School of Engineering and Applied Science (SEAS), Ahmedabad University

Name: _____ Roll no.: _____

M. Tech. (Computer Science and Engineering): Semester II
Subject Name: Deep Learning, End Term Exam (Winter Semester :2025), Marks: 30
CSE602 **May 2, 2025, Friday, 8.30 AM – 9.30 AM**

Instructions:

1 All questions are compulsory.
2 Figures to right indicate full marks.
3 Use white space for calculations and explanations.
4 Assume suitable data wherever necessary and specify it clearly in your answers.

Q-1. Select the appropriate answers. **[10 Marks]**

1. Which among the below is suitable for calculating pixel-wise cross-entropy loss during model training?
   a. $-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\, log(\hat{y}_{i,c})$
   b. $-\sum_{i=1}^{N} log(\hat{y}_{i,y_i})$
   c. $-\sum_{i=1}^{N}\sum_{j=1}^{K} y_{ij} log(\hat{y}_{ij})$
   d. $-\frac{1}{N}\sum_{i=1}^{N}[y_i log(\hat{y}_i) + (1-y_i) log(1-\hat{y}_i)]$

2. What type of layers reduce spatial dimension?
   a. Convolutional layers with padding
   b. Pooling layers
   c. Fully connected layers
   d. Transpose convolution layers

3. How would you know if your model training is suffering from a vanishing gradient issue?
   a. Loss converges to zero
   b. Gradient becomes too small
   c. Loss fluctuates between 0 and 1
   d. Large learning rate

4. What deep learning model is suitable to reduce dimensions and visualise high-dimensional data?
   a. Convolutional neural networks
   b. Multi-headed self-attention
   c. Autoencoder
   d. Long short-term memory network

5. A deep learning model quickly converged to a suboptimal solution even though trained with Adam optimiser. What mistake could the ML engineer have made?

a. Set the learning rate to 0.001
b. Set the learning rate to 0.01
c. Set the learning rate to 0
d. Set the learning rate to 0.6

6. What parameter does a network architecture search technique optimise for EfficientDet?
   a. Input resolution
   b. Network depth
   c. Layer width
   d. All

7. What technique is the transformer network methodology adapted from?
   a. Attention mechanism
   b. CNN style processing
   c. Both
   d. None

8. Considering the query, key and value, how can the attention be represented?
   a. $A(q,k,v) = softmax\left(\frac{qk^T}{\sqrt{d_k}}\right)v$
   b. $A(q,k,v) = softmax\left(\frac{kq^T}{\sqrt{d_k}}\right)v$
   c. $A(q,k,v) = softmax\left(\frac{qk^T}{\sqrt{k_d}}\right)v$
   d. $A(q,k,v) = softmax\left(\frac{qk^T}{\sqrt{v_k}}\right)d$

9. Suppose we have a 5-layer neural network which takes 3 hours to train on a GPU with 4GB VRAM. At test time, it takes 2 seconds for single data point. Now we change the architecture such that we add dropout after 2nd and 4th layer with rates 0.2 and 0.3 respectively. What would be the testing time for this new architecture?
   a. Less than 2 sec
   b. Greater than 2 sec
   c. Exactly 2 sec
   d. Exactly 1 sec

10. Suppose an input to the pooling layer is given. The pooling size of neurons in the layer is (5, 5), and the strategy is max pooling. What would be the output of this pooling layer?

1	2	3	4	5
5	4	3	2	1
6	7	10	9	0
0	9	8	7	6
10	11	12	13	14

a. 10
b. 5
c. 14
d. 1

Q-2. Answer the below questions (four marks for each). **[20 Marks]**

I. Suppose there is a binary classification neural network with one hidden layer. The input $X \in R^{n \times m}$ and output $\hat{y} \in R^{1 \times m}$. The forward propagation equations are below.

$$z^{[1]} = W^{[1]}X + b^{[1]}$$
$$a^{[1]} = \sigma(z^{[1]})$$
$$\hat{y} = a^{[1]}$$
$$J = -\sum_{i=1}^{m} y^{(i)} \log(\hat{y}^{[i]}) + (1 - y^{(i)})(1 - \hat{y}^{(i)})$$
$$\frac{\partial J}{\partial \widehat{W}^{[1]}} = ???$$

(write your answer in terms of matrix product)

II. Suppose you have five convolutional kernels of size 7 x 7 with zero padding and stride 1 in the first layer of a convolutional neural network. You pass an input of dimension 224 x 224 x 3 through this layer. What are the dimensions of the data that the next layer will receive?

III. Suppose we have trained a deep detector. It detects bird species. The dataset consisted of images of 100 bird species. Now, you want to use this detector on a different dataset that only has images of parakeets. The task is to locate the parakeet in an image. What would be your approach?

a. Fine-tune only the last couple of layers and change the classification layer to the regression layer
b. Freeze all the layers except the classification layer, and fine-tune the last layer

IV. Select the data augmentation techniques from the list below?

a. Horizontal flipping
b. Random cropping
c. Random scaling
d. Color jittering
e. Random translation
f. Random shearing
g. Saturation
h. Adding noise

V. Which of the below is a good criteria for a good positional encoding algorithm?

a. It should output a unique encoding for each time-step (word's position in a sentence)
b. Distance between any two time steps should be consistent for all sentence lengths
c. The algorithm should be able to generalise to longer sentences
d. It must be deterministic

All the best
