DS552 – Generative AI

Assignment 4 – Generative Adversarial Networks (GANs)

smadhavankutty@wpi.edu
2/9/2025

1. Theory Questions:

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

Ans: The minimax loss function in Generative Adversarial Networks (GANs) is a mathematical formulation that captures the adversarial relationship between the generator and the discriminator. It ensures competitive training by framing the interaction between the two networks as a two-player minimax game. Here's a detailed explanation:

The minimax loss function for GANs is given by:-

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

Where:

- x is a real data sample from the actual dataset.
- z is a random noise vector input to the generator.
- $G(z)$ is the fake data generated by the generator.
- $D(x)$ is the discriminator's output, representing the probability that x is real.
- $D(G(z))$ is the discriminator's output, representing the probability that $G(z)$ is real.
- $\mathbb{E}$ denotes the expectation over the distribution of real data $p_{\text{data}}(x)$ and noise $p_z(z)$.

The training process involves two main steps:-

Discriminator Training: The discriminator is trained to maximize the probability of correctly classifying real and fake data. In this case, the discriminator aims to maximize the log probability of real data being classified as real and generated data being classified as fake.

Generator Training: The generator is trained to minimize the probability of the discriminator correctly classifying its outputs as fake. This formulation focuses on making the discriminator believe that the generated samples are real.

The minimax game ensures competitive training because of following reasons:

- ✓ The discriminator becomes better at distinguishing real from fake samples, challenging the generator to produce more realistic samples.
- ✓ The generator improves its ability to produce data that can fool the discriminator, forcing the discriminator to become more accurate.

In summary, the minimax loss function in GANs encapsulates the adversarial nature of the training process. By framing the interaction as a minimax game, it ensures that both the generator and discriminator are continuously improving in a competitive manner. This dynamic leads to the generator producing increasingly realistic data, ultimately achieving the goal of generating data that is nearly indistinguishable from real data.

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

Ans: Mode collapse is a common problem that can occur during the training of Generative Adversarial Networks (GANs). It refers to a situation where the generator network produces a limited variety of outputs, effectively collapsing to a single or few modes of the data distribution. This means that instead of generating diverse outputs that cover the entire data distribution, the generator produces outputs that are very similar or identical, thereby failing to capture the full diversity of the target data distribution.

Mode collapse can occur due to several reasons related to the adversarial nature of GAN training:

<u>1. Imbalance Between Generator and Discriminator:</u>

If the discriminator becomes too strong too quickly, it can easily distinguish between real and fake samples. This can cause the generator to find a few specific samples that can fool the discriminator and stick to producing those, leading to a lack of diversity.

<u>2. Loss Function Limitations:</u>

The traditional GAN loss function may not adequately penalize the generator for lack of diversity. The generator might find it easier to minimize the loss by producing a limited set of samples rather than exploring the entire data distribution.

<u>3.Gradient Saturation:</u>

If the discriminator's gradients become saturated, the generator may not receive useful gradient updates, leading it to produce similar outputs.

<u>4. Optimization Challenges:</u>

The training process of GANs involves a delicate balance between the generator and discriminator. If the generator finds a particular mode that consistently fools the discriminator, it may stop exploring other modes, leading to mode collapse.

Strategies to mitigate mode collapse:

1. Employing strategies such as training the generator more frequently than the discriminator or using historical averaging can help in balancing the training dynamics.
2. Use <u>Mini batch Discrimination</u> technique that allows the discriminator to look at multiple samples in a mini-batch and compare them. By doing so, the discriminator can detect if the generator is producing similar samples, encouraging the generator to produce more diverse outputs.
3. Using <u>improved Loss Functions</u> such as Wasserstein loss (WGAN) or Least Squares GAN (LSGAN) can provide more stable training dynamics and reduce mode collapse.
4. Using <u>Regularization Techniques</u> like adding noise to the inputs of the discriminator or using dropout can help prevent the discriminator from becoming too confident and overpowering the generator.
5. Using <u>Feature matching</u> that involves adding an additional term to the generator's loss function that encourages the generated samples to match the statistics of real samples in feature space. This helps the generator to produce samples that are more diverse and realistic.

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

Ans: In the context of Generative Adversarial Networks (GANs), the discriminator plays a crucial role in the learning process. Some key roles as follows:-

1. The primary function of the discriminator is to distinguish between real data samples (from the actual dataset) and fake data samples (generated by the generator network). It acts as a classifier that evaluates the authenticity of the data it receives.
2. During training, the discriminator is trained on a mix of real and fake data. It learns to improve its classification accuracy by minimizing the error in distinguishing between the two types of data. This process involves adjusting the weights of the discriminator network through backpropagation based on the classification error.
3. The discriminator's objective is adversarial in nature because it is in competition with the generator. While the generator tries to create realistic data to fool the discriminator, the discriminator tries to improve its ability to detect fake data.
4. The discriminator provides feedback to the generator through its loss function. The generator uses this feedback to improve its ability to produce more realistic data, thereby making the task of the discriminator more challenging.

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

Ans:

Inception Score (IS)

- ✓ IS valuates the quality and diversity of images generated by GANs.
- ✓ Higher IS score indicates better quality and diversity of the generated images.
- ✓ IS limitation is that it doesn't capture the realism of images and being sensitive to the choice of the Inception model.
- ✓ IS Mathematical Calculation

$$\mathrm{IS} = \exp\left(\mathbb{E}_{x\sim p_g}\left[\mathrm{KL}(p(y|x) \parallel p(y))\right]\right)$$

Where:

- $p(y|x)$ is the conditional class distribution for a generated image x.
- $p(y)$ is the marginal class distribution across all generated images.

For each generated image, the Inception model outputs a probability distribution over the ImageNet classes.

Compute the marginal distribution of class labels by averaging the class probabilities across all generated images.

Calculate the Kullback-Leibler (KL) divergence between the conditional class distribution (for each image) and the marginal class distribution.

The IS is the exponential of the expected KL divergence across all images.

Fréchet Inception Distance (FID)

- ✓ FID measures how similar the statistics of generated images are to real images, capturing both quality and diversity.
- ✓ A lower FID indicates that the generated images are more similar to the real images in terms of feature statistics, implying better quality and realism.
- ✓ FID is considered more robust than IS because it captures both the quality and diversity of images and is less sensitive to the choice of the Inception model.
- ✓ FID Mathematical Calculation

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2})$$

Where:

- μ_r and μ_g are the mean feature vectors of real and generated images, respectively.
- Σ_r and Σ_g are the covariance matrices of the feature vectors of real and generated images, respectively.
- Tr denotes the trace of a matrix.

Pass both real and generated images through the Inception model to extract features from an intermediate layer (typically the pool3 layer).

Compute the mean and covariance matrix of the feature vectors for both real and generated images.

Calculate the Fréchet distance (also known as the Wasserstein-2 distance) between the two multivariate Gaussian distributions defined by the mean and covariance matrices.

Both metrics (IS and FID) provide valuable insights into the performance of GANs, with FID generally being preferred for its robustness and ability to capture image realism more effectively.

2. Coding Tasks:

GitHub Repo Link : https://github.com/smadhavanwpi/DS552_GenAI

Python Source File : GenAI_Assignment4.ipynb

Python Notebook : genai_assignment4.py

Generator & Discriminator Loss over time (50 Epochs) : 332FDD36-BEA6-4EC1-8905-94C773ACA5BD.jpeg