Variational Autoencoder

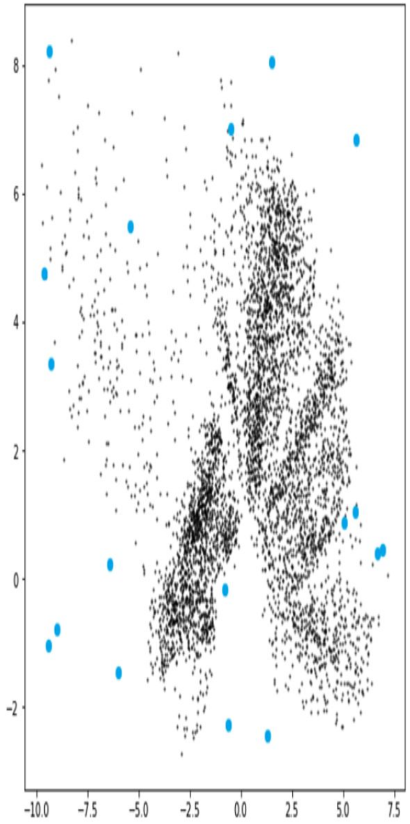
Problem with Standard Autoencoders

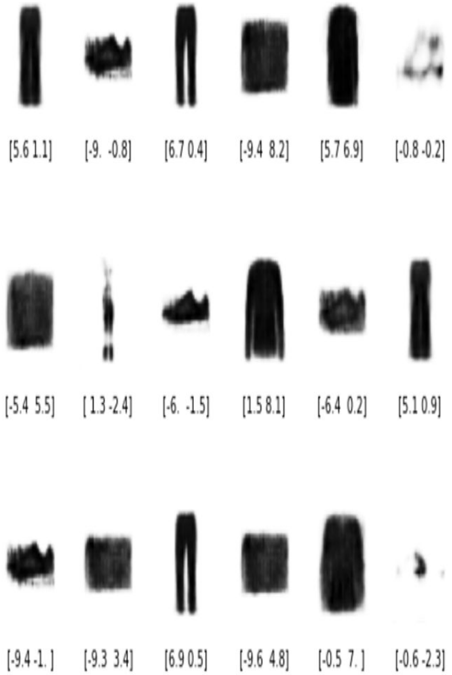
Every item placed at a precise point in wardrobe

Similar clothes often far apart

Random picks = weird, glitchy clothing

Standard Autoencoders have discontinuous latent spaces





The Magic Wardrobe: Intuition Behind Variational Autoencoders

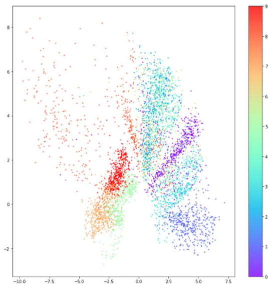
The VAE Makeover Begins...

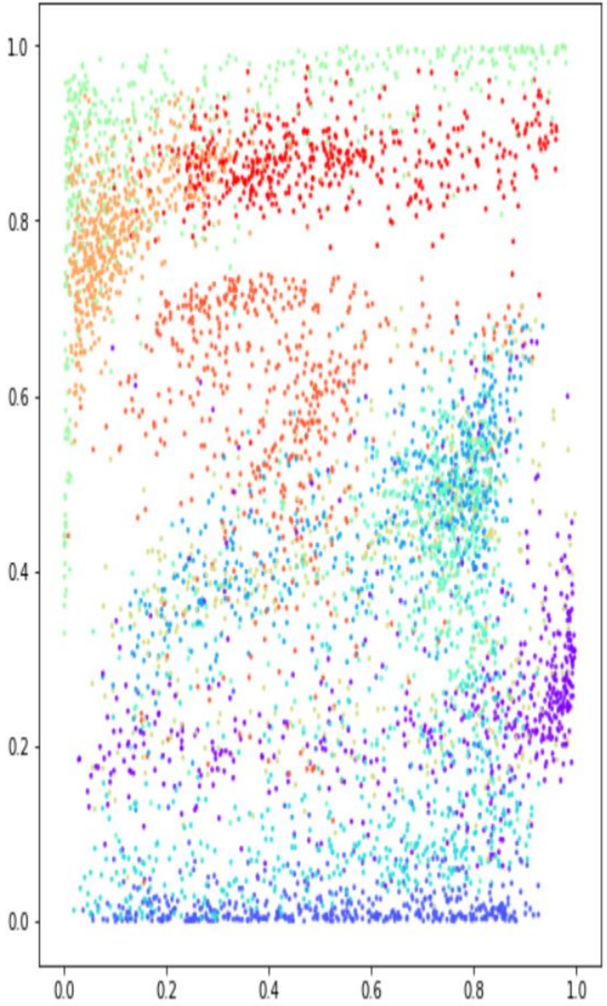
Change 1: Place items in zones, not fixed points

