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

Engineering — Hillary Term, 2017

Senior Sophister

Annual Examinations

Machine Learning with Applications in Media Engineering (EE4C16)

Date: Venue: Time:

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

Question 1

1. Explain why we need a validation set when developing a Machine Learning system.

   [5 marks]

2. What does the backpropagation algorithm compute?

   [3 marks]

3. Explain why backpropagation makes DNN practical.

   [5 marks]

4. Suppose that a credit card company decides to deploy a new AI system for assessing the risk of default of its customers. The new system is using a feed-forward neural network. Suggest, in a form of essay, what should the bank have before the system can be used? Discuss problems associated with this requirement.

   [12 marks]

Question 2

1. The estimation of the weights in a DNN is usually based on the gradient descent optimisation technique. Is the gradient descent technique guaranteed to converge to the global minimum? If not, comment on the conditions and the type of convergence.

   [5 marks]

2. "A classifier trained on less data is less likely to overfit". comment.

   [4 marks]

3. What is a "Batch gradient" descent? What is a "stochastic gradient" descent?

   [4 marks]

4. Why do RNNs have a tendency to suffer from exploding/vanishing gradient?

   [6 marks]

5. There are a few ways of preventing the problem of vanishing gradients. Name one activation function that can be used to reduce the issue. Name one example of CNN architecture designed to address the issue. Name one example of RNN architecture designed to address the issue.

   [6 marks]

Question 3

1. Assuming that the matrix A is symmetric, and denoting I as the identity matrix, and λ is real number. Compute the gradient $\frac{\partial E(\mathsf{w})}{\partial \mathsf{w}}$ of $E(\mathsf{w}) = \|(\mathsf{A} + \lambda \mathsf{I})\mathsf{w}\|^2$.

   [5 marks]

2. We also know that $\mathsf{Bw} \approx \mathsf{b}$. Design a new loss function $E'(\mathsf{w})$ that adapts the loss function $E(\mathsf{w})$ to add a L2 regularisation term to loosely enforce this constraint.

   [5 marks]

3. Derive the gradient $\frac{\partial E'(\mathsf{w})}{\partial \mathsf{w}}$

   [5 marks]

4. Bogus Forensics Ltd. has developed a new test for detecting potential criminals using a new patented hand writing analysis technology. The company claims an 89% accuracy and impresses the investors. Comment.

   [5 marks]

5. After investigation, you obtain the following confusion matrix for their detector:

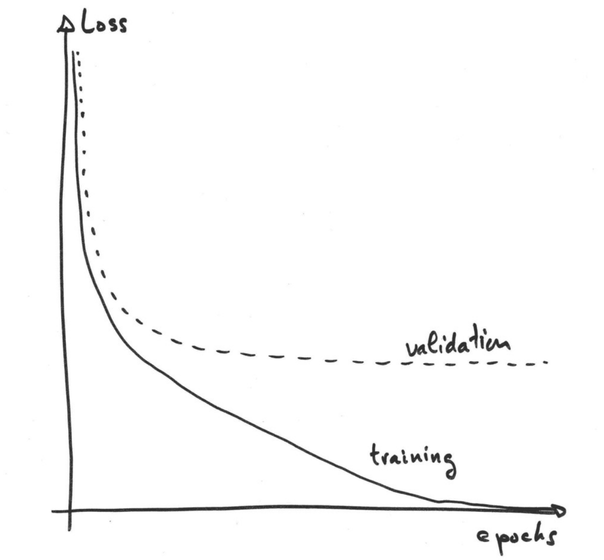
	actual: 0	actual: 1
predicted: 0	TN=80	FN=8
predicted: 1	FP=3	TP=9

   A positive prediction means that we predict that the person will commit a crime. Make your own analysis on the value of the product.

   [5 marks]

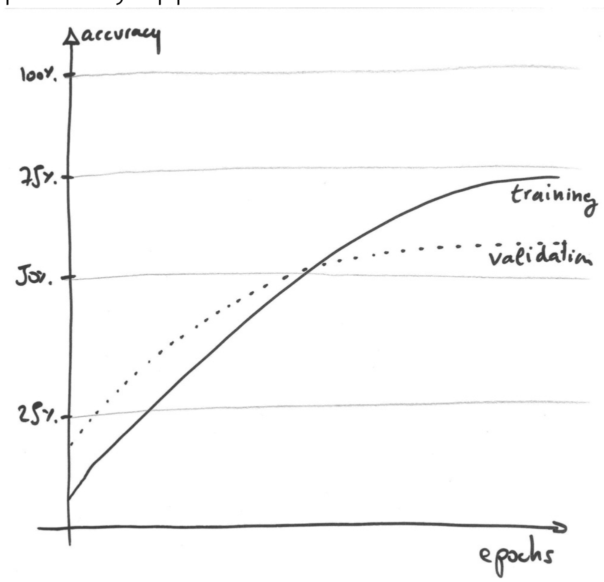
Question 4

1. While training a convolutional neural net, we obtain a loss function graph similar to the one below. Comment what is happening and explain what your next steps should be.



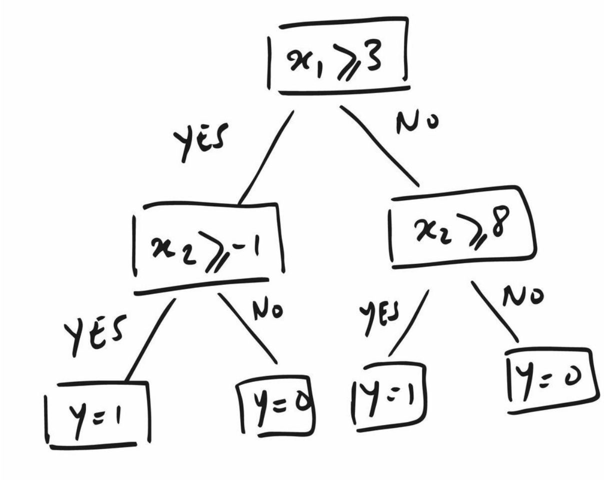
[5 marks]

2. While training a convolutional neural net, we obtain an accuracy graph similar to the one below. Comment on the graph and indicate what kind of training technique was probably applied.



[5 marks]

3. Recall what a Decision Tree is and briefly outline why any Decision Tree can be modelled as a neural network. Illustrate this on the simple example below. You can use the binary step activation $f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$.



[15 marks]

Question 5

1. Discuss, in the form of an essay, the challenges of deploying DNN applications on embedded devices. The answer should include discussions about the importance of embedded systems in DNN applications, the specific requirements of embedded systems and the different strategies available to optimise the networks to these specific constraints.

[25 marks]

Supporting material

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

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