AI61002 - Deep Learning Foundations and Applications

Class Test 1 - 06/02/2025

Instructions

1. Except electronic calculators, no other devices or materials are allowed during the exam.
2. All answers must be supported by calculations and/or explanations unless stated otherwise.

Question 1

Pick the correct option(s) in each of the following questions. Note that there may be more than one correct option associated with each question, and in such cases you must specify all the options to receive full credit. Explanations are not necessary for this question.

1. Suppose a 16 input XOR gate is designed with a multi-layer perceptron by pairing the inputs as follows:? [1]

   $(x_1 \oplus x_2) \oplus (x_3 \oplus x_4) \oplus \cdots \oplus (x_{15} \oplus x_{16})$

   What is the minimum number of neurons in the MLP that implements the above XOR function, and what is its depth?
   A. # of sufficient neurons = 257, depth = 4
   B. # of sufficient neurons = 45, depth = 8
   C. # of sufficient neurons = 75 depth = 4
   D. # of sufficient neurons = 45, depth = 4

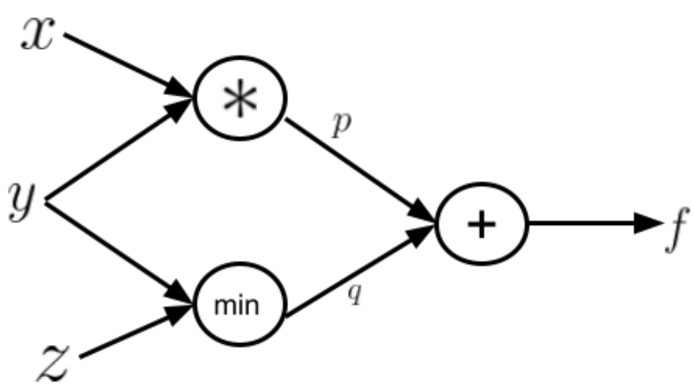
2. The primary difference between LeNet and AlexNet was: [1]
   A. AlexNet had more hidden layers
   B. LeNet uses ReLU activation, while AlexNet uses sigmoid activations
   C. AlexNet was trained using GPUs
   D. AlexNet employed dropout, unlike LeNet

3. Consider a neural network that has 10 layers and employs residual/skip connections between successive layers and also between the last layer and output. How many distinct paths of gradient flow exist between the input of layer 7 and the output obtained from layer 10? [1]
   A. 8

B. 6
C. 16
D. 3

Question 2

Consider the following computational graph and the associated functions it implements: [4]



$$p = xy$$
$$q = \min(y, z)$$
$$f = p + q$$

Given $x = 2$, $y = 0.55$ and $z = -1$, compute $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ and $\frac{\partial f}{\partial z}$ using backpropagation. You should re-draw the computational graph showing the relevant gradients flowing backward along the bottom of the edges in your solution.

Question 3

Consider a fully connected neural network with the following architecture

I. input layer having n input units
II. first hidden layer having k_1 hidden units
III. second hidden layer having k_2 hidden units
IV. the output layer having m output units
V. each unit having an associated bias component.

Further, we aggregate the weights and biases of the i^{th} hidden layer into a matrix W_i and a vector $\mathbf{b}_i$, respectively, such that $\mathbf{h}_i = g(W_i \mathbf{h}_{i-1} + \mathbf{b}_i)$, where the activation function $g(\mathbf{z})$ is applied element-wise on the components of $\mathbf{z}$, and $\mathbf{h}_0 = \mathbf{x} \in \mathbb{R}^n$

Based on the information provided above, answer the following questions about the network.

1. What are the dimensions of the weight matrices $\mathbf{W}_1, \mathbf{W}_2$ and $\mathbf{W}_3$? [2]

2. What are the dimensions of the bias vectors $\mathbf{b}_1, \mathbf{b}_2$ and $\mathbf{b}_3$? [1]

3. Compute the total number of learnable parameters of the network for $n = 5,\ k_1 = 8,\ k_2 = 4$ and $m = 2$. [1]

<u>Question 4</u>

Consider a neural network designed for multi-class classification with three classes. It uses softmax activation on the logits (pre-activations) z of the output layer to compute the class likelihood prediction for the i^{th} class as:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{k=1}^{3} e^{z_k}}, \quad i = 1,2,3$$

The model is trained using the categorical cross entropy loss $\ell(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{k=1}^{3} y_k \log(\hat{y}_k)$.

Given logits $z = \begin{bmatrix} 1.2 \\ 0.5 \\ -0.7 \end{bmatrix}$ and class label $\mathbf{y} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, answer the following questions:

1. Compute the predicted vector of class likelihoods $\hat{y}$. [1]

2. Compute the partial derivative of the softmax activation function with respect to z_i and z_j (for $j \neq i$). Give your answers in terms of $\hat{y}_i$ and $\hat{y}_j$. [3]

3. Compute the value of the loss function for the given label and prediction. [1]

4. Compute the gradient of the loss with respect to z (Hint: apply chain rule using the partial derivatives computed in part 2 of the problem). [4]