Change 2: Center each zone near the wardrobe's middle, with spread ~1m

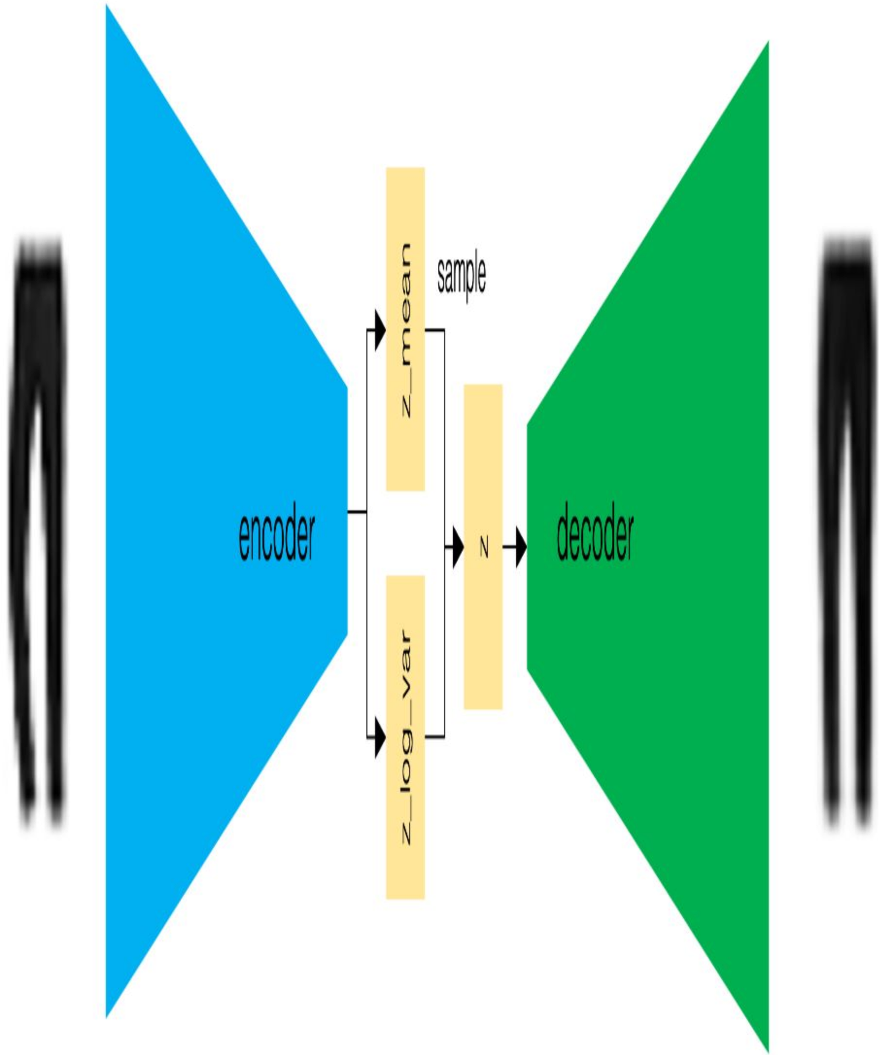
Penalty for violating rules (thanks, Brian!)

