NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER 2 EXAMINATION 2023-2024

SC4001/CE4042/CZ4042 – NEURAL NETWORKS AND DEEP LEARNING

Apr/May 2024 Time Allowed: 2 hours

INSTRUCTIONS

1. This paper contains 4 questions and comprises 7 pages.
2. Answer **ALL** questions.
3. This is an open-book examination.
4. All questions carry equal marks.

---

1. (a) State whether each of the following statements is "TRUE" or "FALSE" with explanations. Each subquestion carries one mark.

   (i) The primary reason that gradient descent, instead of directly solving the optimization problem via matrix inversion, is used to solve neural networks is that it is too computationally expensive for large neural networks.

   (ii) You design a fully connected neural network architecture where all activations are sigmoids. The weights should be initialized with large positive numbers.

   (iii) Whether training data is shuffled or not is unimportant when you are doing gradient descent (GD) using the entire training set.

   (iv) We can use the logistic regression model for regression tasks.

   (v) Dropout is implemented differently during training and testing.

   (vi) Making your network deeper by adding more layers will always reduce the training loss.

Note: Question No. 1 continues on Page 2

(vii) You are training a neural network and notice that the validation error is significantly lower than the training error. This is because you are underfitting.

(7 marks)

(b) Suppose your task is to predict whether a medical image contains a TUMOR (1) or NO TUMOR (0). Now you are given predictions by n doctors, and your objective is to give a single prediction for each medical image as accurate as possible. To this end, you are expected to implement a *majority voting* algorithm – if more than half of the doctors predict TUMOR, then your final prediction should be TUMOR; otherwise, the final prediction should be NO TUMOR.

(i) Design a neural network architecture to implement this majority voting algorithm, with the assumption that we have a total of 4 doctors (named A,B,C,D). Specify the network structure and weights.

(5 marks)

(ii) Detail the learning process for at least one weight with the loss function specified and overfitting taken into consideration. At least one step of back-propagation (BP) should be run.

(5 marks)

(iii) Explain shortly (1-2 lines) how to generalize the network structure and weights to a more general case of n doctors.

(3 marks)

(iv) Suppose that every doctor is based on completely different knowledge, and they all have the same accuracy level (say, 75%) on the given set of medical images. Can majority voting improve the overall accuracy in this case on this same set of medical images? Explain your answer shortly.

(5 marks)

2. (a) Perform forward propagation on the 3-layer neural network in Figure Q2a for $x = 1$ by providing the outputs of the neurons $u_1, u_2, u_3, v_1, v_2, v_3, y$. Note that $u_1, u_2, \ldots, y$ are outputs after performing the appropriate operation as indicated in the node.

(7 marks)

Note: Question No. 2 continues on Page 3

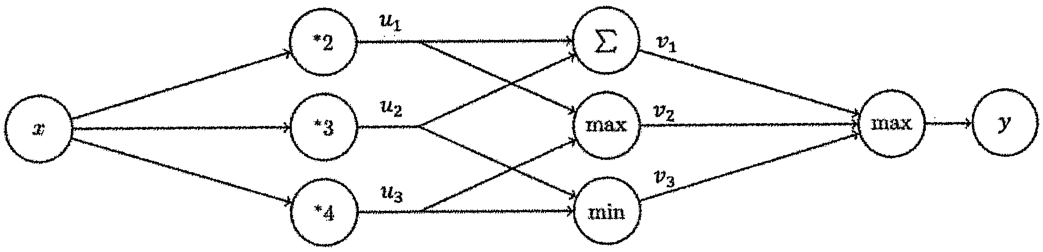

Figure Q2a

(b) Below is a neural network with weights w_1, w_2, w_3, w_4, w_5, and w_6. The inputs are x_1 and x_2.

