Question Paper

Exam Date & Time: 02-May-2024 (10:00 AM - 01:00 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences (MSIS), Manipal
Second Semester Master of Engineering - ME (Artificial Intelligence and Machine Learning) Degree Examination - April / May 2024

Deep Learning [AML 5202]

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

Thursday, May 02, 2024

Answer all the questions.

1) [10 points] [L3, CO1] Consider the following dataset for binary classification using the softmax algorithm: (10)

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

Calculate the raw scores matrix using bias-trick applied to the following weights and bias values (label 0 followed by label 1), and using the calculated raw scores and without any further calculations, identify which class each sample will be predicted to belong to:

$$\mathbf{W} = \begin{bmatrix} 0.2 & 0.4 \\ -0.5 & 0.3 \end{bmatrix},\ \mathbf{b} = \begin{bmatrix} 0.01 \\ 0.01 \end{bmatrix}.$$

2) [10 points] [L5, CO2] Suppose we apply the softmax algorithm for classifying a square colour image (3 channels) into 1 of 10 possible output categories, examples of which are shown below (airplane is label-1, automobile is label-2 etc.) along with the weights matrix with the bias trick applied: (10)

$$W = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 & 10 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 10 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 100 \\ 1 & 1 & 1 & 1 & \dots & 1 & 1 & 10 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 10 \\ -1 & -1 & -1 & -1 & \dots & -1 & -1 & 10 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 10 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1000 \\ 1 & -1 & 1 & -1 & \dots & 1 & -1 & 10 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 10 \end{bmatrix}}_{10\times 3073}$$

(a) Use the shape of the weight matrix to calculate the shape of a sample image.

(b) The softmax model given by the current weight matrix is most biased towards which output category?

(c) The softmax model given by the current weight matrix is most likely to predict a given training sample as what output category? Explain briefly.

(d) Say true or false with a brief justification: the softmax model given by the current weight matrix is more biased towards an airplane than an automobile.

3) (10)

[10 points] [L5, CO1] Suppose $\mathbf{X}$ represents the data matrix (samples along columns) containing information about 100 individuals

with features $\begin{cases} \text{age,} \\ \text{annual income,} \\ \text{credit limit,} \\ \text{gender (female or male),} \\ \text{education level (graduate, high school, post graduate),} \end{cases}$ and categories $\begin{cases} \text{does not default payment,} \\ \text{defaults payment.} \end{cases}$

(a) Suppose we want to apply softmax classifier to the dataset. What will be the shape of the weights matrix $\mathbf{W}$ assuming that the bias trick has been done?

(b) What will be the shape of the raw scores matrix comprising the raw scores of all 100 samples?

(c) In plain English and using the data as context, explain what each of the following represents assuming indexing starts from 0:

$$w_{:,3},\ w_{0,:},\ w_{0,8},\ w_{1,1}.$$

(d) Suppose the gradient of the loss with respect to some weight parameter evaluated at its current value is 0. Justify what will happen to the loss if we increase that weight parameter a little bit while keeping the other parameters fixed? What if we decrease it a little bit?

4) [10 points] [L2, CO2] Consider the following forward propagation through a fully-connected deep neural network architecture (16 nodes in hidden layer) for a 64×64-grayscale image sample represented as vector $\mathbf{x}$ with one-hot encoded correct label 4-vector $\mathbf{y}$ and predicted probability vector $\hat{\mathbf{y}}$ (indexing starts from 0): (10)

$$\boxed{\underbrace{\mathbf{x}_B}_{?} = \begin{bmatrix}\mathbf{x}\\1\end{bmatrix}} \rightarrow \boxed{\underbrace{\mathbf{z}^{[1]}}_{?} = \underbrace{\mathbf{W}^{[1]}}_{?\times?}\mathbf{x}_B} \rightarrow \boxed{\underbrace{\mathbf{a}^{[1]}}_{?} = \tanh\left(\mathbf{z}^{[1]}\right)} \rightarrow \boxed{\mathbf{a}_B^{[1]} = \begin{bmatrix}\mathbf{a}^{[1]}\\1\end{bmatrix}} \rightarrow \boxed{\underbrace{\mathbf{z}^{[2]}}_{?} = \underbrace{\mathbf{W}^{[2]}}_{?\times?}\underbrace{\mathbf{a}_B^{[1]}}_{?}} \rightarrow \boxed{\mathbf{a}^{[2]} = \text{softmax}\left(\mathbf{z}^{[2]}\right)} \rightarrow \boxed{L = \sum_{k=0}^{?} -y_k \log\left(\hat{y}_k\right)}.$$

Identify the missing shapes corresponding to the question marks.

5) [10 points] [L5, CO3] Suppose we have the following raw scores vector corresponding to dense layer 4 of a deep neural network: (10)

$$\mathbf{z}^{[4]} = \begin{bmatrix}1\\-10\\1\\10\end{bmatrix}.$$

(a) Calculate the local gradient of activation layer 4 which is ReLU activated. Round your calculations to 2 decimal places.

(b) Clearly state for which nodes of dense layer 4, learning of the corresponding parameters will be very minimal? Justify your answer briefly. What could you do to improve the learning of the parameters of all nodes in dense layer 4?

