

DEEP LEARNING NO SUPERVISADO

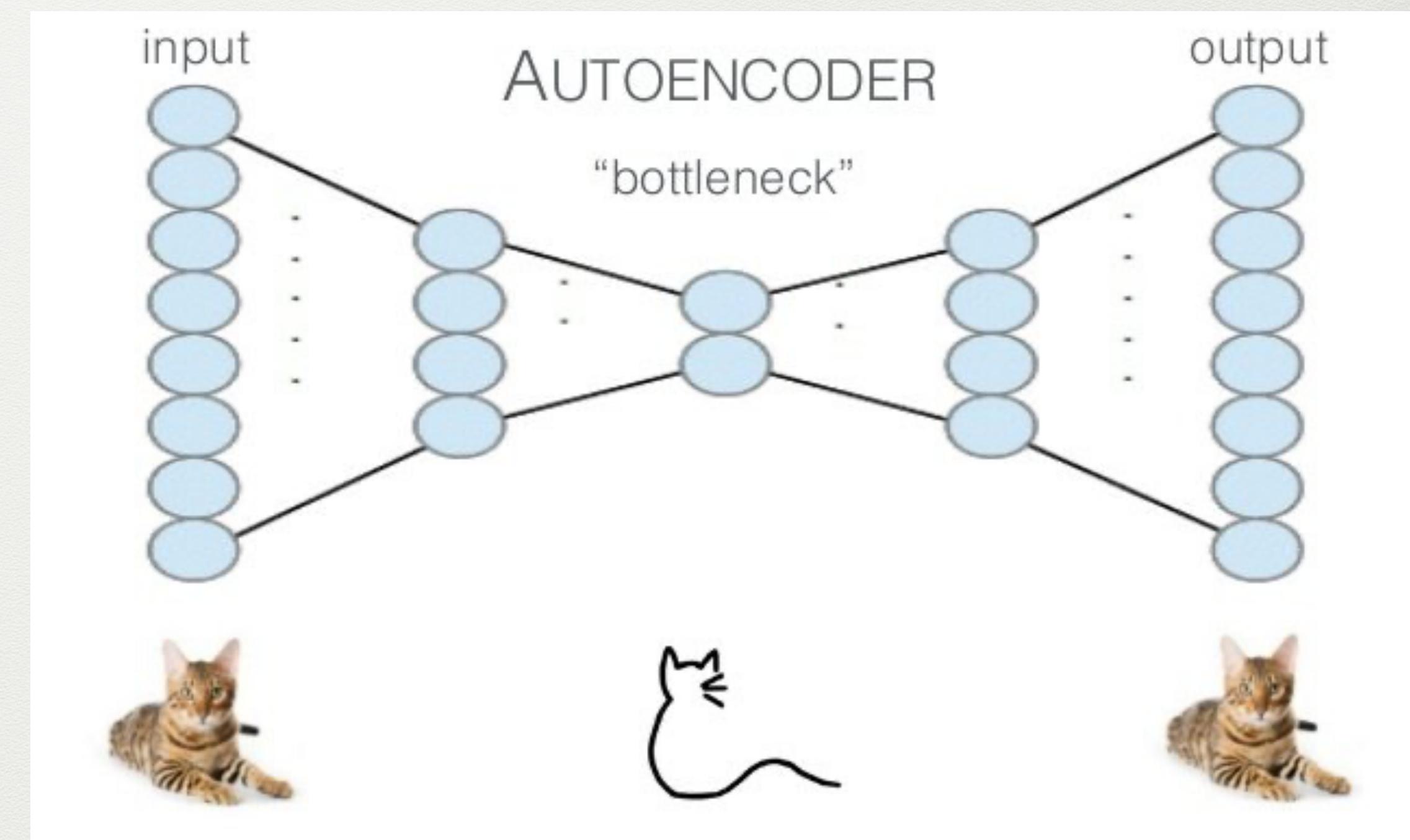
Física Computacional II
Ph.D. Santiago Echeverri Arteaga

