Quiz Questions: GANs for Data Augmentation

Understanding Generative AI Assignment

Prashanth Talwar

December 2025

1 Introduction

These quiz questions assess understanding of Generative Adversarial Networks (GANs) and their application in data augmentation. The questions are designed to test three levels of knowledge:

1. **Recall (Questions 1-5):** Basic concepts, terminology, and fundamental principles
2. **Application (Questions 6-10):** Practical scenarios and implementation decisions
3. **Analysis (Questions 11-15):** Deep evaluation, comparison, and critical thinking

The learning objectives addressed include:

- Understanding GAN architecture and training dynamics
- Recognizing appropriate use cases for GAN-based data augmentation
- Evaluating GAN performance and addressing training challenges
- Comparing GANs with other generative models

2 Recall Questions (Basic Concepts & Terminology)

Question 1: What are the two main components of a GAN?

A) Encoder and Decoder

B) Generator and Discriminator

C) Encoder and Classifier

D) Feature Extractor and Predictor

E) Transformer and Attention Mechanism

Correct Answer: B

Explanation: GANs consist of two neural networks: the Generator, which creates synthetic data from random noise, and the Discriminator, which evaluates whether data is real or fake. This adversarial setup drives both networks to improve continuously. Option A describes autoencoders. Option C is not a standard pairing in generative models. Option D relates to general machine learning architectures. Option E refers to transformer-based models, not GANs.

Reference: Notebook Section 1.1 "What are GANs?" - Introduction to Generative Adversarial Networks

Question 2: In GAN training, what does the Generator network do?

A) Classifies images into different categories

B) Takes synthetic images and determines if they are realistic

C) Takes random noise as input and generates synthetic data

D) Extracts features from input images for classification

E) Optimizes the discriminator's loss function

Correct Answer: C

Explanation: The Generator's role is to transform random noise (typically from a normal distribution) into synthetic data samples that resemble the training data. It does not classify images (Option A), evaluate realism (Option B), extract features for classification (Option D), or directly optimize the discriminator's loss (Option E).

Reference: Notebook Section 1.1 "What are GANs?" and Section 1.3 "The GAN Architecture"

Question 3: What is "mode collapse" in GAN training?

A) When the discriminator becomes too powerful and stops providing useful gradients

B) When the generator produces limited variety of samples, failing to capture the full data distribution

C) When both networks converge to identical weights

D) When training loss increases exponentially

E) When the model runs out of memory during training

Correct Answer: B

Explanation: Mode collapse occurs when the generator learns to produce only a limited variety of samples, failing to capture the diversity of the real data distribution. This is visualized in Section 3.1 where the generator only captures one mode instead of all three modes in a multimodal distribution. Option A describes vanishing gradients. Option C is not a recognized GAN problem. Option D describes divergence, not mode collapse specifically. Option E is a hardware limitation, not a training phenomenon.

Reference: Notebook Section 3.1 "Training Challenges" - Mode Collapse visualization and explanation

Question 4: What activation function is typically used in the output layer of a DCGAN generator for image generation?

A) ReLU

B) Sigmoid

C) Tanh

D) Softmax

E) LeakyReLU

Correct Answer: C

Explanation: The tanh activation function is used in the generator's output layer because it produces values in the range [-1, 1], which matches the normalized range of image data in DCGAN implementations. ReLU (Option A) is used in hidden layers. Sigmoid (Option B) produces values in [0, 1], which is less common for image generation. Softmax (Option D) is for classification. LeakyReLU (Option E) is used in the discriminator and generator's hidden layers.

Reference: Notebook Section 2.1 "Deep Convolutional GAN (DCGAN)" - Architecture guidelines and generator code implementation

Question 5: What type of loss function is commonly used in GAN training?

A) Mean Squared Error (MSE)

B) Binary Cross-Entropy

C) Categorical Cross-Entropy

D) Hinge Loss

E) Huber Loss

Correct Answer: B

Explanation: Binary Cross-Entropy (BCE) is the standard loss function for GANs because the discriminator performs binary classification (real vs. fake). This is explicitly shown in the code implementation where `tf.keras.losses.BinaryCrossentropy()` is used. MSE (Option A) is for regression tasks. Categorical Cross-Entropy (Option C) is for multi-class classification. While Hinge Loss (Option D) is used in some GAN variants like Wasserstein GANs, BCE is more common. Huber Loss (Option E) is for robust regression.

Reference: Notebook Section 1.3 "Define Loss Functions and Optimizers" - Loss function definition and implementation

3 Application Questions (Practical Scenarios)

Question 6: You're training a GAN on the MNIST dataset and notice the generator loss is consistently around 0.75 while the discriminator loss is around 1.35. What does this indicate?