These rules = **structured latent space**



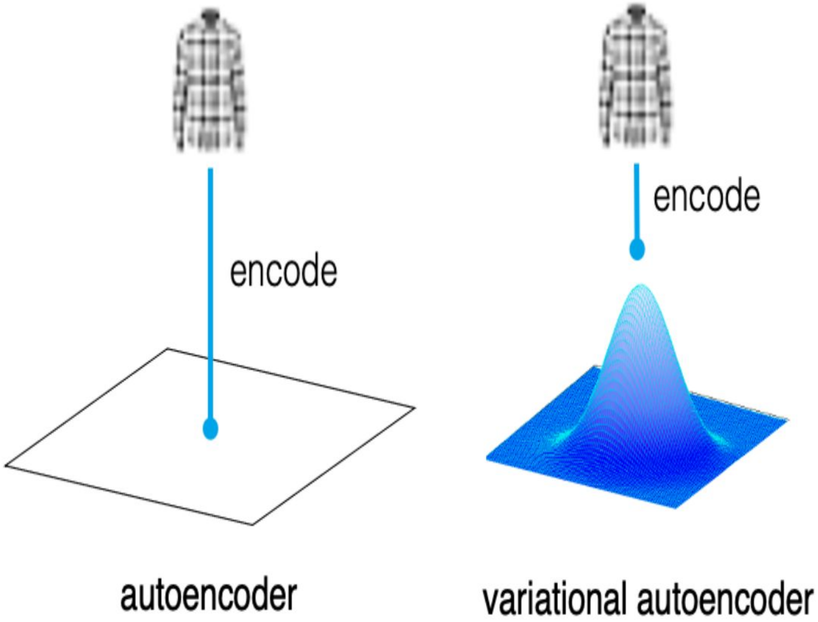


VAE Architecture



VAE Encoder

- Each input mapped to a **distribution** (**not a point**) in latent space
- Encoder outputs:
    - **z_mean**: center of the distribution
    - **z_log_var**: logarithm of variance (spread) per dimension
- Assumes dimensions are uncorrelated → **diagonal covariance matrix**



Why Use Log Variance?

Variance must be **positive** (variance > 0)

Neural networks output unconstrained real numbers → use log variance

Convert log variance back via exponential for positivity:

$$\sigma = \exp(0.5 \times z\_log\_var)$$

Sampling from the Latent Distribution

Sample noise $\epsilon \sim \mathcal{N}(0, I)$

Latent vector: $z = z\_mean + z\_sigma \times \epsilon$

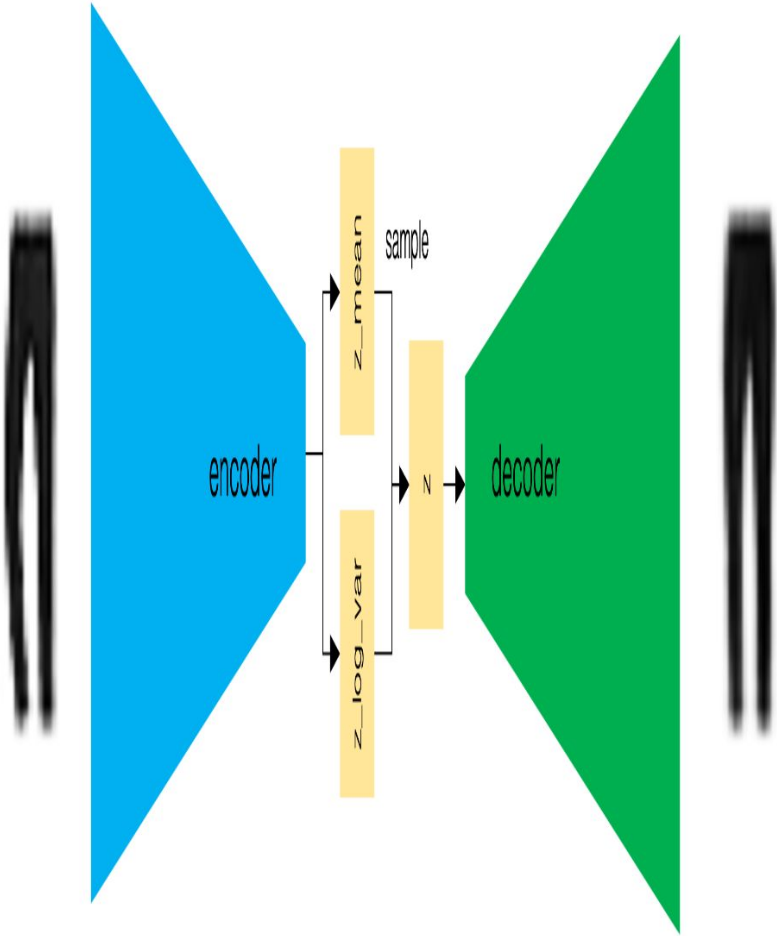
Introduces controlled randomness into encoding

