Week 0e: Generative AI
The Creation Challenge

Machine Learning for Smarter Innovation

BSc-Level Course

October 6, 2025

Outline

Traditional ML: "What is this?"

- Email: spam or not?
- Image: cat or dog?
- Text: positive sentiment?

Limitation: Analysis only

Generative AI: "Create something new"

- Generate realistic images
- Write coherent articles
- Compose music

Power: Creation & innovation

Fundamental shift: from understanding to creating

Discriminative Models
Learn: $P(y|x)$
Examples:

- Logistic regression
- Random Forest, SVM

Goal: Decision boundaries

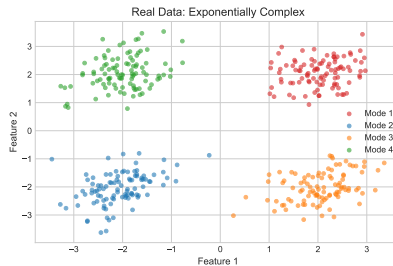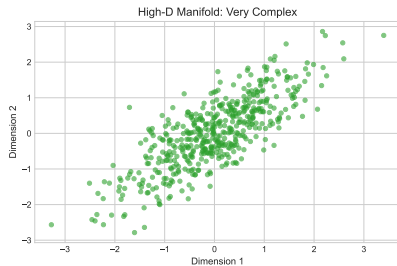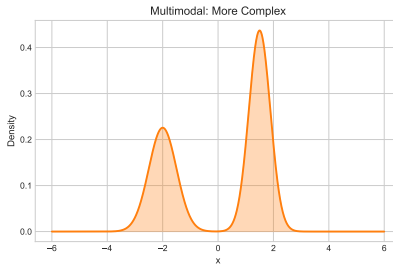**Generative Models**
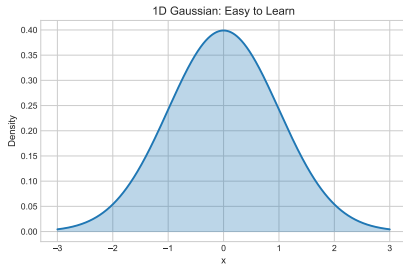Learn: $P(x)$ or $P(x, y)$
Examples:

- VAEs, GANs
- Diffusion models

Goal: Data generation
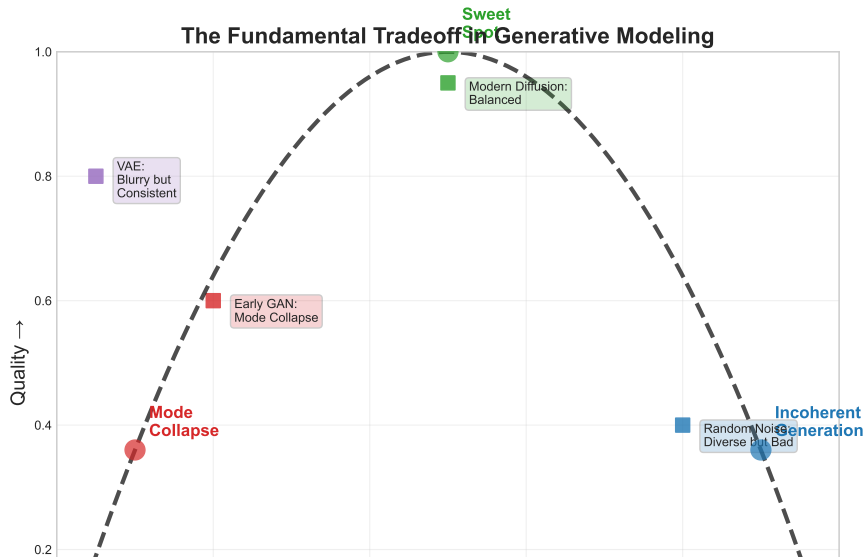
Discriminative: "Is this X?" — Generative: "Create X"