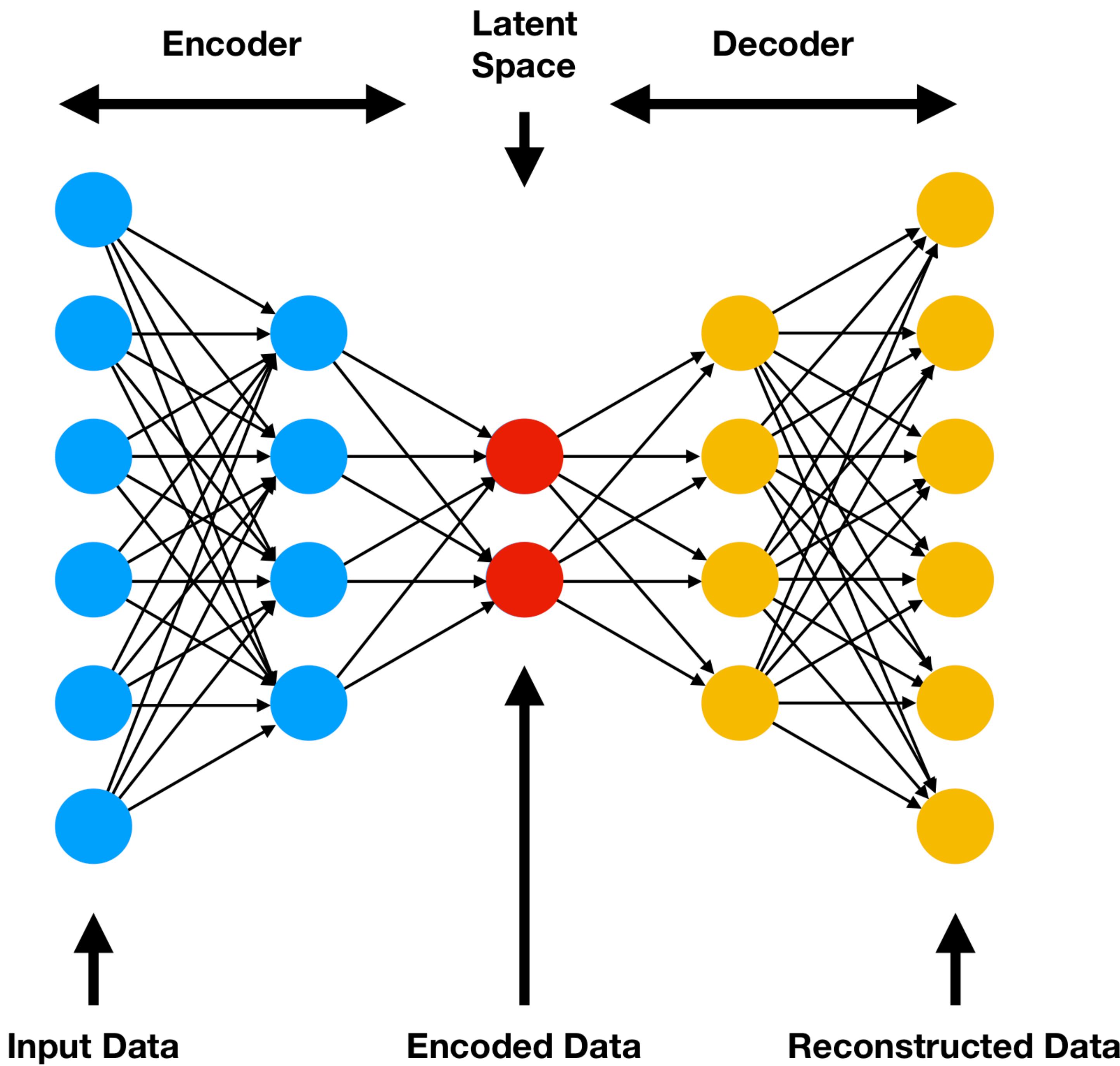
AUTOENCODERS

Continuando la estrategia Encoder-Decoder
Machine Learning Auto-Supervisado

Autoencoder

- **Encoder:** Red feedforward que descompone el input reduciendo el número de neuronas iniciales. El espacio de pocas neuronas al que llega se llama espacio Latente $\phi : \mathcal{X} \rightarrow \mathcal{F}$
- **Decoder:** Red feedforward que toma como input la salida del **encoder** (espacio latente) e intenta reproducir la entrada original al encoder.
 $\psi : \mathcal{F} \rightarrow \mathcal{X}'$
- Forward pass: Pasa la información a través del **encoder** y su salida se la pasa al **decoder**.
- Si no ponemos función de activación, recuperamos PCA
- Loss: El loss se calcula a partir de la diferencia entre los datos originales y la salida del decoder.
- Si el decoder logra reproducir exactamente la imagen original, el encoder está reduciendo la imagen de dimensionalidad sin perder información





