Quiz 1

Prof . Prathosh AP and Chandan J

May 2025

1. Suppose in GAN, the supports of p_{data} and p_g are disjoint. What is the value of the GAN value function?

   (a) 1
   (b) $-\infty$
   (c) 0
   (d) $\log 2$

   Answer: (c)

2. Consider in GAN, a simple setting where $p_{\text{data}} = \delta(x-1)$ and $p_g = \delta(x+1)$, i.e., both are Dirac delta functions located at different points. What is the GAN value function $V(G)$?

   (a) $-\log 2$
   (b) 0
   (c) $-\log(0.5)$
   (d) $\log 4$

   Answer: (b).

3. Consider the following GAN training loop in PyTorch:

```
for _ in range(k):
    optimizer_D.zero_grad()
    loss_D.backward()
    optimizer_D.step()
optimizer_G.zero_grad()
loss_G.backward()
optimizer_G.step()
```

   What does this loop imply?

(a) Generator is updated multiple times per discriminator update.

(b) Discriminator is updated multiple times per generator update.

(c) Both networks are updated simultaneously.

(d) The model uses gradient accumulation.

Answer: (b).

4. Suppose you are training a GAN using `torch.nn.BCELoss()`. During training, the discriminator loss becomes nearly zero and the generator loss increases rapidly. What is the most likely explanation?

   (a) The generator is overfitting to the noise.

   (b) The discriminator has become too strong.

   (c) The learning rate of the generator is too high.

   (d) The noise vector dimension is too large.

   Answer: (b)

5. In Bi-GAN, which of the following expressions represents the correct **joint objective** to be minimized by the encoder E and generator G, and maximized by the discriminator D?

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

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

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

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

   Answer: (a)

6. Assume that the encoder E in a Bi-GAN maps $x \in \mathbb{R}^{784}$ to $z \in \mathbb{R}^{100}$, and G maps z to x. What is a necessary condition for E and G to become **inverses of each other**?

   (a) D must be able to distinguish between fake and real with high confidence.

(b) The composite mapping $G(E(x)) \approx x$ and $E(G(z)) \approx z$ must hold for all x and z.

(c) The generator and encoder must share weights.

(d) The discriminator must estimate pixel-wise binary mask loss.

Answer: (b)

7. A GAN's generator $G(z)$ takes in a latent vector $z \sim \mathcal{N}(0, I)$ of size 100, and produces images of shape 28×28. If the generator consists of three linear layers: $100 \to 128 \to 256 \to 784$, how many trainable parameters are in the generator (excluding biases)?

   (a) 78,400
   (b) 102,400
   (c) 246,272
   (d) 156,000

   Answer: (c)

8. During Bi-GAN training, suppose you compute discriminator output on:

$$D(x, E(x)) = 0.85, \quad D(G(z), z) = 0.15$$

   If using Binary Cross Entropy loss with target labels 1 for real and 0 for fake, what is the total discriminator loss?

   (a) 0.324
   (b) 0.45
   (c) 0.23
   (d) 0.85

   Answer: (a)

9. In a Bi-GAN, you use an encoder E, generator G, and discriminator D. Given batch size 64, and each sample is 784-dimensional, what is the shape of the input to the discriminator?

   (a) $[64, 784]$
   (b) $[64, 100]$
   (c) $[64, 884]$
   (d) $[128, 784]$

   Answer: (c)

10. Let p_r and p_g be two probability distributions supported on a compact metric space $\mathcal{X}$. The Wasserstein-1 distance is defined as:

$$W(p_r, p_g) = \inf_{\gamma \in \Pi(p_r, p_g)} \mathbb{E}_{(x,y)\sim\gamma}[\|x - y\|_2]$$

Which of the following properties is guaranteed for W, unlike the Jensen-Shannon divergence?

(a) It is finite and continuous even when p_r and p_g have disjoint supports.

(b) It is always zero for disjoint supports.

(c) It upper bounds the KL divergence.

(d) It requires the same dimensionality of supports.

Answer: (a)

11. In W-GAN, enforcing the 1-Lipschitz condition on the critic (discriminator) is crucial. Which of the following methods **fails** to properly enforce the Lipschitz constraint?

(a) Weight clipping, lead to capacity underuse and gradient vanishing.

(b) Gradient penalty via $\lambda \mathbb{E}_{\hat{x}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]$.

(c) Spectral normalization to bound each layer's Lipschitz constant.

(d) Enforcing $\|f(x_1) - f(x_2)\| \leq \|x_1 - x_2\|$ directly via constraint optimization.

Answer: (a)

12. In the PyTorch implementation of W-GAN, which of the following code snippets correctly enforces the weight clipping constraint on the critic?

(a)
```
for p in D.parameters():
        p = torch.clamp(p, -0.01, 0.01)
```

(b)
```
for p in D.parameters():
        p.data.clamp_(-0.01, 0.01)
```

(c) `torch.nn.utils.clip_grad_norm_(D.parameters(), 0.01)`

(d) `D.clip(-0.01, 0.01)`

Answer: (b)

13. Consider the following marginal distributions:

$$P = [0.25, 0.25, 0.5], \quad Q = [0.5, 0.3, 0.2]$$

Which of the following transport plans T is valid (i.e., has row sums $= P$ and column sums $= Q$)?

