Week 0e: Generative AI

The Creation Challenge

Machine Learning for Smarter Innovation

BSc-Level Course

September 28, 2025

Outline

Traditional ML:
"What is this?"

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

Limitation: Analysis only

Generative AI:
"Create something new"

- Generate: realistic images
- Write: coherent articles
- Compose: original music
- Design: novel molecules

Power: Creation & innovation

The fundamental shift: from understanding existing data to creating new possibilities

Discriminative Models

Learn: $P(y|x)$
"Given input x, what's the label y?"

Examples:

- Logistic regression
- Random Forest
- Neural networks (classification)
- SVM

Goal: Decision boundaries

Generative Models

Learn: $P(x)$ or $P(x, y)$
"What does the data distribution look like?"

Examples:

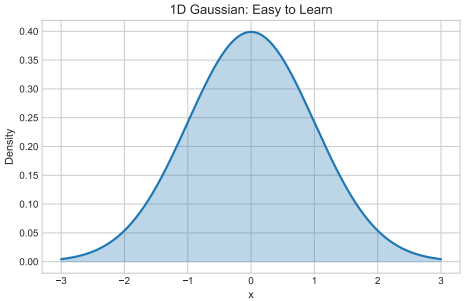
- Gaussian Mixture Models
- Variational Autoencoders
- GANs
- Diffusion models

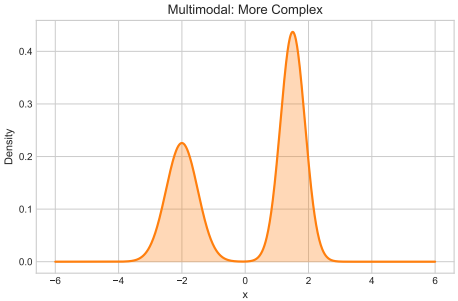
Goal: Data generation

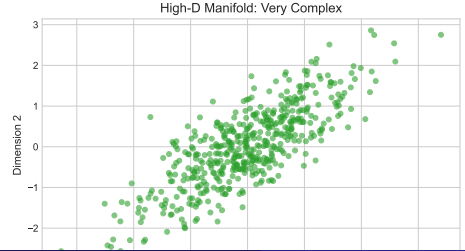
Discriminative: "Is this a cat?" — Generative: "Draw me a cat"

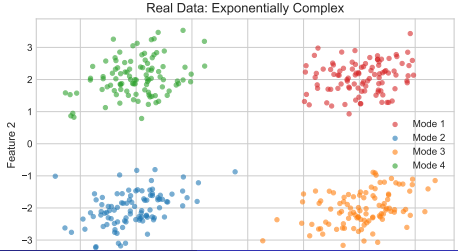
The Hard Problem

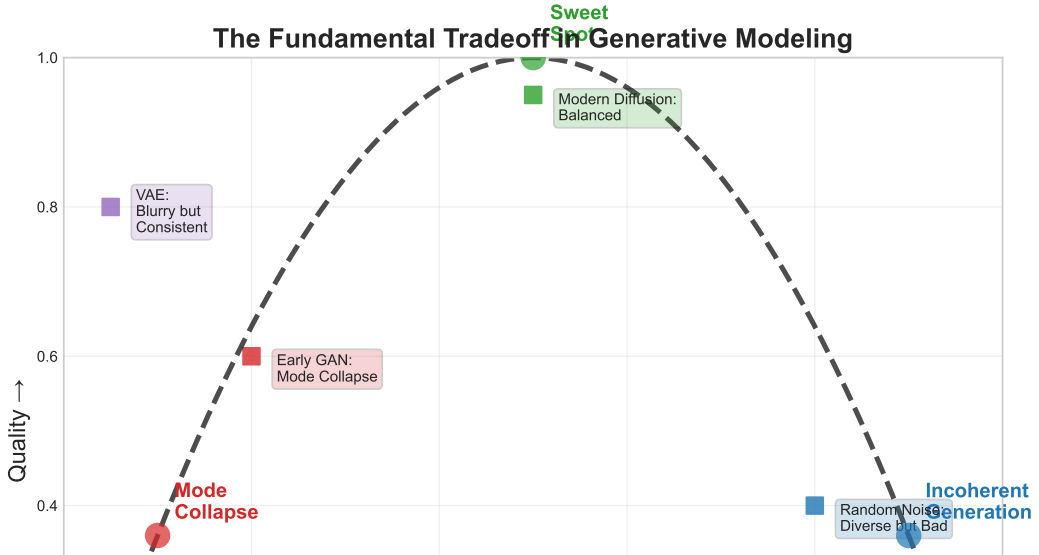
Why Generation is Fundamentally Difficult





