NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER 1 EXAMINATION 2021-2022

CE4042/CZ4042 – NEURAL NETWORKS AND DEEP LEARNING

Nov/Dec 2021 Time Allowed: 2 hours

INSTRUCTIONS

1. This paper contains 4 questions and comprises 6 pages.
2. Answer **ALL** questions.
3. This is an open-book examination.
4. All questions carry equal marks.

---

1. (a) Give brief answers to the following. Each part carries 2 marks.

   (i) State the shape of the following tensor [[[[2, 0, 1]], [[2, 5, 3]]]] written in Python.

   (ii) Consider a trained logistic neuron. Would multiplying all its weights and the bias by a constant change its accuracy? State why?

   (iii) State how you could use a linear neuron to learn a given nonlinear function.

   (iv) State one advantage each for selecting mini-batch gradient descent method over batch gradient descent, and selecting mini-batch gradient descent method over stochastic gradient descent.

   (v) To train a neuron layer, all the weights were initialized to 0.6. Is this good idea? Justify your answer.

Note: Question No. 1 continues on Page 2

(vi) A feedforward neural network receives two-dimensional inputs, has a hidden layer comprising of 5 neurons, and has an output layer comprising of 3 neurons. What is the total number of trainable parameters (i.e., weights and biases) in the network?

(vii) State why you would choose the three-way data split method over the cross-validation method and why you would sometimes have to use the cross-validation over the three-way data split method.

(14 marks)

(b) Figure Q1(b) shows a dataset with each example belonging to one of the two classes, displayed in the space of its two features x_1 and x_2. You are to design a discrete perceptron network with a single hidden layer to classify the examples.

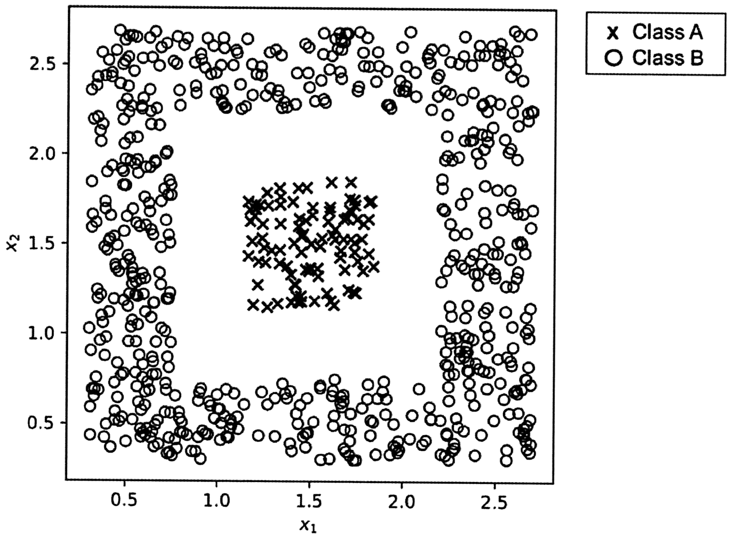

Figure Q1b

(i) Draw a decision boundary separating the two classes to design the network.

(2 marks)

(ii) State the number of neurons in the hidden layer.

(1 marks)

(iii) Draw the network indicating the values of all weights and biases.

(8 marks)

2. A two-layer feedforward neural network shown in Figure Q2 receives two-dimensional inputs $(x_1, x_2) \in \boldsymbol{R}^2$ and produces an output class label $y \in \{1, 2\}$. The hidden layer consists of three perceptrons and the output layer is a *softmax* layer of two neurons. The weights of the network are initialized as indicated in the figure and all the biases are initialized to 0.2.

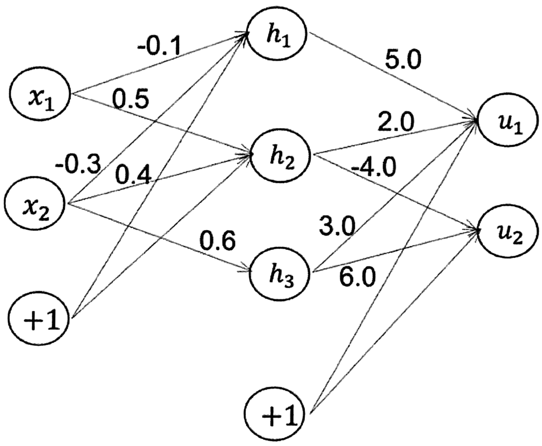

Figure Q2

The network is trained to produce a desired output $d = 2$ for an input $\boldsymbol{x} = \begin{pmatrix} 0.5 \\ 2.0 \end{pmatrix}$. You are to perform one iteration of gradient descent learning with the example $(\boldsymbol{x}, d)$. The learning factor $\alpha = 0.4$. Give your answers up to two decimal places.

(a) Write initial weight matrix $\boldsymbol{W}$ and bias vector $\boldsymbol{b}$ of the hidden layer, and initial weight matrix $\boldsymbol{V}$ and bias vector $\boldsymbol{c}$ of the output layer.

(2 marks)

(b) Find synaptic input $\boldsymbol{z}$ and output $\boldsymbol{h}$ of the hidden layer, and synaptic input $\boldsymbol{u}$ and output y of the output layer.

(5 marks)

(c) State the probabilities that the input $\boldsymbol{x}$ belongs to each class.

(2 marks)

(d) Find the cross-entropy cost J and classification error.

(3 marks)

Note: Question No. 2 continues on Page 4

(e) Find gradients $\nabla_{\boldsymbol{u}} J$ and $\nabla_{\boldsymbol{z}} J$ of the cost J with respect to $\boldsymbol{u}$ and $\boldsymbol{z}$, respectively.