A) The model has completely failed and should be restarted

B) The generator is overpowering the discriminator

C) The training is relatively stable and balanced

D) Mode collapse is occurring

E) The learning rate is too high

Correct Answer: C

Explanation: The training results shown in the notebook demonstrate generator loss stabilizing around 0.74-0.76 and discriminator loss around 1.34-1.36, indicating balanced, stable training. The discriminator is challenged but not overwhelmed, which is ideal. Option A is incorrect as these losses indicate successful training. Option B would show discriminator loss near 0. Option D would show unstable or diverging losses. Option E would typically cause both losses to oscillate wildly.

Reference: Notebook Section 1.5 "Discussion of MNIST Results" - Training Performance analysis and loss curve interpretation

Question 7: When would you choose to use a Conditional GAN (CGAN) over a standard GAN for data augmentation?

A) When you need faster training times

B) When you want to generate samples from specific classes to address class imbalance

C) When working with very large datasets

D) When the discriminator is too powerful

E) When you want to reduce model complexity

Correct Answer: B
Explanation: Conditional GANs allow control over the generation process by conditioning on class labels, making them ideal for generating specific classes to balance imbalanced datasets. This is explicitly discussed in Section 4.2. Option A is incorrect as CGANs are typically more complex. Option C doesn't relate to the conditional aspect. Option D describes a training issue, not a use case for CGANs. Option E is false as CGANs add complexity through embedding layers.
Reference: Notebook Section 2.2 "Conditional GAN (CGAN)" and Section 4.2 "Addressing Class Imbalance"

Question 8: You're implementing a DCGAN and decide between using strided convolutions vs. pooling layers in the discriminator. Which approach follows DCGAN best practices?

A) Use max pooling layers for downsampling

B) Use average pooling layers for downsampling

C) Use strided convolutions for downsampling

D) Use a combination of both pooling and strided convolutions

E) Use neither; keep the spatial dimensions constant

Correct Answer: C
Explanation: DCGAN architecture guidelines explicitly recommend replacing pooling layers with strided convolutions in the discriminator (and transposed convolutions in the generator). This is one of the key architectural innovations that made DCGANs successful. Options A and B contradict DCGAN principles. Option D unnecessarily complicates the architecture. Option E would prevent the discriminator from processing features at multiple scales.
Reference: Notebook Section 2.1 "Deep Convolutional GAN (DCGAN)" - Architecture guidelines stating "Replace pooling layers with strided convolutions"

Question 9: Your GAN-generated images for Fashion-MNIST look blurrier and less distinct than those generated for MNIST. What is the most likely explanation?

A) The Fashion-MNIST dataset is corrupted

B) Fashion-MNIST has higher complexity with greater intra-class variability

C) The learning rate should be increased significantly

D) MNIST data is larger in file size

E) The generator architecture needs more dropout layers

Correct Answer: B

Explanation: The notebook explicitly compares MNIST and Fashion-MNIST results, noting that Fashion-MNIST's higher complexity (more complex structures, greater variability, textural details) makes it harder for GANs to generate clear images. This is a dataset characteristic issue, not a corruption problem (Option A). Increasing learning rate (Option C) could destabilize training. Option D is factually incorrect. Adding dropout to the generator (Option E) would likely worsen results.

Reference: Notebook Section 2.5 "Discussion of Fashion-MNIST Results" and "Comparison of Results Between MNIST and Fashion-MNIST"

Question 10: When implementing data augmentation with GANs for a classification task with severe class imbalance, what should be your primary concern?

A) Ensuring the generated samples are visually perfect

B) Generating samples that preserve the statistical properties and diversity of the minority class

C) Training the GAN for the maximum number of epochs possible

D) Using the largest possible batch size

E) Minimizing generator loss to exactly zero

Correct Answer: B

Explanation: For data augmentation to be effective, generated samples must preserve the statistical properties and diversity of the minority class while being realistic enough to help the classifier generalize. Visual perfection (Option A) is less critical than distribution matching. More training (Option C) can lead to overfitting. Batch size (Option D) is a hyperparameter choice, not the primary concern. Zero generator loss (Option E) typically indicates training problems or discriminator failure.

Reference: Notebook Section 4.1 "Why Use GANs for Data Augmentation?" and Section 4.2 "Addressing Class Imbalance"

4 Analysis Questions (Deep Evaluation)

Question 11: Compare the training stability of GANs on MNIST versus Fashion-MNIST. Which statement best explains the difference in stability?

A) MNIST has more training samples, leading to better convergence

B) Fashion-MNIST's higher intra-class variability and structural complexity make it harder for the GAN to learn a stable mapping from noise to images

C) The MNIST dataset uses a different file format that's easier to process

D) Fashion-MNIST requires different activation functions

E) MNIST digits are colored while Fashion-MNIST is grayscale

Correct Answer: B

Explanation: The notebook's comparison reveals that Fashion-MNIST shows "more oscillation in loss values" and lower image quality due to its inherent complexity: clothing items have complex structures (collars, sleeves), greater variability within classes, and textural information. This makes the generator's task significantly harder, leading to less stable training. Option A is incorrect as both datasets have 60,000 training images. Option C is irrelevant. Option D is false; the same architectures work for both. Option E is backwards; both are grayscale.