The Decoder — Same as Before

Decoder architecture **unchanged**

Receives **sampled** z and **reconstructs** image

Learns to decode regions instead of points



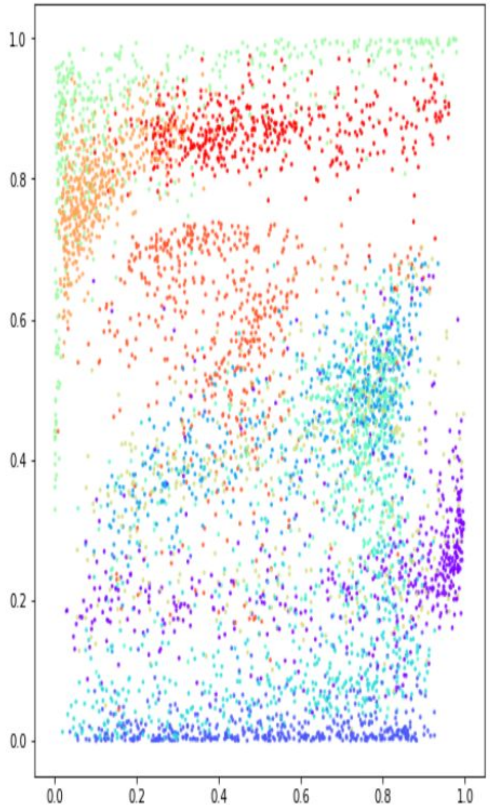
Why Does This Help? — The Continuity Advantage

