National University of Computer and Emerging Sciences

Generative AI (AI-4009)
Date: November 04, 2024
Course Instructor
Akhtar Jamil

Sessional-II
Total Time: 1 Hour
Total Marks: 50
Total Questions: 03

Semester: FALL-2024
Campus: Islamabad

Student Name	Roll No	Section	Signature

Question No 1. MCQ [25 x 1 = 25]

Answer the MCQs on the given answer sheet attached at the end of the question paper. Answers marked on the question paper will not be evaluated.

1. **In GANs, if the distribution of generator (p_g) perfectly matches the distribution of real data p_{data}, then the Kullback-Leibler (KL) divergence between the two distributions $D_{KL}(p_{data} \parallel p_g)$ will be:**
   a). 0
   b). 1
   c). ∞
   d). Cannot be determined
2. **In the context of Diffusion Models, what is the primary purpose of the forward diffusion process?**
   a). To reduce the dimensionality of the data for easier processing
   b). To generate realistic data directly from random noise
   c). To gradually add noise to the data
   d). To improve model efficiency by removing unnecessary features
3. **In ViT, the positional embeddings are added to the patch embeddings to:**
   a). Increase the model's dimensionality for better learning
   b). Facilitate backpropagation through the transformer layers
   c). Improve the convergence speed of the model
   d). Encode the spatial structure of patches

4. **Why are the weights of the Discriminator and Generator updated in an alternating manner during GAN training?**
   a). To prevent premature convergence
   b). To maintain balanced learning between the Generator and Discriminator
   c). To reduce overfitting
   d). To encourage diverse outputs
5. **In StyleGAN, what does "stochastic variation" refer to?**
   a). The random changes in the overall structure of the generated image
   b). The fine details introduced in the image, such as hair, eye color, etc.
   c). The modification of the style vector across layers
   d). The adjustment of resolution at different stages of generation
6. **Fréchet Inception Distance measure can be used to:**
   a). Find similarity between images
   b). Check model training efficiency
   c). Measure resolution of generated images
   d). None of the above
7. **Which of the following is not relevant to Transformers?**
   a). Multi-Head Self-Attention
   b). Positional Encoding
   c). Convolutional Filters
   d). Layer Normalization
8. **Which of the following methods is not effective for addressing mode collapse in GANs?**
   a). Adding Noise to Inputs
   b). Regularization Techniques
   c). Using Different Architectures or Loss Functions
   d). Increasing Batch Size
9. **In Vision Transformers (ViT), what is the role of the class embedding?**
   a). To represent each class label as a unique embedding vector
   b). To generate attention maps for individual patches
   c). To apply positional encoding to each image patch
   d). To aggregate information from all patch embeddings

10. **In the context of a Conditional GAN, where x is the conditional input data, y is the target output, and z is the noise vector, which of the following terms in the loss function ensures that the generated output matches the label y?**

    a). $\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\| y - G(x,z) \|_1]$

    b). $\mathcal{L}_{cos}(G) = \mathbb{E}_{x,y,z}\left[1 - \cos\left(y, G(x,z)\right)\right]$

    c). $\mathcal{L}_{BCE}(G) = -\mathbb{E}_{x,y,z}\left[y \log\left(G(x,z)\right) + (1-y) log\left(1 - G(x,z)\right)\right]$

    d). None of the above

11. **In Vision Transformers (ViT), are the values of classification token learnable?**

    a). Yes

    b). No

12. **In StyleGAN, the "style" vector is introduced to:**

    a). Define the image resolution for the generator output

    b). Apply spatial transformations to image patches

    c). Limit the diversity of generated images

    d). Control the scale and translation of features in each layer

13. **How does disentangling the latent space benefit StyleGAN?**

    a). Reduces training time

    b). Allows independent control over features for generation

    c). Improves resolution

    d). Lowers computational cost

14. **What is the role of noise injection in StyleGAN's generator?**

    a). Control global features

    b). Add fine-grained stochastic details

    c). Ensure consistency

    d). Add redundancy

15. **If the model used for generating input embeddings also encodes positional information during embedding generation, then separate positional encodings are not needed for the transformer.**

    a). **True**

    b). False

16. **In CycleGANs, the condition (x) is applied only to the input of the generator.**

    a). True

b). False

17. In Conditional GANs, if the term $E_{x\sim p_{data}(x)}[\log D\,(x|y)]$ is high, what does this indicate?

a). The discriminator is struggling to distinguish real samples from generated samples.

b). The generator is producing samples that match the condition (y) accurately.

c). **The discriminator is confidently identifying real samples conditioned on (y) as real.**

d). The generator has successfully "fooled" the discriminator for most samples.

