

**Department of Computer Science and Engineering**  
**Indian Institute of Technology Jodhpur**  
**CSL 4020: Deep Learning**  
**Quiz 3**

**Total marks: 10**

**Answer Set**

**Time: 10 minutes**

---

**Instructions:**

- During the exam, invigilators will not address any queries. If you encounter anything unclear or incorrect in a question, make a reasonable assumption and proceed.
  - Each question with options may have multiple correct options. You have to mark all the correct options.
  - Mark the correct choices with a pen. No correction is allowed in the marking.
- 

**Answer all the questions. [Each carries 1 mark]**

1. Find out the correct statements about the basic vision transformers (ViT)
  - A. ViT has an encoder-decoder architecture
  - B. ViT does not require positional encoding
  - C. ViT does not require masked self-attention**
  - D. ViT does not have any skip connection
2. Find out the correct statements
  - A. Mode collapse is predominantly observed for multi-modal distributions**
  - B. Vanishing gradient problem does not occur in GANs
  - C. The discriminator of WGAN-GP takes the synthetic data from the generator directly as input
  - D. FID does not consider the distribution of real data
3. In DCGANs, what is used for upsampling in the generator network?
  - A. Max unpooling
  - B. Transposed convolutions (sometimes called deconvolutions)**
  - C. Bilinear interpolation
  - D. None of the others
4. Which components are essential for the functioning of a standard Transformer encoder block?
  - A. Multi-head self-attention**
  - B. Positional encoding**
  - C. Masked self-attention
  - D. Feed-forward network layers**

- 
5. What are the primary advantages of transformer-based models compared to traditional RNN-based approaches in NLP?
- A. Multiple contextual correlations of tokens can be explicitly learnt in transformers**
  - B. Tokens can be processed parallelly in transformers**
  - C. Reduced memory consumption in transformers
  - D. None of the others
6. What are the main advantages of WGAN-GP over WGAN?
- A. Provides more stable training dynamics by eliminating the weight clipping step**
  - B. None of the others
  - C. Eliminates the need for a critic network
  - D. Reduces the number of hyperparameters to tune
7. In WGAN-GP, suppose the required norm of the gradient of the critic's output with respect to an interpolated sample is calculated as 1.3. If the gradient penalty coefficient ( $\lambda$ ) is 10, what is the contribution of this sample to the total gradient penalty term?
- A. 0.82
  - B. 0.65
  - C. 0.90**
  - D. None of the others
8. Which of the following is/are the advantages of unsupervised pretraining?
- A. Better generalization with small labeled datasets**
  - B. Improved initialization for deep models**
  - C. Minimized overfitting without using regularization
  - D. Reduced number of trainable parameters
9. Which of the following defines the primary goal of domain adaptation
- A. Using labeled data to infer missing labels in the same domain
  - B. Adapting a model trained on data from one domain to perform well on data from a different domain**
  - C. Sharing model parameters across different tasks
  - D. None of the others
10. Zero-shot learning typically relies on
- A. Attribute-based class descriptions**
  - B. Extensive labeled training data for target classes
  - C. Class-attention
  - D. None of the others