Latent space becomes **continuous and smooth**

Neighboring points **decode** to similar outputs

Eliminates gaps and “**dead zones**” in latent space

Enables **realistic generation** from new latent points



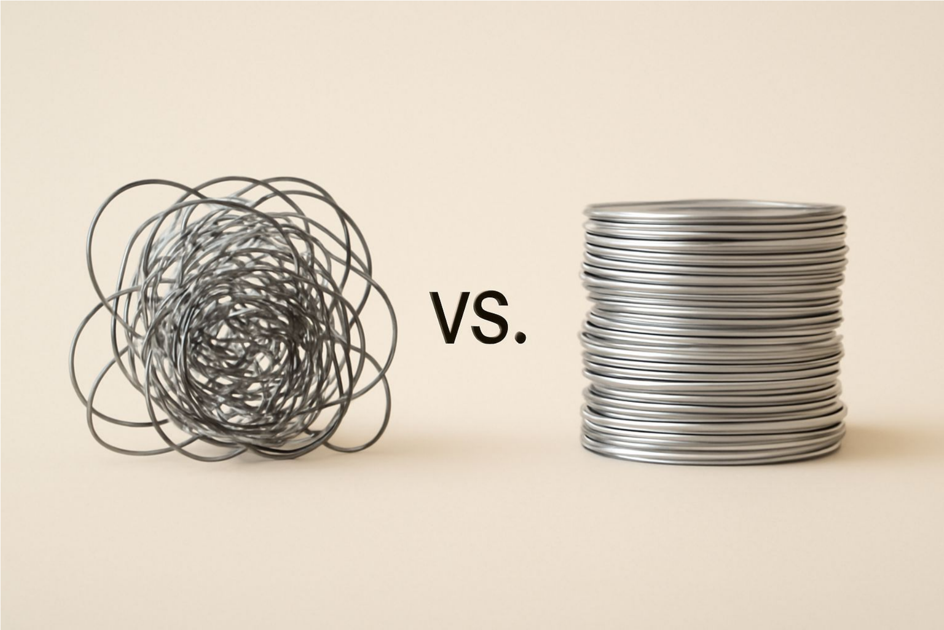
Understanding the VAE Loss Function

Recap – Standard Autoencoder Loss

Measures how **close** output is to the input

Usually MSE or Binary Cross-Entropy

Doesn’t care how latent space is structured

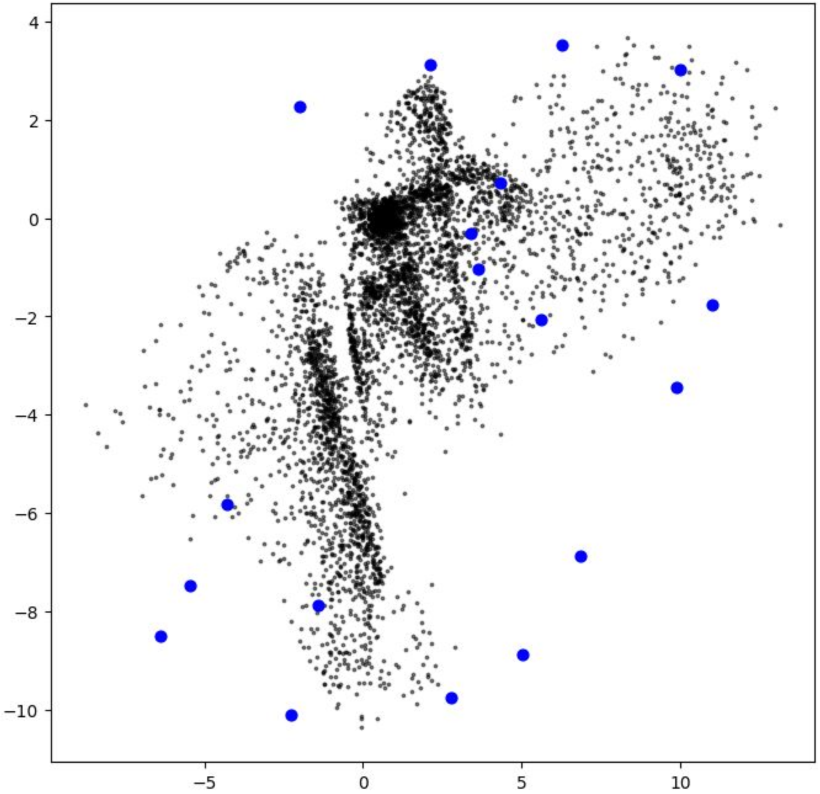


The Problem with Just Reconstruction Loss

Latent space can become discontinuous

No control over how inputs are organized in the latent space

Hard to sample new points

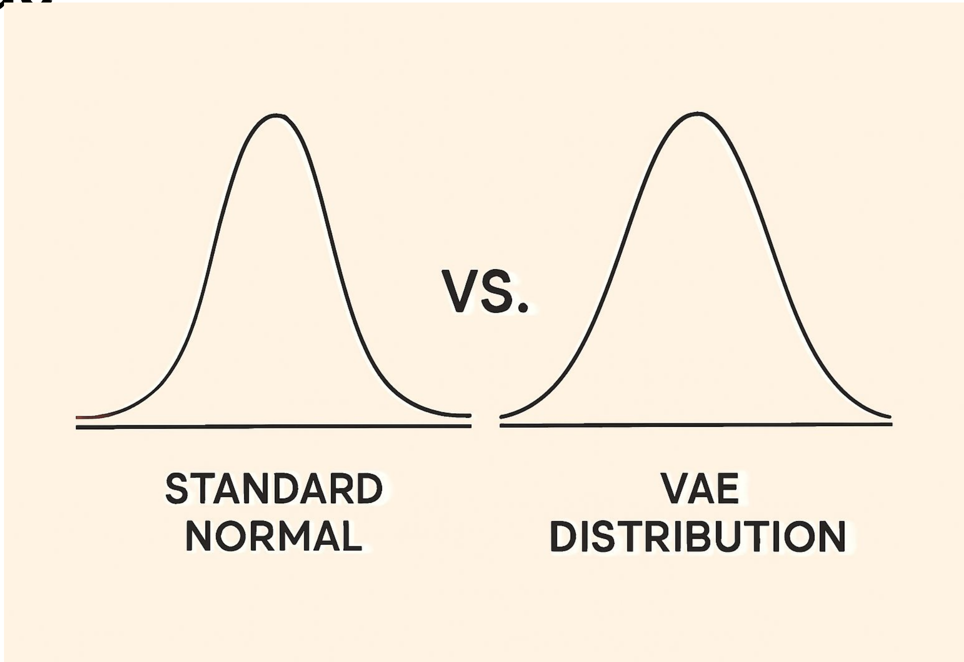


KL Divergence to the Rescue

Measures **how much one distribution differs from another**

In VAEs: compares encoded distribution to standard normal

Encourages smooth, continuous latent space

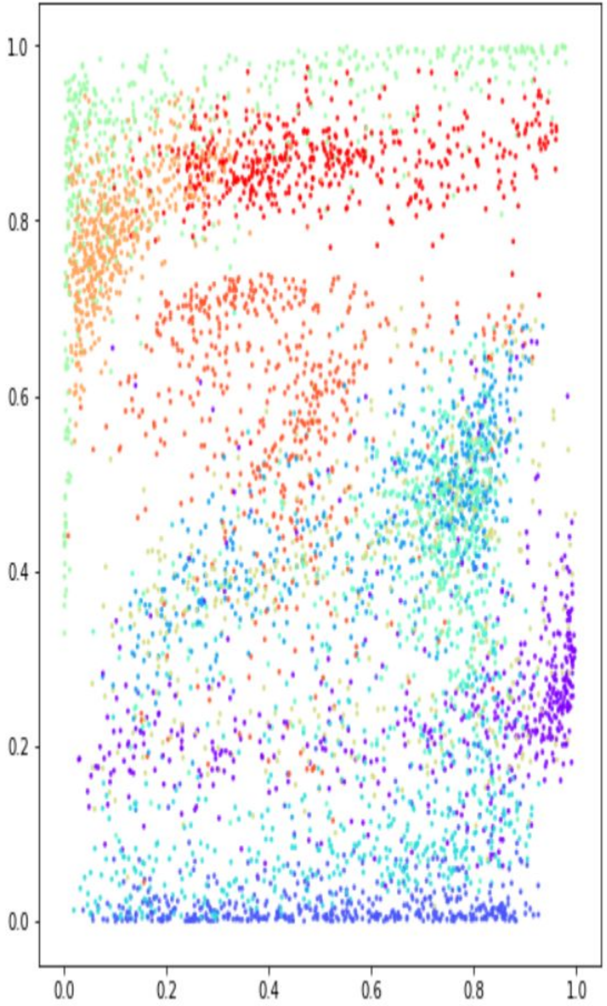


KL Divergence as a "Regularizer"

Encourages **z_mean ≈ 0** and **z_log_var ≈ 0**

Prevents overfitting to the training data

Makes latent space usage symmetric and efficient



The VAE Loss Function

Formula:

```
Total Loss = Reconstruction Loss + KL Divergence
```

KL Divergence term:

```
−0.5 * ∑ (1 + z_log_var − z_mean² − exp(z_log_var))
```

Why Is This Useful?

Enables **smooth interpolation** between points

Supports **sampling** and generative use

Reduces gaps in latent space

Introducing β-VAE

- β-VAE: `Loss = Reconstruction + β × KL`
- $\beta < 1$ → focus on reconstruction
- $\beta > 1$ → stronger regularization