18. Which of the following problems cannot be effectively addressed using Conditional GANs?

a). Image-to-image translation

b). Text-to-image generation

c). Supervised image classification

d). Image inpainting with labels

19. How is Instance Normalization typically performed on a feature map in neural networks?

a). The entire batch is normalized along each column independently.

b). The entire batch is normalized across row independently.

c). Each feature map averaged across all instances

d). None of the above

20. In CycleGAN, let $G: X \rightarrow Y$ represent the generator mapping facial contours x to human faces (y) and ($F: Y \rightarrow X$) represent the inverse mapping. Which of the following is a suitable cycle consistency loss term to ensure that mapping x to y and back to x preserves the original facial contours?

a). $E_{x\sim p_{data}(x)}\left[|G\left(F(x)\right)-x|_1\right]$

b). $E_{x\sim p_{data}(x)}\left[|F\left(G(x)\right)-x|_1\right]$

c). $E_{x\sim p_{data}(x)}\left[|F\left(G(x)\right)+x|_1\right]$

d). $E_{x\sim p_{data}(x)}[|G(x)-x|_1]$

21. How many generators and discriminators are involved in the CycleGAN architecture?

a). 1 generator and 1 discriminator

b). 2 generators and 1 discriminator

c). 1 generator and 2 discriminators

d). 2 generators and 2 discriminators

22. GANs cannot be used for video generation.

a). True

b). False

23. After training a GAN, we can remove the discriminator and provide a Gaussian noise vector to produce an image during testing.

a). True

b). False

24. Assume that after training the generator of a GAN for 100 epochs, the probability distribution $p_g(x)$ of generated data matches with the real data probability distribution p_{data}. At this stage, if we randomly select a sample from real data and pass to the discriminator, what is the most probable output probability $D(x)$ that the discriminator should assign to x at this point?

a). 0

b). 0.5

c). 1

d). Undefined

25. Which of the following features is not available in Transformer models?

a). Recurrence and convolutions

b). Attention

c). Positional encoding

d). Sequence-to-sequence processing

Question No 2. [3 x 5=15]

Write short answers to the following questions.

1. Consider the Input vector $\mathbf{x} = [4, 6, 8]$, Style scaling vector $\mathbf{y_s} = [2.0, 1.5, 1.0]$ and Style bias vector $\mathbf{y_b} = [1.0, -1, 0.5]$.

Calculate the Adaptive Instance Normalization (AdaIN) output using the formula:

$$\text{AdaIN}(x, y) = y_s \cdot \frac{x - \mu(x)}{\sigma(x)} + y_b$$

Note: show all steps and in the last line write the final output.

Solution:

1. Calculate the mean $\mu(x)$:

$\mu(x) = \frac{4+6+8}{3} = 6$

2. Calculate the standard deviation $\sigma(x)$:

$$\sigma(x) = \sqrt{\frac{(4-6)^2 + (6-6)^2 + (8-6)^2}{3}} = \sqrt{\frac{4+0+4}{3}} = \sqrt{\frac{8}{3}} \approx 1.63$$

3. Normalize (x):

$$\frac{x - \mu(x)}{\sigma(x)} = \left[\frac{4-6}{1.63}, \frac{6-6}{1.63}, \frac{8-6}{1.63}\right]$$

$= [-1.23, 0, 1.23]$

4. Apply AdaIN formula:

$\text{AdaIN}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y_s} \cdot \frac{\boldsymbol{x} - \mu(\boldsymbol{x})}{\sigma(\boldsymbol{x})} + \boldsymbol{y_b}$

$= [2, 1.5, 1] \cdot [-1.23, 0, 1.23] + [1, -1, 0.5]$

$= [-2.46, 0, 1.23] + [1, -1, 0.5]$

$= [-1.46, -1, 1.73]$

AdaIN $\approx [-1.46, -1, 1.73]$.

2. In a transformer model for language translation, the decoder is trained using embeddings of both source and target language. However, during testing, we do not have access to the target sentence. Since the decoder still expects an input at each step, how can we handle this situation to enable the decoder to generate the translation correctly without having the true target embeddings?

During testing the decoder starts with a special START token. Using this token it actually generates the desired first output token. This process continues in an autoregressive manner where START token and generated tokens are feed back into the decoder to predict next token. The process stops when model predicts the END token.

3. What is Mixing Regularization as used in StyleGAN?

Mixing regularization in StyleGAN involves using two different latent codes. A random crossover point is selected among the layers of the model. Up to this crossover point, the model applies the first latent code, while for the remaining layers, it applies the second latent code. This process helps introduce diversity and improve the quality of generated images.

4. How Minibatch Discrimination introduces diversity in the generator's output in GANs?

In minibatch discrimination, we calculate similarities among the samples within the minibatch. These similarities are then incorporated as an additional input to the discriminator. If the discriminator detects high similarity scores among those samples, then the generator is penalized to create more diverse outputs. This motivates the generator to generate a variety of outputs that differ not only from real data but also from other generated samples.

5. Given two vectors, $\mathbf{v_1} = [2.1,\ 3.6,\ 4.5]$ and $\mathbf{v_2} = [1.9,\ 4.1,\ 2.8]$, and a threshold of 3, perform binarization on both vectors and then calculate the L2 norm between them to measure similarity.