Reference: Notebook "Comparison of Results Between MNIST and Fashion-MNIST" section with loss curve analysis and "Key observations from the comparison"

Question 12: Why might a GAN with excellent FID (Fréchet Inception Distance) scores still be unsuitable for data augmentation in a classification task?

A) FID only measures image quality, not whether generated samples cover the full diversity of the data distribution

B) FID scores are always inversely related to classification performance

C) Low FID automatically means the GAN has experienced mode collapse

D) FID cannot be calculated for augmented datasets

E) Classification tasks don't benefit from synthetic data

Correct Answer: A
Explanation: FID measures how similar generated and real distributions are in feature space but doesn't guarantee coverage of all modes in the distribution. A GAN could achieve low FID by generating high-quality samples from only a subset of classes or modes (mode collapse), making it poor for augmentation. Option B is incorrect; FID and classification performance are related but not inversely. Option C is backwards; low FID typically indicates better diversity. Option D is false. Option E contradicts the entire premise of GAN-based augmentation discussed in Section 4.
Reference: Notebook Section 5 "Evaluation Metrics for GANs" - FID description and Section 5.3 "Precision and Recall" discussing quality vs. coverage tradeoffs

Question 13: Analyze the following scenario: After training a GAN for data augmentation, you add 5,000 synthetic samples to a dataset with only 500 real minority class samples. The classifier's performance on the minority class decreases. What is the most likely explanation?

A) GANs always harm classification performance

B) The generated samples are of poor quality or don't capture the true distribution, introducing noise that confuses the classifier

C) 5,000 samples is too few for effective augmentation

D) The synthetic samples are too similar to the real samples

E) The classifier needs a different architecture

Correct Answer: B
Explanation: When synthetic samples are low quality or fail to capture the true distribution (possibly due to mode collapse or insufficient training), they can mislead the classifier rather than help it. The 10:1 ratio of synthetic to real data amplifies any distributional mismatch. Option A is contradicted by Section 4.3's demonstration of augmentation benefits. Option C suggests more data, but poor quality data doesn't help regardless of quantity. Option D would not typically harm performance. Option E deflects from the augmentation quality issue.
Reference: Notebook Section 4.3 "Evaluating Augmentation with a Classifier" and Extension 2 "Data Augmentation for a Small Classifier" discussing quality requirements

Question 14: In the minimax objective function for GANs, what does reaching a Nash equilibrium theoretically represent?

A) The generator produces perfect copies of training samples

B) The discriminator achieves 100% accuracy

C) The generator produces samples indistinguishable from real data, and the discriminator can only guess randomly (50% accuracy)

D) Both networks have identical weights

E) Training loss reaches exactly zero for both networks

Correct Answer: C
Explanation: At Nash equilibrium, the generator has learned to perfectly replicate the real data distribution, making its outputs indistinguishable from real samples. At this point, the discriminator can only guess randomly, achieving $\sim$50% accuracy. This is the theoretical goal described in Section 1.4. Option A describes overfitting/memorization. Option B would indicate the generator has failed. Option D is not a GAN concept. Option E misunderstands the loss dynamics; discriminator loss around 0.69 (log(2)) indicates random guessing.
Reference: Notebook Section 1.4 "GAN Training Dynamics" - Minimax game explanation and convergence criteria

Question 15: When comparing evaluation metrics for GANs (Inception Score, FID, Precision/Recall), which statement best describes their complementary nature?

A) All three metrics measure exactly the same thing, so only one is needed

B) Inception Score measures quality and diversity, FID measures distribution similarity, and Precision/Recall separately measure quality vs. coverage—using multiple metrics provides a more complete picture

C) FID is superior and makes the other metrics obsolete

D) Precision/Recall can only be used for classification tasks, not generation

E) These metrics are incompatible and cannot be used on the same GAN

Correct Answer: B
Explanation: The notebook Section 5 explains that different metrics capture different aspects: Inception Score assesses both quality and diversity but doesn't compare to real data. FID compares statistical properties of generated vs. real distributions in feature space. Precision measures what fraction of generated samples are realistic (quality), while Recall measures what fraction of the real distribution is covered (diversity/coverage). Using all three provides comprehensive evaluation. Options A, C, and E incorrectly suggest redundancy or incompatibility. Option D misunderstands Precision/Recall in the generative context.
Reference: Notebook Section 5 "Evaluation Metrics for GANs" - Subsections 5.1 (IS), 5.2 (FID), and 5.3 (Precision and Recall) with explanations of each metric's strengths and limitations

5 References

All questions reference the Jupyter Notebook: "Generative Adversarial Networks (GANs) for Data Augmentation" with specific sections noted in each explanation.