Clase 157 — Variational Autoencoders (VAE)

Parte: 2 — Deep Learning · Fuente: Géron, cap. 17 § Variational Autoencoders + Kingma & Welling (2014). Duración estimada: 80 min.

Clase 157 — Variational Autoencoders (VAE)

> *Parte: 2 — Deep Learning · Fuente: Géron, cap. 17 § Variational Autoencoders + Kingma & Welling (2014). Duración estimada: 80 min.*

Objetivo

Construir un VAE (Variational Autoencoder, Kingma & Welling 2014) — variante probabilística del AE que aprende una distribución sobre el latent en lugar de un punto: encoder outputs μ, σ de una gaussiana; sampling + reparametrization trick para mantener gradientes. Resultado: latent space continuo y estructurado → permite generación, interpolación entre samples.

Resultados de aprendizaje

Al finalizar, el estudiante podrá:

- Implementar encoder que devuelve (μ, log σ²); sample con z = μ + σ · ε (reparametrization).
- Loss = reconstruction_loss + β · KL(N(μ,σ²) || N(0,I)).
- Generar samples nuevos: muestrear z ~ N(0,I), pasar por el decoder.
- Interpolar en el latent space y verificar transiciones suaves.
- Reconocer que VAE produce outputs borrosos (consecuencia del MSE/BCE) — motivó GANs (132).

Temas

- ELBO (Evidence Lower BOund): log p(x) ≥ E[log p(x|z)] - KL(q(z|x) || p(z)).
- Reparametrization trick: para back-propagar a través del sample.
- β-VAE: subir β fuerza latent más disentangled.
- Posterior collapse: cuando el decoder ignora z.

Definiciones y características

- q(z|x): encoder, devuelve μ(x), σ(x).
- p(z): prior, típicamente N(0, I).
- p(x|z): decoder.
- ELBO: lower bound de log-likelihood que se maximiza.
- KL divergence: KL(N(μ,σ²) || N(0,I)) = 0.5 · Σ (1 + log σ² - μ² - σ²) con signo.
- β-VAE: scale del término KL, controla trade-off reconstrucción vs estructura latente.

Dataset / recursos

- Fashion-MNIST / MNIST / Celeb-A (cara).
- Librerías: tensorflow, keras.

Ejercicios

1. VAE básico: encoder → (z_mean, z_log_var) → sample → decoder. Loss combinada. Entrenar en

MNIST.

2. Sampling: muestrear z ~ N(0,I) de tamaño (100, latent_dim). Pasar por decoder. Visualizar las 100 imágenes generadas.
3. Interpolación: dos imágenes A y B → z_A, z_B. Generar 10 imágenes en interpolación lineal entre z_A y z_B. Visualizar.
4. β-VAE: probar β=1, β=5, β=10. Comparar disentanglement vs blurriness.
5. Posterior collapse: con LR alto, el encoder colapsa a μ=0, σ=1. Diagnosticar mirando z_mean.std() cerca de 0.

Homework verificable

VAE sobre Fashion-MNIST:

1. Encoder convolucional, latent_dim=10.
2. Decoder simétrico.
3. Loss: BCE + KL.
4. Entrenar 30 épocas; generar 64 muestras nuevas y visualizar grid.
5. Interpolación entre 2 prendas distintas.

Criterio de aceptación: muestras generadas son reconocibles como prendas (aunque borrosas); interpolación es suave.

Errores comunes

Síntoma / mensaje	Causa y cómo arreglar
Posterior collapse: encoder predice μ=0, σ	KL domina. Fix: KL annealing (subir β grad
Outputs borrosos	Consecuencia inherente de VAE + MSE/BCE. F
KL = 0 → AE puro	β=0 → no es VAE. Fix: β > 0.
Sampleo con z ~ N(μ, σ) en lugar de N(0,I)	Eso es reconstrucción + ruido, no generaci
Generación muestra solo 1 modo	Posterior collapse parcial. Fix: arquitect

Preguntas frecuentes

¿VAE en 2026?

Como generador end-to-end: superado por difusión. Como encoder dentro de Stable Diffusion (latent space comprimido), sí.

¿Por qué reparametrization trick?

Sin él, no se puede back-propagar a través del sample (operación estocástica). Con z = μ + σ·ε, ε es noise externo independiente, gradientes fluyen por μ y σ.

¿Qué es disentanglement?

Cada dimensión latente captura un factor independiente (rotación, color, tamaño). β-VAE y FactorVAE lo persiguen explícitamente.

¿VAE para texto?

Sí, VAE de texto histórico. Hoy LLMs autoregresivos lo cubren mejor.

¿VQ-VAE?

VAE con codebook discreto. Base de muchos modelos modernos: DALL-E 1, Jukebox, EnCodec (audio). Importante.

Referencias

- Géron, cap. 17 — Variational Autoencoders.
- Kingma & Welling (2014), Auto-Encoding Variational Bayes, ICLR.
- Higgins et al. (2017), β-VAE, ICLR.
- van den Oord et al. (2017), Neural Discrete Representation Learning (VQ-VAE), NeurIPS.

Siguiente clase

Clase 158 — GANs: DCGAN, Progressive GAN, StyleGAN

Apéndice: notebook (primer bloque)

Primera celda ejecutable del notebook de la clase.

```
# Imports y configuración inicial
```

Archivos complementarios

- notebook.ipynb