

Resumen sobre Autoencoders, Segmentación y RAGs: conceptos y arquitecturas

Andrés Sánchez Rojas
Escuela de Ingeniería en Computación
Instituto Tecnológico de Costa Rica
16/10/2025

Resumen—Estos son los apuntes de la segunda clase de semana 11 del curso de IA. Los autoencoders son modelos de aprendizaje no supervisado que aprenden a reconstruir sus entradas a través de una representación comprimida (espacio latente). En este documento se resumen conceptos clave: arquitectura encoder/decoder, variational autoencoders (VAE) y la reparametrización, funciones de pérdida típicas, y aplicaciones como detección de anomalías y denoising. También se introduce la segmentación de imagen y la arquitectura U-Net, y se discuten conceptos relacionados con RAGs (Retrieval Augmented Generation), agentes basados en LLM, tokenización y embeddings. El objetivo es ofrecer una visión compacta y legible para un lector que busca una introducción técnica y aplicada.

Index Terms—Autoencoders, Variational Autoencoders, U-Net, Segmentación de imagen, RAG, Tokenización, Embeddings

I. INTRODUCCIÓN

Este documento sintetiza los principios y aplicaciones prácticas de los autoencoders, describiendo su entrenamiento sin etiquetas por reconstrucción, sus usos en reducción de dimensionalidad, compresión, detección de anomalías, denoising y superresolución, y la extensión hacia variantes relevantes como los variational autoencoders; además presenta la conexión con tareas de visión por computador (p. ej., segmentación y arquitecturas tipo U-Net) y la extensión a representaciones para texto mediante tokenización y embeddings, así como su papel en sistemas más amplios como RAGs y agentes basados en LLM. El texto ofrece una guía práctica con definiciones, fórmulas y recomendaciones operativas para implementar experimentos en imágenes y texto.

II. AUTOENCODERS

II-A. Definición y propósito

Los autoencoders son una arquitectura novedosa en el ámbito del aprendizaje automático que se caracteriza por comparar sus salidas con las mismas entradas, lo que permite entrenarlos sin necesidad de etiquetas, clasificándolos como modelos no supervisados. Su principal utilidad radica en la reducción de dimensionalidad, ofreciendo representaciones más potentes que técnicas clásicas como el Análisis de Componentes Principales (PCA). Esta capacidad de compresión y reconstrucción los hace especialmente útiles en tareas como la detección de anomalías, la identificación de transacciones fraudulentas, la eliminación de ruido en datos, el aumento de resolución (upsampling) y el reconocimiento facial. En esencia, los autoencoders aprenden una codificación eficiente de los datos, lo que les permite capturar

patrones relevantes y reconstruir las entradas con alta fidelidad. Esta versatilidad los convierte en herramientas valiosas tanto en aplicaciones de seguridad como en procesamiento de imágenes y señales.

II-B. Encoder y Decoder

El *encoder* es un conjunto de bloques convolucionales que extraen la información más relevante de la entrada y descartan lo que no aporta, comprimiendo los datos a través de un “cuello de botella” para eliminar ruido y características innecesarias; la salida de ese proceso es el vector o espacio latente, una representación de baja dimensionalidad que conserva los rasgos útiles para diferenciar patrones. El *decoder* es la parte que toma el espacio latente y reconstruye la señal o imagen original, expandiendo la información comprimida para producir una salida lo más fiel posible a la entrada; su objetivo es invertir la codificación del encoder y permitir tareas como denoising, upscaling o detección de anomalías mediante la comparación entre entrada y reconstrucción.

II-C. Aplicaciones

Entre las aplicaciones prácticas destacan:

- Detección de anomalías y fraude .
- Eliminación de ruido (denoising).
- Aumento de resolución (upsampling).
- Reconocimiento facial y compresión de imágenes.

III. VARIATIONAL AUTOENCODER

Son una variante probabilística de los autoencoders que generan una representación latente continua modelada como una distribución. En lugar de devolver un único vector latente determinista, el encoder estima parámetros de una distribución: la media $\mu(x)$ y la log-varianza $\log \sigma^2(x)$.

III-A. Reparametrización

La reparametrización permite que la aleatoriedad se aísle en una variable independiente, de forma que los gradientes puedan fluir hacia los parámetros que predicen la media y la varianza.

III-B. Funciones de pérdida

La pérdida de un variational autoencoder combina dos términos:

1. **Reconstruction loss:** mide la discrepancia entre la entrada y la reconstrucción producida por el decoder.
2. **KL divergence:** Compara qué tanto se parecen dos distribuciones.

La pérdida total habitual es la suma: Reconstruction loss + KL Divergence.

III-C. Espacio latente y generación

Un beneficio importante de los variational autoencoders es que el espacio latente resultante es continuo: puntos cercanos en el espacio latente generan observaciones similares, lo que permite interpolación y generación de nuevas muestras mediante muestreo.

IV. SEGMENTACIÓN DE IMAGEN

La segmentación consiste en localizar y etiquetar píxeles que pertenecen a objetos de interés dentro de una imagen. Devuelve un mapa donde cada píxel tiene una etiqueta, siendo útil en aplicaciones que requieren alta resolución espacial como el análisis médico o el conteo de células.

IV-A. Arquitectura U-Net

U-Net es una arquitectura en forma de "U" similar a un autoencoder pero con skip connections entre las capas de encoder y decoder. Estas conexiones permiten conservar información durante el upsampling, mejorando significativamente la precisión de los mapas de segmentación. U-Net ha demostrado ser especialmente útil en tareas médicas como la identificación de células cancerígenas.

V. RAGS Y AGENTES

V-A. RAG: Retrieval Augmented Generation

Los RAGs combinan recuperación de documentos relevantes con generación de lenguaje. El flujo general es:

1. Convertir la consulta y fragmentos de texto en embeddings.
2. Buscar textos relevantes en una base de conocimiento mediante medidas de similitud en el espacio de embeddings.
3. Pasar los fragmentos recuperados como contexto a un modelo de lenguaje para generar respuestas más fundamentadas.

V-B. Agentes basados en LLM

Los agentes usan un LLM como núcleo de decisión para orquestar pasos (consultar fuentes, ejecutar APIs, leer memoria). Un agente integra recuperación, gestión del contexto y conectores a herramientas externas para resolver tareas complejas de forma autónoma.

VI. TOKENIZACIÓN Y EMBEDDINGS

VI-A. Tokenización

La tokenización convierte texto en secuencias de identificadores. Estrategias comunes de tokenización son: por palabra, por subword, por carácter, por bytes. Cada estrategia tiene trade-offs en cobertura, eficiencia y manejo de formas raras.

VI-B. Embeddings

Los embeddings son vectores densos que representan tokens o secuencias en un espacio donde la proximidad indica similitud semántica. Al agregar embeddings de tokens (por ejemplo mediante promedio o modelos que producen representaciones de secuencia) se obtienen vectores de frases/consultas útiles para búsquedas semánticas y recuperación en RAGs, y como entrada para razonamiento en agentes.

VII. CONCLUSIONES

Los puntos vistos en esta clase y resumidos en este documento ofrecen una síntesis compacta de conceptos relevantes en autoencoders, variational autoencoders, segmentación de imágenes con arquitecturas como U-Net, y RAGs y agentes basados en LLMs. Estos avances han permitido la expansión del uso de la inteligencia artificial en ambientes en los que antes no se hubiera considerado útil. Sin embargo, debemos ser responsables al decidir qué tareas realmente requieren un agente o pueden usar un sistema más ligero de machine learning.