(a)
$$\begin{bmatrix} 0.25 & 0 & 0 \\ 0.25 & 0 & 0 \\ 0 & 0.3 & 0.2 \end{bmatrix}$$

(b)
$$\begin{bmatrix} 0.2 & 0.05 & 0 \\ 0.3 & 0 & 0 \\ 0 & 0.25 & 0.25 \end{bmatrix}$$

(c)
$$\begin{bmatrix} 0.1 & 0.1 & 0.05 \\ 0.2 & 0.2 & 0.1 \\ 0.2 & 0 & 0.1 \end{bmatrix}$$

(d)
$$\begin{bmatrix} 0.3 & 0 & 0 \\ 0.2 & 0.2 & 0.2 \\ 0 & 0.1 & 0.3 \end{bmatrix}$$

Answer: (b)

14. Let $P = [0.2, 0.3, 0.5]$ and $Q = [0.3, 0.3, 0.4]$. Consider the following transport plan matrix T:

$$T = \begin{bmatrix} 0.2 & 0.0 & 0.0 \\ 0.1 & 0.2 & 0.0 \\ 0.0 & 0.1 & 0.4 \end{bmatrix}$$

and cost matrix

$$D = \begin{bmatrix} 0 & 1 & 3 \\ 1 & 0 & 2 \\ 3 & 2 & 0 \end{bmatrix}$$

Compute the total cost (EMD).

(a) 1.1
(b) 1.5
(c) 0.8
(d) 0.9

Answer: (d)

15. Given two discrete distributions:

$$P = [0.5, 0.5], \quad Q = [0.25, 0.75]$$

and an f-divergence defined using $f(u) = u \log u$ (i.e., Kullback-Leibler divergence), compute $D_f(P\|Q)$.

(a) 0.056
(b) 0.22
(c) 0.31
(d) 0.13

Solution:

$$D_{\mathrm{KL}}(P\|Q) = 0.5 \log \frac{0.5}{0.25} + 0.5 \log \frac{0.5}{0.75} = 0.5(1) + 0.5(-0.5849) \approx 0.13$$

Answer: (d)

16. Let $P = [0.1, 0.2, 0.7]$ and $Q = [0.2, 0.3, 0.5]$. Approximate the total variation distance $D_{TV}(P, Q) = \frac{1}{2} \sum_i |p_i - q_i|$.

(a) 0.05
(b) 0.2
(c) 0.3
(d) 0.5

Solution:

$$D_{TV} = \frac{1}{2}(|0.1 - 0.2| + |0.2 - 0.3| + |0.7 - 0.5|) = \frac{1}{2}(0.1 + 0.1 + 0.2) = 0.2$$

Answer: (b)

17. If a domain classifier achieves 100% accuracy in distinguishing between source and target domain samples, what can we say about domain adaptation?

(a) The domains are perfectly aligned
(b) The classifier is overfitting
(c) The domains are misaligned
(d) Nothing can be inferred

Answer: (c)

18. In adversarial domain adaptation, a gradient reversal layer is used with a coefficient $\lambda = 0.5$. If the original gradient is $\nabla = [2, -4]$, what is the reversed gradient used to update the feature extractor?

(a) $[-1, 2]$
(b) $[1, -2]$
(c) $[-2, 4]$
(d) $[0, 0]$

Solution:
$$\nabla_{\text{reversed}} = -\lambda\nabla = -0.5 \cdot [2, -4] = [-1, 2]$$

Answer: (a)

19. Given two multivariate Gaussians with means μ_r, μ_g and covariances Σ_r, Σ_g, the Frechet Inception Distance (FID) is defined as:
$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}\right)$$
Suppose:
$$\mu_r = [1, 1], \quad \mu_g = [2, 3], \quad \Sigma_r = I, \quad \Sigma_g = 4I$$
What is the FID?

(a) 1
(b) 10
(c) 8
(d) 6

Solution: $\|\mu_r - \mu_g\|^2 = (1-2)^2 + (1-3)^2 = 1 + 4 = 5$ $\text{Tr}(I + 4I - 2\sqrt{I \cdot 4I}) = \text{Tr}(5I - 2 \cdot 2I) = \text{Tr}(5I - 4I) = \text{Tr}(I) = 1 \Rightarrow \text{FID} = 5 + 1 = 6$

Answer: (d)

20. If the Inception network produces activations with mean and covariance for real data as μ_r, Σ_r and for generated data as μ_g, Σ_g, which of the following is true?

(a) If $\mu_r = \mu_g$ and $\Sigma_r = \Sigma_g$, FID is negative.
(b) FID is non-negative and equals 0 when distributions are identical.
(c) FID can be undefined if Σ_r is not positive definite.
(d) FID equals the KL divergence between the distributions.

Answer: (b)