NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER 2 EXAMINATION 2023-2024

EE7207 – NEURAL NETWORKS & DEEP LEARNING

April / May 2024 Time Allowed: 3 hours

INSTRUCTIONS

1. This paper contains 4 questions and comprises 5 pages.
2. Answer all 4 questions.
3. All questions carry equal marks.
4. This is a closed book examination.
5. Unless specifically stated, all symbols have their usual meanings.
6. A list of Formulae is provided in Appendix A on page 5, respectively.

---

1. There are four vectors which are given below:

$$P_1 = [1 \quad -1 \quad 1 \quad 1]^T$$

$$P_2 = [1 \quad 1 \quad -1 \quad -1]^T$$

$$P_3 = [1 \quad 1 \quad 1 \quad 1]^T$$

$$P_4 = [-1 \quad -1 \quad -1 \quad -1]^T$$

(a) If a Hopfield neural network is used to store the four vectors, sketch the architecture of the Hopfield neural network, determine the weight matrix, and check whether the four vectors could be correctly retrieved by the Hopfield neural network. If some vectors cannot be correctly retrieved, analyse the causes of the problem.

(10 Marks)

Note: Question No. 1 continues on page 2.

(b) If a bi-directional associative memory (BAM) neural network is used to store the four vectors, sketch the architecture of the BAM neural network, determine the weight matrices, and check whether the four vectors could be correctly retrieved by the BAM neural network. If some vectors cannot be correctly retrieved, analyse the causes of the problem.

(10 Marks)

(c) Explain why Hopfield and BAM neural networks can be used to store fundamental memories.

(5 Marks)

2.

(a) Assume that Gaussian kernel support vector machine (SVM) is adopted in a study. After training the SVM, four support vectors are obtained:

$$\mathbf{s}(1) = [0.2575 \quad 0.0598]^T$$

$$\mathbf{s}(2) = [4.5052 \quad -0.4733]^T$$

$$\mathbf{s}(3) = [3.5651 \quad -1.3827]^T$$

$$\mathbf{s}(4) = [-0.2015 \quad -2.1464]^T$$

The Lagrange multipliers corresponding to the four support vectors are:

$$\alpha(1) = 0.3646, \alpha(2) = 0.0067, \alpha(3) = 0.029, \alpha(4) = 0.3289$$

The class labels of the four support vectors are $1, -1, -1, -1$, respectively.

Assume that the width of the Gaussian kernel is 1. Express the Gaussian kernel SVM classifier using the information given above, and use it to classify the following sample:

$$\mathbf{x} = [-0.2033 \quad -2.2553]^T$$

(8 Marks)

(b) Centre selection for hidden layer neurons is a key issue in RBF neural network training. List and briefly describe three methods of centre selection and discuss the advantages and disadvantages of each method.

(12 Marks)

(c) Shallow architecture is often adopted for multilayer perceptron (MLP) neural networks. Explain the rationale of using shallow architecture instead of deep architecture in MLP neural networks.

(5 Marks)

3.

(a) Discuss the challenges associated with traditional RNNs in learning long-term dependencies and elaborate on the advantages offered by LSTM and GRU.

(7 Marks)

(b) Explain the concept of transfer learning and discuss its benefits. Provide in detail an example application where transfer learning might be effective.

(12 Marks)

(c) Draw the encoder architecture of the typical transformer model. Explain the difference between BERT and GPT in terms of their model architectures. Give two examples of common use cases for BERT and GPT, respectively.

(6 Marks)

4.

(a) The choice of aggregation method is important for the expressiveness of a GCN model. Three common choices are:

$$\text{AGGREGATE}\left(\left\{h_u^{(k-1)}, \forall u \in \mathcal{N}(v)\right\}\right) = \max_{u \in \mathcal{N}(v)}\left(h_u^{(k-1)}\right) \quad (1)$$

$$\text{AGGREGATE}\left(\left\{h_u^{(k-1)}, \forall u \in \mathcal{N}(v)\right\}\right) = \frac{1}{|\mathcal{N}(v)|}\sum_{u \in \mathcal{N}(v)}\left(h_u^{(k-1)}\right) \quad (2)$$

$$\text{AGGREGATE}\left(\left\{h_u^{(k-1)}, \forall u \in \mathcal{N}(v)\right\}\right) = \sum_{u \in \mathcal{N}(v)}\left(h_u^{(k-1)}\right) \quad (3)$$

Please give an example of two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ and their node features, such that for a node $v_1 \in V_1$ and a node $v_2 \in V_2$ with the same initial features $h_{v_1}^0 = h_{v_2}^0$, the updated features $h_{v_1}^1$ and $h_{v_2}^1$ are equal if we use the aggregation methods (1) and (2), but are different if we use the aggregation method (3).

HINT: Your node features can be scalars or vectors. Also, you are free to choose any number of nodes (e.g., 3 nodes) and edges in your example.

(8 Marks)

Note: Question No. 4 continues on page 4.

(b) Consider a GAT layer where the attention mechanism is applied to a node i with two neighbours $j = \{1, 2\}$. Each node is associated with a 2-dimensional feature vector.

The feature vectors are $v_i = [2, 3]$, $v_1 = [1, 2]$, and $v_2 = [3, 1]$. The attention coefficients are calculated as $e_{ij} = LeakyReLU\left(a^T[Wh_i||Wh_j]\right)$, where a^T is an all-one vector, $W = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, and $||$ denotes concatenation. Please compute the normalized attention coefficients using the SoftMax function and then calculate the updated feature vector for node i after the attention-based aggregation.

(12 Marks)

(c) Consider the process of GAT facilitated by both the learnable weight matrix **W** and attention coefficients $\boldsymbol{e}$. Please explain the differences between **W** and $\boldsymbol{e}$.

HINT: You could discuss in terms of definition, purpose, or scope, etc.

(5 Marks)

Appendix A

$$\text{softmax}\,(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}}$$

$$\text{LeakyReLU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \text{negative\_slope} \times x, & \text{otherwise} \end{cases}$$

END OF PAPER

EE7207 NEURAL NETWORKS & 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.