Week-3 Questions

Prof . Prathosh AP and Chandan J

May 2025

1. Given a fixed generator G, what is the expression for the optimal discriminator $D^*(x)$?

   (A) $D^*(x) = \frac{p_{gen}(x)}{p_{\text{data}}(x)}$

   (B) $D^*(x) = \frac{p_{\text{data}}(x)}{p_{gen}(x)}$

   (C) $D^*(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x)+p_{gen}(x)}$

   (D) $D^*(x) = \log \frac{p_{\text{data}}(x)}{p_{gen}(x)}$

   Correct Answer: (C)

2. Does the failure of Critic network to discriminate between data guarantee the matching of those data distribution?

   (A) Yes

   (B) No

   Correct Answer: (B)

3. You are given a DC-GAN generator with the following architecture:

   - Linear layer: $100 \rightarrow 512 \times 4 \times 4$
   - Reshape to $(512, 4, 4)$
   - ConvTranspose2d$(512, 256, 4, 2, 1)$
   - ConvTranspose2d$(256, 128, 4, 2, 1)$
   - ConvTranspose2d$(128, 64, 4, 2, 1)$
   - ConvTranspose2d$(64, 3, 4, 2, 1)$

   What is the spatial resolution (height $\times$ width) of the final output image?

   (A) 32×32

   (B) 64×64

   (C) 128×128

(D) 256×256

Correct Answer: (B) 64×64

4. Let the generator $G_\theta : \mathbb{R}^n \to \mathbb{R}^{3\times 64\times 64}$ and discriminator $D_\phi : \mathbb{R}^{3\times 64\times 64} \to [0, 1]$. Assume latent noise $z \sim p_z(z) = \mathcal{N}(0, I)$ and real data from p_{data}.

   The Jensen-Shannon divergence is defined as:

$$\text{JSD}(p_{\text{data}} \parallel p_g) = \frac{1}{2}\text{KL}(p_{\text{data}} \parallel m) + \frac{1}{2}\text{KL}(p_g \parallel m), \quad m = \frac{1}{2}(p_{\text{data}} + p_g)$$

   Which of the following statements is **TRUE**?

   (A) The standard GAN discriminator directly estimates $\text{JSD}(p_{\text{data}} \parallel p_g)$

   (B) The optimal discriminator maximizes the Wasserstein distance between p_{data} and p_g

   (C) The training objective of DC-GAN approximates $\min_G \text{JSD}(p_{\text{data}} \parallel p_g)$

   (D) The generator minimizes $\text{KL}(p_{\text{data}} \parallel p_g)$

   Correct Answer: (C) The training objective of DC-GAN approximates $\min_G \text{JSD}(p_{\text{data}} \parallel p_g)$

5. In a conditional GAN, both the generator $G(z, y)$ and discriminator $D(x, y)$ are conditioned on label y. Assume $y \in \{1, \ldots, K\}$ is a one-hot encoded class label.

   The discriminator objective is:

$$\max_D \; \mathbb{E}_{(x,y)\sim p_{\text{data}}}[\log D(x, y)] + \mathbb{E}_{z\sim p_z(z), y\sim p_y(y)}[\log(1 - D(G(z, y), y))]$$

   Assuming the optimal discriminator is given by:

$$D^*(x, y) = \frac{p_{\text{data}}(x|y)}{p_{\text{data}}(x|y) + p_g(x|y)}$$

   Which of the following is the generator's optimal objective?

   (A) $\min_G \; \text{KL}(p_g(x|y) \parallel p_{\text{data}}(x|y))$

   (B) $\min_G \; \text{KL}(p_{\text{data}}(x|y) \parallel p_g(x|y))$

   (C) $\min_G \; \text{JSD}(p_{\text{data}}(x|y) \parallel p_g(x|y))$

(D) $\min_G \ \mathbb{E}_{y \sim p_y}[\mathrm{KL}(p_g(x|y) \parallel p_{\text{data}}(x|y))]$

Correct Answer: (C)

6. In a standard PyTorch implementation of a Vanilla GAN using ‘nn.BCELoss()‘ and ‘sigmoid‘ in the discriminator output, which of the following conditions may cause **vanishing gradients** for the generator?

   (A) The discriminator is too weak and always outputs 0.5

   (B) The discriminator is too strong and outputs values near 0 for fake samples

   (C) The generator uses ‘tanh‘ as the final activation

   (D) The noise vector is sampled from a uniform distribution instead of a Gaussian

   Correct Answer: (B)

7. In a PyTorch conditional GAN generator, the label y is one-hot encoded as a tensor of shape $[k]$, and the noise vector z has shape $[n]$. If they are concatenated to form the generator input, what must be done in PyTorch to ensure this works during batch training?

   (A) Use ‘torch.cat([z, y], dim=1)‘ without reshaping

   (B) Use ‘torch.cat([z.unsqueeze(1), y.unsqueeze(1)], dim=1)‘

   (C) Ensure both ‘z‘ and ‘y‘ are shaped ‘[batch_size, *]‘ before concatenation along dim=1

   (D) Broadcast ‘y‘ to match ‘z‘’s shape

   Correct Answer: (C)

8. Which of the following practices are generally adopted to improve the stability and performance of training GANs?

   (A) Using batch normalization in both the generator and discriminator

   (B) Using sigmoid activation at the output of the generator

   (C) Updating the generator more frequently than the discriminator

   Choose the best combination:

   (I) Only (A) and (C)

   (II) Only (A), (B), and (C)

   (III) Only (B) and (C)

   (IV) Only (A)

   Correct Answer: (I) Only (A) and (C)

9. Let $G(z, y)$ be a conditional generator where $z \sim \mathcal{N}(0, I)$ and y is one-hot class label. Suppose the generator is trained with mismatched (z, y) pairs due to a data loading bug (labels are randomly shuffled across samples).

   Which of the following behaviors is most likely?

   (A) The generator collapses to a single class output

   (B) The generator becomes non-convergent due to conflicting gradients

   (C) The generator perfectly models the joint distribution $p(x, y)$

   (D) The discriminator compensates and restores conditional mapping

   Correct Answer: (B)

10. A conditional GAN generator can be conditioned on class label y using:

    (a) One-hot concatenation: $[z, y]$

    (b) Learned embedding: $z' = z + E(y)$

    Which of the following is an advantage of using learned embeddings over naive one-hot conditioning?

    (A) Embeddings remove the need for explicit conditioning

    (B) Embeddings allow the generator to learn soft semantic relationships between classes

    (C) Embeddings reduce overfitting by collapsing similar class labels

    (D) Embeddings eliminate mode collapse in the generator

    Correct Answer: (B)