DS552/CS552- Generative AI - Assignment 4

Theory Questions

Q1. Explain the minimax loss function in GANs and how it ensures competitive training between the generator and discriminator.

The minimax loss function in GANs is designed to establish a competitive framework between the generator (G) and the discriminator (D). Mathematically, it is represented as:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[1 - \log D\big(G(z)\big)]$$

- The discriminator (D) tries to maximize this function by correctly distinguishing between real data (x) and fake data (G(z)). It aims to assign high probabilities to real data and low probabilities to generated data.
- The generator (G) tries to minimize this function by generating samples that fool the discriminator, making it classify generated data as real.

This adversarial process drives both networks to improve:

- If D becomes too strong, G fails to generate convincing samples.
- If G dominates, D fails to differentiate real from fake data.
- Ideally, they reach an equilibrium where G(z) is indistinguishable from real data.

Q2. What is mode collapse, why can mode collapse occur during GAN training? And how can it be mitigated?

Mode collapse occurs when the generator produces only a limited variety of outputs instead of diverse samples. This means the generator finds a small set of outputs that consistently fool the discriminator and fails to explore the full data distribution.

Why Does Mode Collapse Occur?

The generator finds a shortcut by producing a few specific outputs that the discriminator struggles to classify as fake. The discriminator then learns to identify these outputs, leading the generator to shift to another set of outputs, creating instability. This cycle results in a lack of diversity in generated samples.

How to Mitigate Mode Collapse?

- Minibatch Discrimination: Encourages diversity by allowing the discriminator to compare samples within a batch.
- Wasserstein GAN (WGAN): Uses a different loss function (Wasserstein distance) to improve stability and diversity.
- Feature Matching: Instead of directly fooling the discriminator, the generator tries to match statistical features of real data.
- Unrolled GANs: Allows the generator to take into account future updates of the discriminator, preventing it from sticking to a narrow set of outputs.

Q3. Explain the role of the discriminator in adversarial training?

The discriminator (D) is a binary classifier that distinguishes between real and fake samples. It plays a crucial role in GAN training by providing feedback to the generator about the quality of generated samples. If the discriminator becomes too strong, the generator receives minimal gradient updates and struggles to improve. On the other hand, if the discriminator is too weak, the generator's outputs remain unrealistic. Thus, maintaining balance is key to stable training.

Key Responsibilities:

1. Classifies Real vs. Fake Data: D(x) should be close to 1 for real samples and close to 0 for fake samples.
2. Guides the Generator's Learning: By minimizing the discriminator's success rate, the generator improves.
3. Prevents Overfitting: A well-trained discriminator ensures that the generator doesn't simply memorize and replicate real data.

Q4. How do metrics like IS and FID evaluate GAN performance?

GANs do not have a straightforward evaluation metric like accuracy in supervised learning. Two common metrics are:

1. **Inception Score (IS)**

$$IS = \exp\left(\mathbb{E}_x D_{KL}(p(y|x)||p(y))\right)$$

   - Measures image quality and diversity.
   - Uses a pretrained Inception model to classify generated images.
   - A high IS means that the generated images belong to distinct classes and are easily recognizable.
   - Does not compare with real images; only measures diversity and recognizability.

2. **Fréchet Inception Distance (FID)**

$$FID = ||\mu_r - \mu_g||^2 + Tr(\Sigma_r - 2\sqrt{\Sigma_r \Sigma_g} + \Sigma_g)$$

   - Compares statistics of real and generated images using the Inception model's feature space.
   - Lower FID means the generated images are closer to real data.
   - More robust than IS since it accounts for both diversity and realism.