The Hard Problem

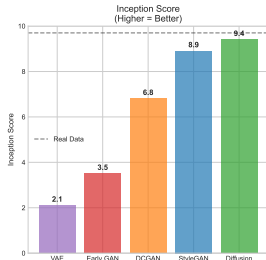Why Generation is Fundamentally Difficult

The Fundamental Tradeoff in Generative Modeling

Plot labels:
- **Sweet Spot** (green)
- Modern Diffusion: Balanced (green square)
- VAE: Blurry but Consistent (purple square)
- Early GAN: Mode Collapse (red square)
- **Mode Collapse** (red)
- Random Noise: Diverse but Bad (blue square)
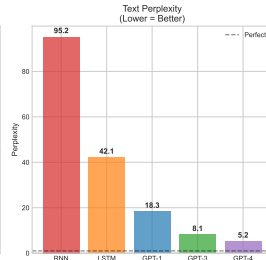- **Incoherent Generation** (blue)

Y-axis: Quality →

Inception Score
- Higher = better
- Quality & diversity

FID Score
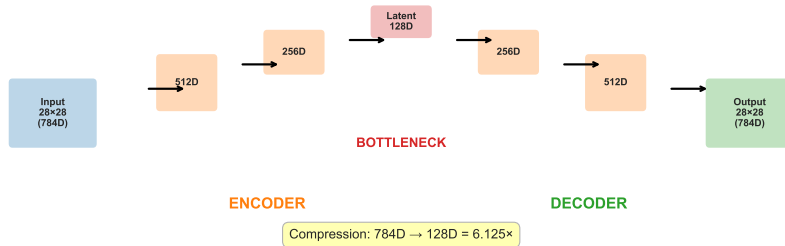- Lower = better
- Feature distance

Perplexity (Text)
- Lower = better
- Language fluency

IS¿300 excellent, FID¡10 photorealistic, PPL¡20 human-like

Autoencoder Architecture: Compression Through Reconstruction

| Input 28×28 (784D) | 512D | 256D | Latent 128D | 256D | 512D | Output 28×28 (784D) |

BOTTLENECK

ENCODER　　　　　**DECODER**

Compression: 784D → 128D = 6.125×

Encoder

- 784D -¿ 128D
- $z = f_{enc}(x)$

Force information through bottleneck, learn to reconstruct

Latent Space

- Bottleneck: 128D
- Key features

Decoder
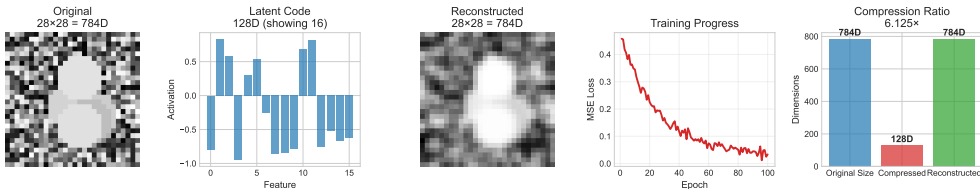
- 128D -¿ 784D
- $\hat{x} = f_{dec}(z)$

Architecture Details:

- Input: 28×28 = 784 pixels
- Encoder: 784 -¿ 512 -¿ 256 -¿ 128
- Decoder: 128 -¿ 256 -¿ 512 -¿ 784
- Activation: ReLU (hidden), Sigmoid (output)

Training Process:

- Loss: $L = ||x - \hat{x}||^2$
- Optimizer: Adam, lr=0.001
- Epochs: 100
- Compression ratio: $784/128 = 6.125x$

Reconstruction loss: MSE drops from 0.45 to 0.03 over 100 epochs

Autoencoder Successes

Visualization Placeholder

(Chart 12)

[+] SUCCESSES:
- Dimensionality reduction: 784D -¿ 128D

Quantitative Results:
- MSE: 0.031, Compression: 6.125x

Autoencoder Failures

Visualization Placeholder

(Chart 13)