Solution:

1. Semantic Hash:

$s_1 = [0,1,1]$

$s_2 = [0,1,0]$

2. Calculate L2 Norm (Euclidean distance):

$$\text{L2 norm} = \sqrt{(0-0)^2 + (1-1)^2 + (1-0)^2} = \sqrt{0+0+1} = \sqrt{1} = 1$$

Question No 3. [10]

Prepare the input for the transformer model. First generate the positional embeddings for the given input sequence and add with input embeddings. Using these final embeddings, calculate the attention scores according to the attention formula given below. Assume that all linear layer weights are all set to 0.5 and biases are set to 0 for each required layer.

$$\text{Input Embeddings} = \begin{bmatrix} 0.1 & 0.2 \\ 0.5 & 0.6 \\ 0.9 & 1.0 \end{bmatrix}$$

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

Step 1: Input Embeddings

$$\text{Input Embeddings} = \begin{bmatrix} 0.1 & 0.2 \\ 0.5 & 0.6 \\ 0.9 & 1.0 \end{bmatrix}$$

Positional Encodings

$$\text{PE}_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right)$$

$$\text{PE}_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d}}}\right)$$

$$\text{Positional Encodings} = \begin{bmatrix} 0 & 1 \\ 0.8415 & 0.5403 \\ 0.9093 & -0.4161 \end{bmatrix}$$

$$\text{Final Embeddings} = \begin{bmatrix} 0.1 & 0.2 \\ 0.5 & 0.6 \\ 0.9 & 1.0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0.8415 & 0.5403 \\ 0.9093 & -0.4161 \end{bmatrix}$$

$$\text{Final Embeddings} = \begin{bmatrix} 0.1 & 1.2 \\ 1.3415 & 1.1403 \\ 1.8093 & 0.5839 \end{bmatrix}$$

Compute Q, K, and V Matrices, using the weights (0.5) and biases (0), we calculate Q, K, and V from the final embeddings.

$$Q = K = V = 0.5 \times \text{Final Embeddings} = 0.5 \times \begin{bmatrix} 0.1 & 1.2 \\ 1.3415 & 1.1403 \\ 1.8093 & 0.5839 \end{bmatrix}$$

$$Q = K = V = \begin{bmatrix} 0.05 & 0.6 \\ 0.67075 & 0.57015 \\ 0.90465 & 0.29195 \end{bmatrix}$$

$$QK^T = \begin{bmatrix} 0.3625 & 0.3756275 & 0.2204025 \\ 0.3756275 & 0.7749765 & 0.773154 \\ 0.2204025 & 0.773154 & 0.903617 \end{bmatrix}$$

Approximating each division:

$$\frac{QK^T}{\sqrt{d_k}} \approx \begin{bmatrix} 0.2562 & 0.2656 & 0.1558 \\ 0.2656 & 0.5479 & 0.5467 \\ 0.1558 & 0.5467 & 0.6388 \end{bmatrix}$$

Apply Softmax

Next, we apply the softmax function row-wise to get the attention scor

Example for the first row:

$$\text{Softmax}(0.2562, 0.2656, 0.1558) = \left(\frac{e^{0.2562}}{\sum e^{\text{row}}}, \frac{e^{0.2656}}{\sum e^{\text{row}}}, \frac{e^{0.1558}}{\sum e^{\text{row}}} \right)$$

$$\text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \approx \begin{bmatrix} 0.343 & 0.346 & 0.310 \\ 0.346 & 0.424 & 0.389 \\ 0.310 & 0.389 & 0.443 \end{bmatrix}$$

$$\text{Attention Output} = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V$$

$$\text{Attention Output} = \begin{bmatrix} 0.343 & 0.346 & 0.310 \\ 0.346 & 0.424 & 0.389 \\ 0.310 & 0.389 & 0.443 \end{bmatrix} \times \begin{bmatrix} 0.05 & 0.6 \\ 0.67075 & 0.57015 \\ 0.90465 & 0.29195 \end{bmatrix}$$

$$\text{Attention Output} \approx \begin{bmatrix} 0.5297 & 0.4936 \\ 0.6536 & 0.5629 \\ 0.6772 & 0.5371 \end{bmatrix}$$

_____	_____	_____	_____
Roll No	**Name**	**Section**	**Student Signature**

Answer Sheet MCQs

Mark (X) for the correct option. Only one option must be selected. Selection of multiple options or overwriting will result in ZERO marks.

S.No.	A	B	C	D	S.No.	A	B	C	D
1.	X				14		X		
2.			X		15	X			
3.				X	16		X		
4.		X			17			X	
5.		X			18			X	
6.	X				19		X		
7.			X		20		X		
8.				X	21				X
9.				X	22		X		
10.	X				23	X			
11.	X				24		X		
12.				X	25	X			
13.		X							