University of Glasgow

Thursday 4 May 2023
14.00 – 15.30 BST
Duration: 1 hour 30 minutes
Additional time: 30 minutes
Timed exam – fixed start time

DEGREE OF MSc

Deep Learning for MSc (M)

COMPSCI5103

Answer all 3 questions

This examination paper is an open book, online assessment and is worth a total of 60 marks.

1. You should answer the following questions by manually examining the following models with consecutive PyTorch layers as given (there will not be time to create these models in code during the exam and this would not answer the questions anyway):

```
model1 consisting of PyTorch layers:
  ('flatten', nn.Flatten()),
  ('fc1', nn.Linear(224*224,28*28)),
  ('sig', nn.Sigmoid())

model2 consisting of PyTorch layers:
  ('conv1', nn.Conv2d(1,2,3,stride=2,padding=1)),
  ('sig1', nn.Sigmoid()),
  ('conv2', nn.Conv2d(2,4,3,stride=2,padding=1)),
  ('sig2', nn.Sigmoid()),
  ('conv3', nn.Conv2d(4,1,3,stride=2,padding=1)),
  ('sig3', nn.Sigmoid()),
  ('flatten', nn.Flatten())
```

(a) The input tensor to these models is of shape: `torch.size([3,1,224,224])` What type of real-world data does this represent and explain what each dimension of the tensor is used for. [2]

(b) Determine the total number of parameters in `model1`. Show and briefly describe your working for full credit. [4]

(c) Determine the total number of parameters in `model2`. Show and briefly describe your working for full credit. [4]

(d) If the tensor given in part (a) is put through these networks, calculate the shapes of tensor outputs from successive PyTorch layers for each model. Write down the name of each layer and put the size of the tensor output next to it. [4]

(e) Comment on advantages and disadvantages of both these models in terms of code and also processing the type of data highlighted in part (a). [3]

(f) Describe the size of the output you would get if you put in a tensor of shape `torch.size([3,1,28,28])` into `model1` and `model2`. [3]

2. **(a)** Calculate the result of applying the following kernel to the given example image. Assume no padding and a stride of 1. [3]

(2x3 kernel)

1	0	1
0	1	0

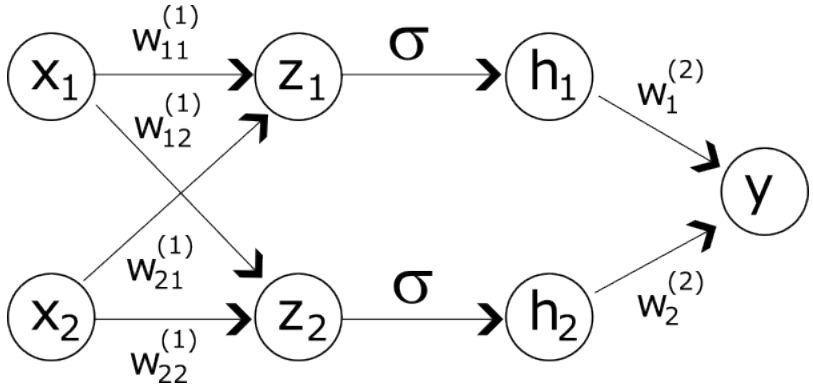
(Image values for 3x5 pixel image)

0	12	15	12	0
12	15	15	15	12
0	12	20	25	30

(b) If a 2x2 max pooling layer with a stride of 1 was then applied to the result from part (a), determine the output image from that layer. [1]

(c) The following shows a neural network, using a sigmoid activation function, with a structure and notation that should be familiar to you. Let $\boldsymbol{x} \in \mathbb{R}^2$, $\boldsymbol{z} \in \mathbb{R}^2$, $\boldsymbol{h} \in \mathbb{R}^2$ denote the relevant vectors. Let $\boldsymbol{M} \in \mathbb{R}^{2\times 2}$ and $\boldsymbol{N} \in \mathbb{R}^{1\times 2}$ denote relevant matrices.

In vector/matrix notation, write out the equations for y as a function of $\boldsymbol{h}$, $\boldsymbol{h}$ as a function of $\boldsymbol{z}$, and $\boldsymbol{z}$ as a function of $\boldsymbol{x}$. [3]



(d) Given the equations you wrote in part (c), write out the matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ in terms of their individual matrix elements given in the figure. [4]

(e) Calculate in terms of symbols: $\nabla_{\boldsymbol{h}} y$ [2]

(f) By using back-propagation, determine: $\nabla_{\boldsymbol{z}} y$ [3]

(g) Again using back-propagation, determine: $\frac{\partial y}{\partial x_1}$ [4]

3. A large city is considering using security cameras to detect illegal and antisocial behaviour by people in the city centre. You have been asked to design a system which could segment people from a series of image frames and classify their behaviour as 'normal', 'antisocial' or 'criminal'. In cases of antisocial or criminal activity, the police will be notified, and potentially deployed, and the recordings will be stored for later analysis in criminal proceedings. The recordings will also be added into future training data as automatically annotated data. If the classifier is uncertain, it is to send the video sequence to a group of human annotators who will then classify the scene and add it to the training data for future optimisation.

(a) What are the characteristics of this sequence learning problem? [3]

(b) How could *neural attention mechanisms* improve the performance and interpretability of the surveillance system? [4]

(c) Suggest one alternative visualisation technique which could be used, and discuss its advantages and disadvantages. [3]

(d) What are the risks associated with the proposed approach for augmenting the training data? [3]

(e) What are the other ethical and practical risks associated with the use of machine learning in such a system? [7]

END OF QUESTION PAPER