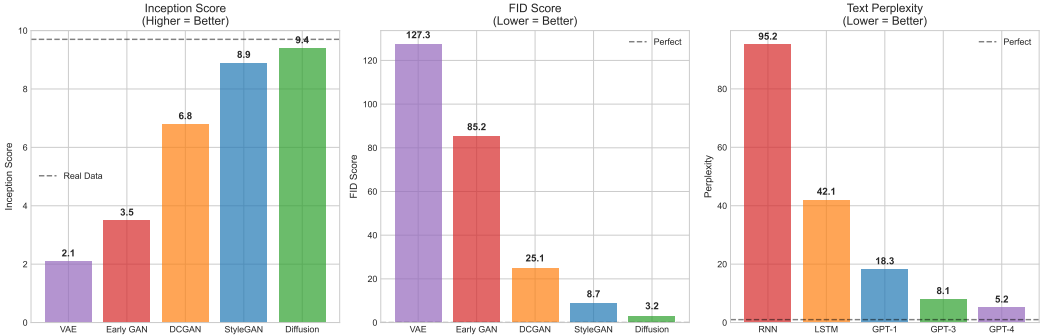





The Fundamental Tradeoff in Generative Modeling
Sweet Spot
Modern Diffusion: Balanced
VAE: Blurry but Consistent
Early GAN: Mode Collapse
Mode Collapse
Random Noise: Diverse but Bad
Incoherent Generation
Quality →
1.0
0.8
0.6
0.4

Measuring Generation Quality

Metrics for Evaluating Generative Models



Inception Score (IS)

- Range: 1-1000+
- Higher = better
- Quality & diversity
- $IS = \exp(E[KL(p(y|x)||p(y))])$

FID Score

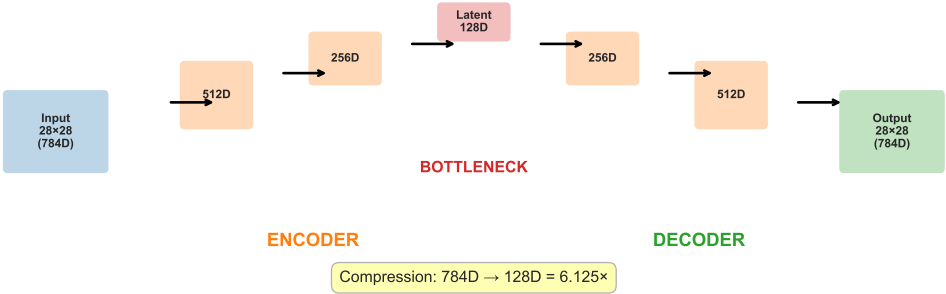
- Range: 0-500+
- Lower = better
- Feature distance
- Real vs generated

Perplexity (Text)

- Range: 1-10,000+
- Lower = better
- Predictability
- Language fluency

Quantitative evaluation: IS=300+ (excellent), FID¡10 (photorealistic), Perplexity¡20 (human-like text)

Autoencoder Architecture: Compression Through Reconstruction



Encoder

- Input: 784D (28x28 image)

Latent Space

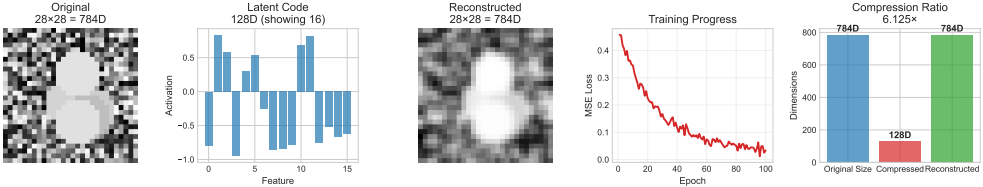
- Bottleneck: 128D

Decoder

- Latent: 128D

Worked Example: MNIST Compression

From 784 Pixels to 128 Features



Architecture Details:

- Input: $28 \times 28 = 784$ pixels
- Encoder: $784 \rightarrow 512 \rightarrow 256 \rightarrow 128$
- Decoder: $128 \rightarrow 256 \rightarrow 512 \rightarrow 784$
- Activation: ReLU (hidden), Sigmoid (output)

Training Process:

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

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

Autoencoder Successes

Visualization Placeholder

(Chart 12)

Autoencoder Failures

Visualization Placeholder

(Chart 13)

Averaging Problem

Visualization Placeholder

(Chart 14)

Vae Framework

Visualization Placeholder

(Chart 15)

Artist Learning Process

Visualization Placeholder

(Chart 16)

Two Approaches

Visualization Placeholder

(Chart 17)

Forger Detective Analogy

Visualization Placeholder

(Chart 18)

Reverse Corruption Analogy

Visualization Placeholder

(Chart 19)

Gan Geometric Dynamics

Visualization Placeholder

(Chart 20)

Gan Training Walkthrough

Visualization Placeholder

(Chart 21)

Diffusion Mathematics

Visualization Placeholder

(Chart 22)

Latent Space Interpolation

Smooth Transitions in Generated Content

Latent Interpolation

Visualization Placeholder

(Chart 23)

Diffusion Denoising Visualization

From Noise to Image in 1000 Steps

Denoising Steps

Visualization Placeholder

(Chart 24)

Adversarial Theory

Visualization Placeholder

(Chart 25)

Quality Metrics Over Time

Visualization Placeholder

(Chart 26)

Stable Diffusion Api

Visualization Placeholder

(Chart 27)

Generative Landscape

Visualization Placeholder

(Chart 28)

Generative Tradeoffs

Visualization Placeholder

(Chart 29)

Modern Applications

Visualization Placeholder

(Chart 30)

Ethics Summary

Visualization Placeholder

(Chart 31)