Clase 156 — Autoencoders: undercomplete, stacked, denoising, sparse

Parte: 2 — Deep Learning · Fuente: Géron, cap. 17 § Autoencoders, GANs, and Diffusion Models. Duración estimada: 70 min.

Clase 156 — Autoencoders: undercomplete, stacked, denoising, sparse

> *Parte: 2 — Deep Learning · Fuente: Géron, cap. 17 § Autoencoders, GANs, and Diffusion Models. Duración estimada: 70 min.*

Objetivo

Entender autoencoders — red Encoder → bottleneck → Decoder entrenada a reconstruir su input. Variantes que cubrimos: undercomplete (dim_latent < dim_input, fuerza compresión), stacked (deep), denoising (input ruidoso → output limpio), sparse (penaliza activaciones latentes). Saber qué problemas resuelven (compresión, anomaly detection, pretraining) y cuándo VAEs/GANs/Diffusion los superan en generación.

Resultados de aprendizaje

Al finalizar, el estudiante podrá:

- Construir un undercomplete AE con MLP/CNN y entrenarlo con MSE.
- Diferenciar latent_dim < input_dim (compresión real) de latent_dim >> input_dim con regularización (sparse).
- Implementar Denoising AE: input x + noise, target x.
- Aplicar AE como anomaly detector: alta reconstruction error → anomalía.
- Reconocer que autoencoders no son buenos generadores (latent space irregular) → motivó VAE (clase 131).

Temas

- Encoder + Decoder simétricos.
- Bottleneck: latent space.
- Undercomplete: dimensión chica.
- Stacked: varias capas Dense/Conv.
- Denoising: robusto a ruido.
- Sparse: penalizar ||latent||_1 para que pocas neuronas activas.
- AE para anomaly detection.

Definiciones y características

- Bottleneck: capa con menor dimensión que input, fuerza al modelo a comprimir.
- Reconstruction loss: MSE o BCE pixel-wise.
- Denoising AE: input contaminado → target original. Aprende invariancias.
- Sparse AE: agrega L1 sobre las activaciones latentes a la loss.
- Tied weights: usar W^T del encoder como decoder. Reduce parámetros.

Dataset / recursos

- Fashion-MNIST / MNIST.
- Librerías: tensorflow, keras.

Ejercicios

1. AE simple: Encoder: 784 → 64; Decoder: 64 → 784. Entrenar en MNIST. Visualizar reconstrucciones.
2. Latent space 2D: latent_dim=2. Plot scatter de las representaciones de 1000 imágenes coloreadas por clase.
3. Denoising: noise = 0.5 * rng.normal(x.shape), target = x. Mostrar que reconstruye limpio aunque input está ruidoso.
4. Sparse: agregar keras.regularizers.l1(1e-3) sobre la capa latente. Inspeccionar activaciones.
5. Anomaly detection: entrenar AE solo sobre clase "normal"; calcular reconstruction error en clase "anomalía"; usar como score.

Homework verificable

AE como anomaly detector en Fashion-MNIST:

1. Entrenar AE convolucional solo sobre clase 0 (T-shirt).
2. Calcular MSE de reconstrucción para todas las imágenes test.
3. Plotear histograma del MSE separado por "T-shirt" vs "no T-shirt".
4. ROC-AUC de "es T-shirt" usando MSE como score (negativo).

Criterio de aceptación: ROC-AUC ≥ 0.85; el histograma muestra clara separación.

Errores comunes

Síntoma / mensaje	Causa y cómo arreglar
latent_dim muy grande → AE copia el input	No comprime nada. Fix: latent_dim < input_
MSE pierde detalle	Lo borronea. Fix: BCE pixel-wise o percept
Decoder con Dense para imágenes → mucho pa	Fix: Conv2DTranspose para upsampling.
Visualizar latent de latent_dim=64 directa	No se puede visualizar 64D. Fix: t-SNE / U
AE para generar nuevas imágenes → outputs	Latent space no es continuo. Fix: VAE (131

Preguntas frecuentes

¿AE moderna?

Sí, especialmente como encoder backbone para auto-supervised pretraining (MAE — Masked Autoencoders en visión, He et al. 2022).

¿AE vs PCA?

Linear AE con MSE = PCA. AE con activaciones no lineales = PCA no lineal. Buena ganancia con datos no lineales.

¿Tied weights?

Reducen params 2× sin perder mucho. Usado históricamente, hoy raro porque GPUs y datos son abundantes.

¿Denoising AE en producción?

Sí, para imagen denoising, audio. Pero modelos de difusión (133) lo hacen mejor.

¿Sparse AE relevante?

Más histórico. Hoy se usan Vector Quantized AE (VQ-VAE) para representaciones discretas (audio, video → tokens).

Referencias

- Géron, cap. 17 — Autoencoders.
- Vincent et al. (2008), Extracting and Composing Robust Features with Denoising Autoencoders.
- He et al. (2022), Masked Autoencoders Are Scalable Vision Learners (MAE), CVPR.

Siguiente clase

Clase 157 — Variational Autoencoders (VAE)

Apéndice: notebook (primer bloque)

Primera celda ejecutable del notebook de la clase.

```
# Imports y configuración inicial
```

Archivos complementarios

- notebook.ipynb