How β Affects the Model

Low β → good reconstructions, poor latent structure

High β → structured latent space, blurry outputs

β must be tuned for your application

Key Takeaways

VAE = Reconstruction Loss + KL Divergence

KL divergence forces latent space toward standard normal

β balances reconstruction vs. structure

Helps with interpolation, sampling, and generation

Quiz Time!

Q1. What is the main difference between a standard autoencoder and a variational autoencoder (VAE)?

a. VAEs use convolutional layers instead of dense layers

b. VAEs encode inputs as a probability distribution instead of a fixed vector

c. VAEs don’t use a decoder

d. VAEs only reconstruct images, not data

Quiz Time!

Q2. What is the purpose of the KL divergence term in the VAE loss function?

a. To improve image quality

b. To prevent the decoder from overfitting

c. To force the latent distribution closer to a standard normal

d. To reduce training time

Quiz Time!

Q3. In a VAE, what do the variables z_mean and z_log_var represent?

a. Output image dimensions

b. Random noise values

c. Parameters of the learned latent distribution

d. Encoder input values

Quiz Time!

Q4. True or False: In VAEs, the decoder directly uses z_mean as input to reconstruct the data.

Quiz Time!

Q5. Why is β (in β-VAE) introduced in the loss function?

a. To penalize the decoder’s complexity

b. To control the balance between reconstruction accuracy and latent space regularization

c. To scale the reconstruction loss

d. To reduce the number of parameters in the encoder