Question Paper

Exam Date & Time: 23-Jun-2022 (02:00 PM - 05:00 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences (MSIS), Manipal
Second Semester Master of Engineering - ME (Artificial Intelligence and Machine Learning) Information Science Degree Examination - June 2022

Deep Learning [AML 5202]

Marks: 100 **Duration: 180 mins.**

Thursday, June 23, 2022

Answer all the questions.

1) [10 points] [TLO 1.1, CO 1] Suppose we have 10^3 samples corresponding to 3 output labels and that each sample is a 3 × 28 × 28 color image. If we apply the softmax algorithm to this data, what are the dimensions of the following quantities assuming that the bias-trick pre-processing has been performed: (10)

- Data matrix;
- Weight matrix;
- Probability matrix;
- Adjusted probability matrix;
- Total average data loss;
- Regularization loss;
- Gradient of total loss w.r.t. the weight matrix?

2) [10 points] [TLO 1.2, CO 1] Consider the following dataset for binary classification: (10)

$$x^{(1)} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \; x^{(2)} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \; x^{(3)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$
$$y^{(1)} = 1, \; y^{(2)} = 0, \; y^{(3)} = 0.$$

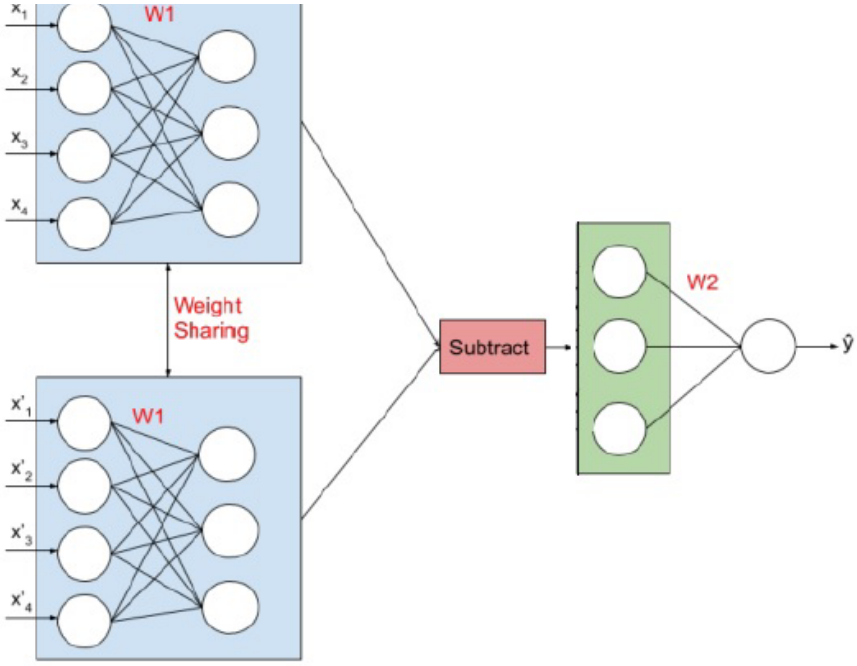
Calculate the SVM loss using bias-trick for the following weights and bias values:

$$W = \begin{bmatrix} 0.1 & -0.4 \\ 0.5 & -0.3 \end{bmatrix}, \; b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

3) [10 points] [TLO 3.2, CO 3] Consider the same dataset, weights, and bias values from the previous problem. Calculate the Softmax loss using a regularization strength of 0.3. (10)

4) [10 points] [TLO 2.2, CO 2] Using the same setup from the previous problem, perform one step of gradient descent with learning rate = 10^{-2} . Round all numbers to 2 decimal places. (10)

5) [10 points] [TLO 3.1, CO 2] A Siamese neural network consists of twin networks which accepts distinct inputs but share the same weights. The outputs of the twin networks are usually joined later on by one or more layers. The following image shows such a network with a pair of distinct input samples x and x^1 both of which are of size 4 × 1: (10)



Note the following:

- $x^{(i)}$ and $x'^{(i)}$ together constitute the ith input sample for the network;
- The weights for the first hidden layer is the same (W_1) as shown in the blue regions.

Fill in the question marks below for the forward propagation for the ith sample leading to the loss L_i:

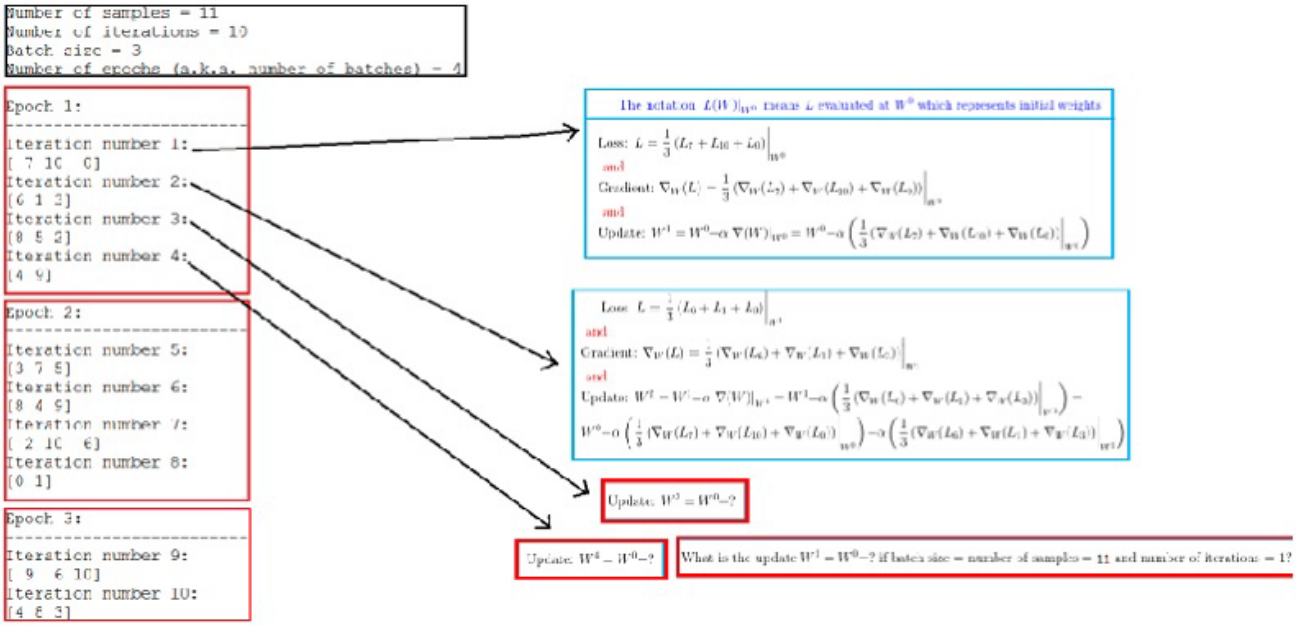
$$
\begin{aligned}
? &= W_1 x^{(i)} + b_1, \\
a_1 &= ?, \\
z_2 &= W_1 ? + ?, \\
? &= ReLU(z_2), \\
? &= a_1 - a_2, \\
z_3 &= ? a + b_2, \\
a_3 &= ?, \\
L_i &= -\left(y^{(i)} \times ? + (1 - ?) \times \log(1 - a_3)\right).
\end{aligned}
$$

6) [10 points] [TLO 3.1, CO 2] For the Siamese neural network from the previous problem, calculate the gradient $\nabla_{w2}(L_i)$ (10)

7) [10 points] [TLO 2.1, CO 2] During backpropagation, as the gradient flows backward through a sigmoid node, explain in not more than 2-3 lines as to whether the gradient will (10)

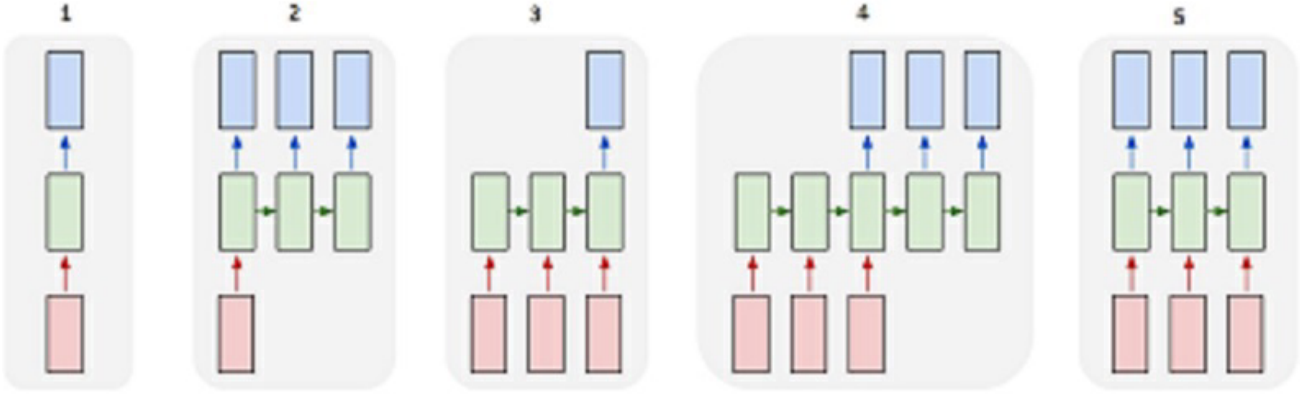
- increase or decrease in magnitude;
- maintain or reverse polarity (sign)

8) [10 points] [TLO 4.1, CO 3] In the schematic given below for batch processing, fill in the question marks in the red boxes: (10)



9) [10 points] [TLO 4.2, CO 4] What are the different types of recurrent neural networks? Give any two applications for many-to-many architecture when T_x ! = T_y. (10)

10) (10)



[10 points] [TLO 4.2, CO 4] Answer the questions following the RNN architectures shown below:

-
- Classify the type of connection in Network-2.
- Which RNN network is suitable for sentiment analysis?
- Which application suits for Network-3?
- Which network is suitable for named-entity recognition?
- True/false: Network-4 is an example for many-to-many RNN network.

-----End-----