(7 marks)

(f) Find gradients $\nabla_{\boldsymbol{V}} J$, $\nabla_{\boldsymbol{c}} J$, $\nabla_{\boldsymbol{W}} J$, and $\nabla_{\boldsymbol{b}} J$ of the cost J with respect to $\boldsymbol{V}$, $\boldsymbol{c}$, $\boldsymbol{W}$, and $\boldsymbol{b}$, respectively.

(4 marks)

(g) Find the updated values of $\boldsymbol{V}$, $\boldsymbol{c}$, $\boldsymbol{W}$, and $\boldsymbol{b}$.

(2 marks)

3. (a) Given an input volume of size $3 \times 225 \times 225$, we have 128 convolution filters each with a size of $3 \times 3 \times 3$, a stride $= 2$, and no padding.

(i) What is the output volume size?

(2 marks)

(ii) What is the number of parameters in this layer? Be reminded to account for the bias terms.

(2 marks)

(iii) What is the size of output volume if depthwise convolution is performed? Assume the spatial size of a filter is 3×3, stride $= 2$, and no padding.

(2 marks)

(b) (i) Give a reason why one would use a 1×1 convolution.

(2 marks)

(ii) What would you set the padding of a two-dimensional convolutional layer to be (as a function of the filter width f) to ensure that the output has the same dimension as the input? Assume the stride is 1.

(2 marks)

(iv) You wish to train a convolutional neural network for classifying 12 different classes of flowers. You only have very limited training data, say 100 samples per class. Describe the steps to perform transfer learning to leverage a large-scale training dataset like ImageNet (which has 1000 classes).

(5 marks)

Note: Question No. 3 continues on Page 5

(c) An autoencoder has three neurons at the input layer and two neurons at the hidden layer. All the neurons have *sigmoid* activation functions. The weight matrix $\boldsymbol{W}$ of the hidden layer, the bias vector $\boldsymbol{b}$ of the hidden layer and the bias vector $\boldsymbol{c}$ of the output layer are given by

$$\boldsymbol{W} = \begin{pmatrix} 0.2 & -0.2 \\ 0.0 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 0.5 \\ -0.5 \end{pmatrix} \text{ and } \boldsymbol{c} = \begin{pmatrix} -0.5 \\ 0.2 \\ -0.5 \end{pmatrix}.$$

Reverse mapping from the hidden layer to the output is constrained to be the same as the input to hidden-layer mapping.

Consider the two inputs $\boldsymbol{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ and $\boldsymbol{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$,

(i) Find the hidden layer activations.

(4 marks)

(ii) Find the outputs of the autoencoder. Assume the decision threshold is 0.5.

(3 marks)

(iii) Describe a way to encourage an autoencoder to learn sparse hidden structure.

(3 marks)

4. (a) Consider a Jordan-type recurrent neural network (RNN) that receives 2-dimensional input patterns $\boldsymbol{x} \in \boldsymbol{R}^2$ and has one hidden layer. The RNN has two neurons in the hidden layer (which are initialized to zeros) and one neuron in the output layer. The hidden layer neurons have *tanh* activation functions and the output layer neurons use *sigmoid* activation functions.

The weight matrices $\boldsymbol{U}$ connecting the input to the hidden layer, the top-down recurrence weight matrix $\boldsymbol{W}$, and $\boldsymbol{V}$ connecting the hidden output to the output layer are given by

$$\boldsymbol{U} = \begin{pmatrix} 0.4 & 0.1 \\ 0.5 & 0.4 \end{pmatrix}, \boldsymbol{W} = \begin{pmatrix} 1.0 & 0.5 \end{pmatrix} \text{ and } \boldsymbol{V} = \begin{pmatrix} 0.3 \\ -0.3 \end{pmatrix}$$

All bias connections to neurons are set to 0.1. The output layer is initialized to an output of 2.0.

Note: Question No. 4 continues on Page 6

Find the output of the network for an input sequence $(\boldsymbol{x}(1), \boldsymbol{x}(2), \boldsymbol{x}(3))$:

$$\boldsymbol{x}(1) = \begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix}, \boldsymbol{x}(2) = \begin{pmatrix} 0.5 \\ -1.0 \end{pmatrix}, \text{ and } \boldsymbol{x}(3) = \begin{pmatrix} 2.0 \\ -2.0 \end{pmatrix}.$$

(13 marks)

(b) Answer “TRUE” or “FALSE” to the following statements:

(i) Layer normalization is usually used in Transformers.

(ii) Positional encoding helps to stabilize the training of Transformers.

(iii) Self-attention is performed in the decoder of Transformers.

(iv) Transformer model pays attention to a single most important word in a sentence.

(v) The feedforward network in the self-attention layer is not applied independently to each position.

(vi) Positional encoding has the same dimension as the input embedding.

(6 marks)

(c) Explain an advantage of attention-based models over recurrent-based ones.

(3 marks)

(d) Describe a method to remedy mode collapse during training of Generative Adversarial Networks (GANs).

(3 marks)

END OF PAPER

CE4042 NEURAL NETWORK & DEEP LEARNING
CZ4042 NEURAL NETWORK & DEEP LEARNING

Please read the following instructions carefully:

1. **Please do not turn over the question paper until you are told to do so. Disciplinary action may be taken against you if you do so.**

2. You are not allowed to leave the examination hall unless accompanied by an invigilator. You may raise your hand if you need to communicate with the invigilator.

3. Please write your Matriculation Number on the front of the answer book.

4. Please indicate clearly in the answer book (at the appropriate place) if you are continuing the answer to a question elsewhere in the book.