover most use cases.

2. another question

   [4 marks]

3. another question

   [4 marks]

4. another question

   [5 marks]

Question 3

1. Consider a binary classifier with the following confusion matrix:

	actual: 0	actual: 1
predicted: 0	TN=16	FN=4
predicted: 1	FP=10	TP=70

Comment on the performance of the classifier.

For this kind of questions, you are first expected to compute a few metrics (such as accuracy, recall, false positive rate, etc.).

Accuracy: 86%. However the number of actual positives is 26 and the number of actual negatives is 74. There is an imbalance in the dataset, hence it is no surprise that the accuracy is high.

True Negative Rate: 10/(10+16)=38.5% is pretty high, showing a potential problem.

Remember it all depends on your particular application. You probably won't have enough information to make any definitive statement about the performance, you just need to highlight the potential issues.

Question 4

Note: all "mathematical" sub-questions will be contained in a single **Question**.

1. An autoregressive model is when a value from a time series is regressed on previous values from that same time series.

$$x_t = w_0 + \sum_{i=1}^{p} w_i x_{t-i} + \varepsilon_t$$

write the design matrix for this problem.

[13 marks]

Say the data x is available from time $t = 0$, ie that we have access to $x_0, x_1, \cdots, x_n$. Then we can only start using the prediction model from $t = p$ onward.

$$X = \begin{pmatrix} 1 & x_{p-1} & x_{p-2} & \cdots & x_0 \\ 1 & x_p & x_{p-1} & \cdots & x_1 \\ 1 & x_{p+1} & x_p & \cdots & x_2 \\ 1 & x_{p+2} & x_{p+1} & \cdots & x_3 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n-1} & x_{n-2} & \cdots & x_{n-p} \end{pmatrix}$$

The size of the matrix is $(n - p + 1) \times (p + 1)$

For reference, the expected output **y** is

$$\mathbf{y} = \begin{pmatrix} x_p \\ \vdots \\ x_n \end{pmatrix}$$

2. Consider the linear model $y = w_0 + w_1 x$. We want to bias w_1 towards the value $\hat{w}_1$. Write a loss function that achieves this.

[7 marks]

Assume that we have n samples $(x_i, y_i)_{1 \leq i \leq n}$. We can start with a L2 loss function on the model as follows:

$$E(w_0, w_1) = \sum_{i=1}^{n} (w_0 + w_1 x_i - y_i)^2$$

We can add a L2 regularisation term to achieve the desired bias as follows:

$$E'(w_0, w_1) = \sum_{i=1}^{n} (w_0 + w_1 x_i - y_i)^2 + \lambda (w_1 - \hat{w}_1)^2$$

Where $\lambda > 0$ is used to tune the bias.

3. some question on gradient derivation

[5 marks]

Question 5

1. Remember that one question (25 marks) will be a short essay on the keynotes from Xilinx and Intel. Write around 1-1.5 pages outlining the importance/problematic/optimisations, etc.

[25 marks]

Supporting material

You will be provided this supporting material at the exam.

Assuming $\mathbf{a}, \mathbf{b}, \mathbf{A}$ are independent of $\mathbf{w}$, below is a list of useful gradient computations:

$$
\begin{aligned}
\frac{\partial \mathbf{a}^\top \mathbf{w}}{\partial \mathbf{w}} &= \mathbf{a} \\
\frac{\partial \mathbf{b}^\top \mathbf{A}\mathbf{w}}{\partial \mathbf{w}} &= \mathbf{A}^\top \mathbf{b} \\
\frac{\partial \mathbf{w}^\top \mathbf{A}\mathbf{w}}{\partial \mathbf{w}} &= (\mathbf{A} + \mathbf{A}^\top)\mathbf{w} \qquad \text{(or } 2\mathbf{A}\mathbf{w} \text{ if } A \text{ symmetric)} \\
\frac{\partial \mathbf{w}^\top \mathbf{w}}{\partial \mathbf{w}} &= 2\mathbf{w} \\
\frac{\partial \ \mathbf{a}^\top \mathbf{w}\mathbf{w}^\top \mathbf{b}}{\partial \ \mathbf{w}} &= (\mathbf{a}\mathbf{b}^\top + \mathbf{b}\mathbf{a}^\top)\mathbf{w}
\end{aligned}
$$