Autoencoder

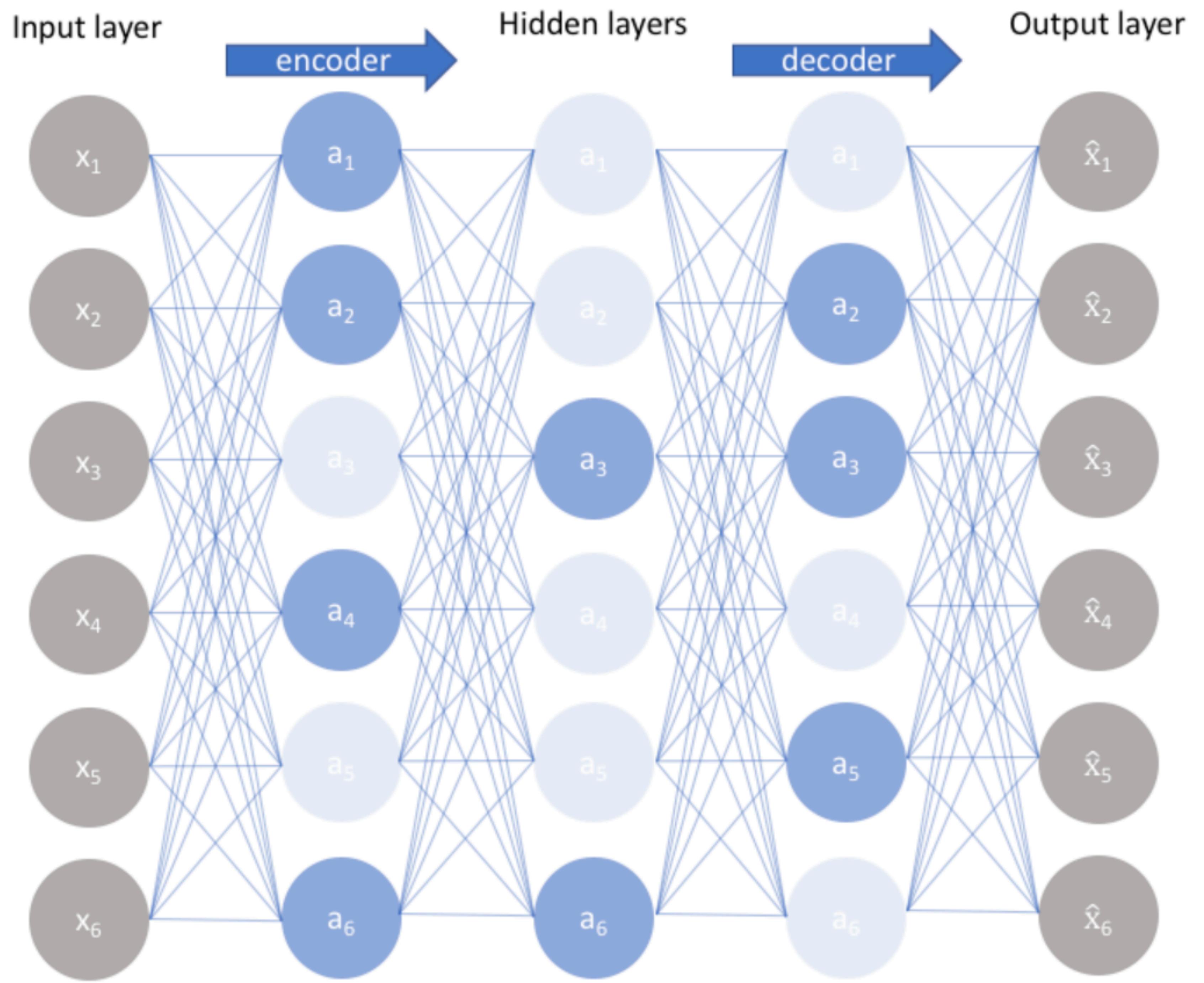
- Usos:
 - Reducción de dimensionalidad
 - Preprocesamiento para clasificación.
 - Identificar los elementos esenciales de los datos de entrada filtrando el ruido
 - Compresión de información
- PCA tiene limitaciones, pues las nuevas características deben ser combinaciones LINEALES de las features originales. Además que la mejor representación depende del problema estudiado.
- No suele ser usado para generar imágenes

Autoencoder de eliminación de ruido

Agrega ruido blanco a la imagen previo al encoder pero en el entrenamiento compara con la imagen original

Sparse Autoencoder

- Agrega al loss un término de regularización L1 (o L2) que penaliza al vector de activación de cada layer h para cada observación i
$$L(x, \hat{x}) \rightarrow L(x, \hat{x}) + \lambda \sum_i |a_i^{(h)}|$$
- Es usado en sistemas de recomendación profundos
- En general es una mejora al autoencoder normal a costa de un hiperparámetro



Problemas de los Autoencoders

- Gaps en el espacio latente
- Topologías no conexas en el espacio latente
- Espacio latente discreto

Autoencoder Variacionales (2013)

<https://arxiv.org/abs/1401.4082>

<https://arxiv.org/abs/1312.6114>

- Se asume que los datos son representados por un conjunto de variables aleatorias **normalmente distribuidos** ($N(\mu_1, \sigma_1), \dots, N(\mu_N, \sigma_N)$)
- El encoder genera los parámetros de esas distribuciones (μ, σ) los cuales se suman (y el sigma se multiplica por un vector de ruido aleatorio ϵ). A partir de esas **distribuciones** el decoder reconstruye la imagen
- Mejora la interpretabilidad
- Se pueden generar imágenes a partir de las distribuciones
- Imágenes similares están cerca en el espacio latente
- Al Loss se agrega el término de divergencia KL (Divergencia de (μ, σ) de una distribución normal estándar $\frac{1}{2} (e^{\log(\sigma)} - (\log(\sigma) + 1) + \mu^2)$) para penalizar distribuciones alejadas de la normal
- El objetivo es que el espacio latente de representaciones sea topológicamente conexo y con buenas propiedades de interpolación, en donde no se ajuste un solo punto sino toda una región