The first hidden layer computes $r_1 = max(w_1 \cdot x_1 + w_3 \cdot x_2, 0)$ and $r_2 = max(w_2 \cdot x_1 + w_4 \cdot x_2, 0)$.
The second hidden layer computes $s_1 = \frac{1}{1+\exp(-w_5 \cdot r_1)}$ and $s_2 = \frac{1}{1+\exp(-w_6 \cdot r_2)}$.
The output layer computes $y = s_1 + s_2$. Note that the weights w_1, w_2, w_3, w_4, w_5, and w_6 are indicated along the edges of the neural network here.
Suppose the network has inputs $x_1 = 1, x_2 = -1$.
The weight values are $w_1 = 4, w_2 = 1, w_3 = 2, w_4 = 2, w_5 = 1, w_6 = 1$.
Forward propagation then computes $r_1 = 2, r_2 = 0, s_1 = 0.9, s_2 = 0.5, y = 1.4$. Note: some values are rounded.

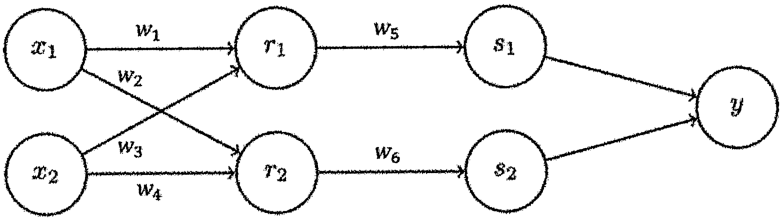

Figure Q2b

Using the values computed from forward propagation, use backpropagation to numerically calculate the partial derivatives $\frac{\partial y}{\partial w_k}$ $(k = 1, 2, \ldots, 6)$. Write your answers as single numbers (not expressions). You do not need a calculator. Use scratch paper if needed.

(12 marks)

Note: Question No. 2 continues on Page 4

(c) Below are two plots with horizontal axis x_1 and vertical axis x_2 containing data labelled × and •. For each plot, we wish to find a function $f(x_1, x_2)$ such that $f(x_1, x_2) \geq 0$ for all data labelled × and $f(x_1, x_2) < 0$ for all data labelled •. Give at least two possible expressions such that all the data is labelled correctly for each plot.

(6 marks)

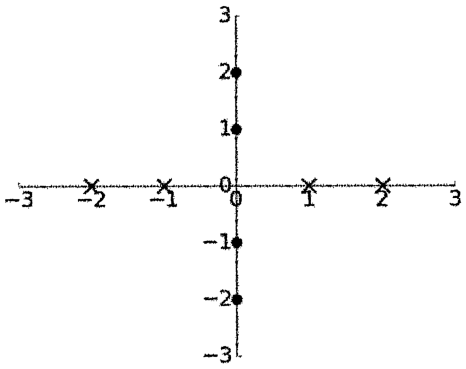

Figure Q2c

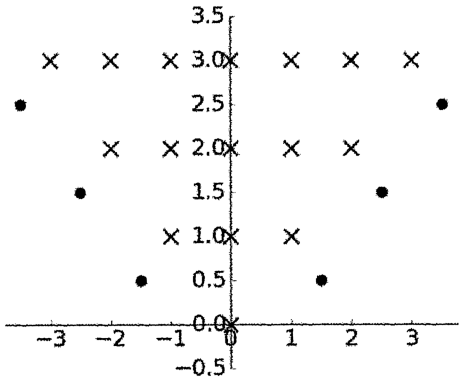

Figure Q2d

3. (a) Consider a convolutional neural network (CNN) designed for image classification. The network consists of the following layers in order:

Input: Grayscale images of size 32x32 pixels.

1) Convolutional layer 1: Uses 6 filters of size 5x5 with a stride of 1 and no padding. A ReLU activation function is used.
2) Max pooling layer: Uses a 2x2 filter with a stride of 2.
3) Convolutional layer 2: Uses 16 filters of size 5x5 with a stride of 1 and no padding. A ReLU activation function is used.
4) Max pooling layer: Uses a 2x2 filter with a stride of 2.
5) Fully connected layer: Has 120 neurons with ReLU activation functions.
6) Fully connected layer: Has 10 neurons (corresponding to 10 classes) with softmax activation functions.

(i) Calculate the dimensions of the output feature map after each layer.

(6 marks)

(ii) How many parameters are there in this CNN? Show your calculations. Be reminded to account for bias terms.

(5 marks)

Note: Question No. 3 continues on Page 5

(iii) Explain the role of the ReLU activation function in this network. What advantage does it have over other activation functions like sigmoid or tanh?