6) [10 points] [L5, CO3] Consider the following initial weights matrix (assuming that the bias trick has been done) for dense layer l in a deep neural network: (10)

$$\mathbf{W} = \begin{bmatrix}0.01 & -0.01 & 0.08 & 0.02 & -0.05 & 0.1\\0.01 & -0.01 & 0.08 & 0.02 & -0.05 & 0.1\\0.01 & -0.01 & 0.08 & 0.02 & -0.05 & 0.1\\0.01 & -0.01 & 0.08 & 0.02 & -0.05 & 0.1\\0.01 & -0.01 & 0.08 & 0.02 & -0.05 & 0.1\end{bmatrix}.$$

(a) How many nodes are there in layer $l-1$ and layer l?

(b) Is there any issue with the initial values of the weights given here? Justify your answer in 1-2 lines.

(c) Calculate the $L2$-regularization loss for dense layer l.

7) (10)

[10 points] [L5, CO3] Suppose we want to implement a dropout layer after dense layer l of a deep neural network with a dropout probability of 0.2. Consider the following dropout-matrix:

$$\mathbf{D} = \begin{bmatrix}0.49 & 0.87 & 0.7 & 0.99 & 0.23 & 0.92\\0.86 & 0.89 & 0.66 & 0.76 & 0.21 & 0.01\\0.13 & 0.08 & 0.9 & 0.84 & 0.19 & 0.34\\0.48 & 0.16 & 0.96 & 0.32 & 0.91 & 0.88\\0.62 & 0.15 & 0.23 & 0.98 & 0.89 & 0.06\end{bmatrix}$$

(a) What is the number of neurons in dense layer l?

(b) What is the batch size?

(c) Each batch sample contributes to the learning of specific neurons of dense layer l. Identify those neurons for each batch sample and fill in the table below (counting starts from 0):

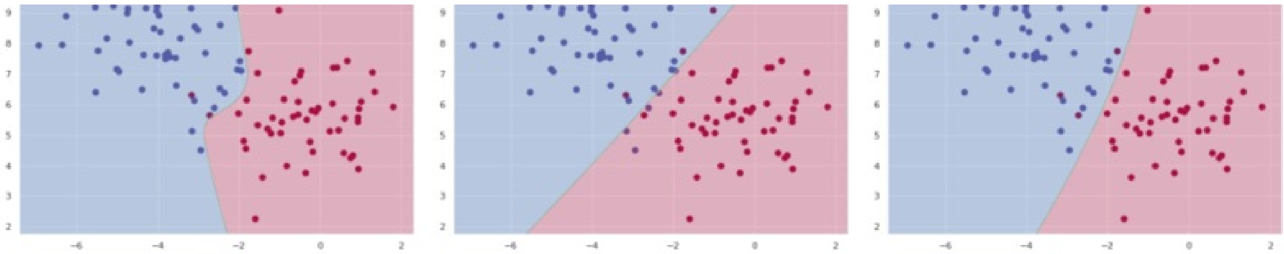
Batch sample	Neurons
0	?, ?, ...
⋮	⋮

(d) How does the activations vector for dense layer l for the 3rd sample denoted as $\mathbf{a}^{[l](3)}$ gets forward propagated through this dropout layer? Your answer should be a vector whose elements involve the elements of the vector $\mathbf{a}^{[l](3)}$.

(e) Compare and contrast dropout vs. loss-based regularization in not more than 2-3 lines.

(f) Is dropout applied to test data? In one line, justify your answer.

8) [10 points] [L5, CO3] Suppose we train three neural network models: the first one has 1 hidden layer, second one has 2 hidden layers, and the third one has 4 hidden layers. All three models have the same number of nodes in each of their hidden layers. Match the models with the outputs shown below and briefly justify your answer. Which model will benefit the most from regularization? Justify your answers briefly. (10)



9) [10 points] [L3, CO4] Consider the convolutional neural network defined by the layers in the left column in the table below. Fill in the shape of the output volume and the number of parameters corresponding to each layer using the notations below: (10)

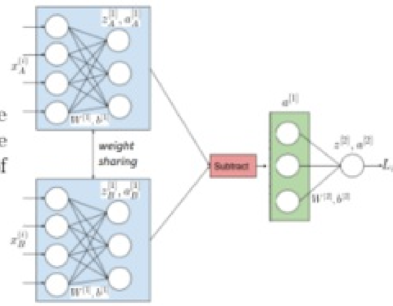
- CONVx-N denotes a convolutional layer with N filters with kernel height and width both equal to x. Padding is 2, and stride is 1.
- POOL-N denotes an $N \times N$ max-pooling layer with stride of N and no zero padding.
- FLATTEN flattens its inputs.
- FC-N denotes a fully-connected layer with N neurons.

Layer	**Output Volume Shape**	**Number of Parameters**
Input	$28 \times 28 \times 3$	0
CONV5-8	?	?
Leaky ReLU	?	?
POOL-2	?	?
FLATTEN	?	?
FC-4	?	?

10) (10)

[10 points] [L3, CO4]

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 image to the right shows such a network with a pair of distinct input samples $x_A^{(i)}$ and $x_B^{(i)}$ both of which are of size 4×1:



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

$$
\begin{aligned}
z_A^{[1]} &= ? + ?, \\
? &= ReLU\left(z_A^{[1]}\right), \\
z_B^{[1]} &= ? + b^{[1]}, \\
a_B^{[1]} &= ? \\
a^{[1]} &= ? - ?, \\
? &= W^{[2]} a^{[1]} + b^{[2]}, \\
a^{[2]} &= ?, \\
L_i &= -\left(y^{(i)} \times \log(?) + (1 - ?) \times \log\left(1 - a^{[2]}\right)\right).
\end{aligned}
$$

-----End-----