Generation Metrics:

Metric	Score
IS	2.1

Averaging Problem

Visualization Placeholder

(Chart 14)

The Averaging Problem:
- Loss: $L = ||x - \hat{x}||^2$

Mathematical Insight:
- $\hat{x} = \arg\min E[||x - \hat{x}||^2]$

Vae Framework

Visualization Placeholder

(Chart 15)

Key Innovation:

- Encode to distribution, not point

VAE Loss:

$$\mathcal{L} = -E[\log p_\theta(x|z)] + KL(q||p)$$

Artist Learning Process

Visualization Placeholder

(Chart 16)

Traditional Art Education:

- Student creates artwork

Key Insights:

- Adversarial feedback drives improvement

Two Approaches

Visualization Placeholder

(Chart 17)

Approach 1: Adversarial
- Two networks compete

Approach 2: Diffusion
- Iterative denoising

Forger Detective Analogy

Visualization Placeholder

(Chart 18)

Forger (Generator):
- Creates fakes from noise

Detective (Discriminator):
- Examines: real or fake?

Reverse Corruption Analogy

Visualization Placeholder

(Chart 19)

Forward (Corruption):
- Clean image -¿ pure noise

Reverse (Generation):
- Pure noise -¿ clean image

Gan Geometric Dynamics

Visualization Placeholder

(Chart 20)

Generator:

- Maps noise z to data x

Discriminator:

- Separates real from fake

Gan Training Walkthrough

Visualization Placeholder

(Chart 21)

Epoch 1:
- D_loss: 1.386

Epoch 100:
- D_loss: 0.695

Diffusion Mathematics

Visualization Placeholder

(Chart 22)

Forward (Fixed):

$$q(x_t|x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

Noise schedule: β_t (0.0001 → 0.02)

Reverse (Learned):

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(\mu_\theta, \Sigma_\theta)$$

Network ϵ_θ predicts noise

Latent Interpolation

Visualization Placeholder

(Chart 23)

GAN Interpolation:

- Sample $z_1, z_2 \sim \mathcal{N}(0, I)$

Applications:

- Style transfer, face morphing

Denoising Steps

Visualization Placeholder

(Chart 24)

Key Time Steps:

- T=1000: Pure noise

Process Control:

- Guidance scale

Adversarial Theory

Visualization Placeholder

(Chart 25)

Theory:

- Minimax convergence

Benefits:

- Sharp, realistic images

Quality Metrics Over Time

Visualization Placeholder

(Chart 26)

GAN Progress:

- Start: IS=1.2, FID=450

Diffusion Progress:

- 100k: FID=200

Stable Diffusion Api

Visualization Placeholder

(Chart 27)

Basic Usage:

```
import requests

response = requests.post(
    api_url,
    headers={"Authorization": key},
```

Parameters:

- `cfg_scale`: Adherence (1-20)
- `steps`: Quality (10-150)
- `seed`: Reproducible

Cost: $0.004 per image

Generative Landscape

Visualization Placeholder

(Chart 28)

VAEs: Probabilistic, smooth latent, blurry
GANs: Adversarial, sharp outputs, unstable

Diffusion: Iterative denoising, high quality, slow
Transformers: Sequential, excellent text, scalable

Each approach has unique strengths - modern systems combine techniques

Generative Tradeoffs

Visualization Placeholder

(Chart 29)

Training Stability:

- VAEs, Diffusion: Stable
- GANs: Unstable

Quality:

- Diffusion, GANs: Excellent
- VAEs: Blurry

Modern Applications

Visualization Placeholder

(Chart 30)

Image Generation:

- DALL-E 3, Midjourney
- Stable Diffusion, Firefly

Text Generation:

- GPT-4, Claude, Gemini
- Llama 2 (open)

Ethics Summary

Visualization Placeholder

(Chart 31)

Capabilities:
- Realistic images from text
- Human-like writing

Challenges:
- Deepfakes, misinformation
- Copyright issues