(3 marks)

(b) Consider a simple convolutional neural network with a single convolutional layer. The input to this layer is a 3x4 grayscale image, and the layer uses a single 2x2 filter with weights $\boldsymbol{w} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and no bias. The convolution operation uses a stride of 1 and no padding.

(i) Given the input image:

$$\begin{pmatrix} -0.3 & 0.7 & -0.2 & 0.4 \\ 0.6 & -0.8 & 0.9 & -0.1 \\ -0.5 & 0.2 & -0.6 & 0.1 \end{pmatrix}$$

Compute the output feature map resulting from the convolution operation.

(4 marks)

(ii) If a ReLU activation function is applied to the output feature map, what would be the final output?

(2 marks)

(c) The statements below are all related to autoencoders. Answer "TRUE" or "FALSE" to the following statements. Each part carries 1 mark.

(i) Autoencoders are primarily used for dimensionality reduction and cannot be used for classification tasks.

(ii) Autoencoders can be used for anomaly detection by reconstructing input data and measuring the reconstruction error.

(iii) Denoising autoencoders are trained to reconstruct the original input data from corrupted versions of the input.

(iv) Overcomplete autoencoders are more likely to capture the most salient features than undercomplete autoencoders.

(v) Sparse autoencoders encourage neuron sparsity via architecture design instead of loss function.

(5 marks)

4. (a) Consider an Elman-type recurrent neural network (RNN) that receives 2-dimensional input patterns $\boldsymbol{x}$ 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 neuron use a *Sigmoid* activation function.

The weight matrices $\boldsymbol{U}$ connecting the input to the hidden layer, $\boldsymbol{W}$ connecting the previous hidden state to the next hidden state, and $\boldsymbol{V}$ connecting the hidden output to the output layer are given by

$$\boldsymbol{U} = \begin{pmatrix} 0.5 & -0.3 \\ 1.0 & 0.8 \end{pmatrix}, \boldsymbol{W} = \begin{pmatrix} 1.0 & -0.5 \\ 0 & 0.6 \end{pmatrix}, \text{ and } \boldsymbol{V} = \begin{pmatrix} 0.2 \\ 1.0 \end{pmatrix}$$

The hidden layer bias vector $\boldsymbol{b}$ and the output layer bias c are given by

$$\boldsymbol{b} = \begin{pmatrix} 0.5 \\ 1.0 \end{pmatrix}, \text{ and } c = 0.2$$

Determine the output sequence of the RNN for an input sequence of $(\boldsymbol{x}(1), \boldsymbol{x}(2), \boldsymbol{x}(3))$ when

$$\boldsymbol{x}(1) = \begin{pmatrix} 1.0 \\ -0.5 \end{pmatrix}, \boldsymbol{x}(2) = \begin{pmatrix} -2.0 \\ 1.5 \end{pmatrix}, \text{ and } \boldsymbol{x}(3) = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}.$$

Provide your answers rounded to three decimal places.

(12 marks)

(b) The statements below are all related to Transformers. Answer "TRUE" or "FALSE" to the following statements. Each part carries 1 mark.

(i) Dividing the dot product of the query vector with the key vector by the square root of the dimension of the key vectors helps to stabilize gradients during training.

(ii) Multi-head attention inherently understands the sequential nature of the data, making positional encoding unnecessary.

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

(iv) Positional encoding only provides information about the relative position of the tokens in the sequence.

Note: Question No. 4 continues on Page 7

(v) When performing language translation using a Transformer, the model processes each word one at a time in a recurrent neural network manner.

(5 marks)

(c) (i) Explain the roles of the generator and the discriminator in a GAN. How do they interact during the training process?

(6 marks)

(ii) In a scenario where a GAN has been trained on a dataset containing images of digits 0-9 (such as the MNIST dataset), it is observed that the trained GAN generates images representing only the digits 0-8, with the digit 9 consistently missing from the generated outputs. What is the name of this phenomenon?

(2 marks)

END OF PAPER

CE4042 NEURAL NETWORK & DEEP LEARNING
CZ4042 NEURAL NETWORK & DEEP LEARNING
SC4001 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.