



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA DE SISTEMAS INFORMÁTICOS  
UNIVERSIDAD POLITÉCNICA DE MADRID

# El auto-encoder variacional

## Un análisis del espacio latente

CÉSAR PÉREZ CUIEL

Máster en Ciencias y Tecnologías de la Computación

# Contenidos

## 1 Introducción

- Objetivos
- ¿Qué es un auto-encoder variacional?

## 2 El auto-encoder variacional

- La arquitectura de un VAE
- La función de coste
- El *reparameterization trick*

## 3 Resultados de los experimentos

- Reconstrucción de imágenes
- Generación de imágenes a partir de los centroides
- Generación de imágenes a partir de la distribución a priori
- Visualización del espacio latente
- Separación de las clases

## 4 Conclusiones

# Objetivos

## A nivel de conocimientos

- Entender el funcionamiento del auto-encoder variacional
- ¿Existe una estructura en el espacio latente?

## A nivel metodológico

- Desarrollar de forma autónoma un proyecto de investigación.

# ¿Qué es un auto-encoder variacional?

- Un modelo no supervisado
- Un modelo generativo
- Un modelo probabilístico

## En pocas palabras...

El auto-encoder variacional es un modelo que genera datos a partir de valores obtenidos como muestras de una distribución de probabilidad.

# Arquitectura

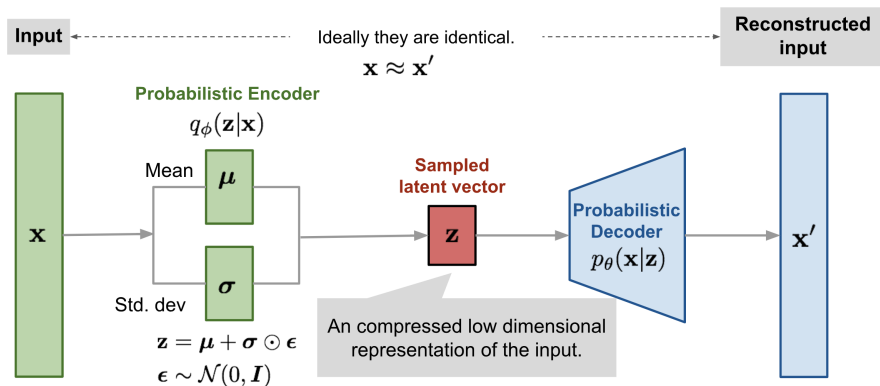


Figura: <https://lilianweng.github.io/posts/2018-08-12-vae/>

# La función de coste

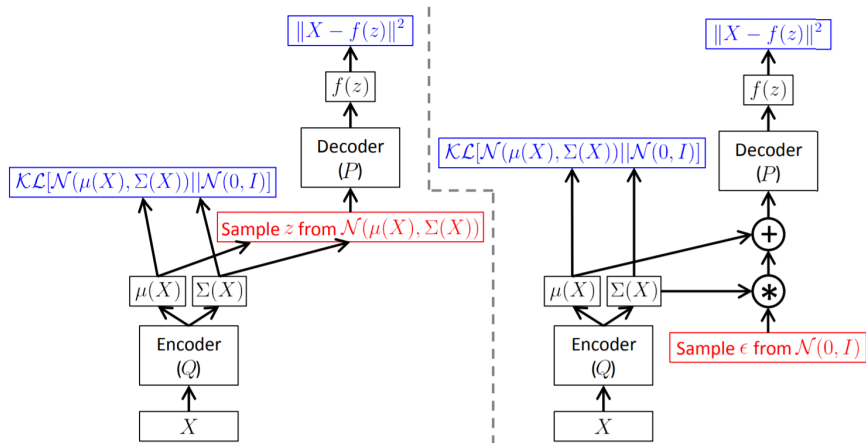
$$\log p_{\theta}(x) = \underbrace{\mathbb{E}_{z \sim q_{\phi}(z|x)} \left[ \log \frac{p_{\theta}(x, z)}{q_{\phi}(z|x)} \right]}_{\mathcal{L}_{\theta, \phi}(x)} + \underbrace{\mathbb{E}_{z \sim q_{\phi}(z|x)} \left[ \log \frac{q_{\phi}(z|x)}{p_{\theta}(z|x)} \right]}_{D_{\text{KL}}(q_{\phi}(z|x) || p_{\theta}(z|x))} \quad (1)$$

$$\mathcal{L}_{\theta, \phi}(x) = \mathbb{E}_{z \sim q_{\phi}(z|x)} [\log p_{\theta}(x|z)] - \underbrace{\mathbb{E}_{z \sim q_{\phi}(z|x)} \left[ \log \frac{q_{\phi}(z|x)}{p_{\theta}(z)} \right]}_{D_{\text{KL}}(q_{\phi}(z|x) || p_{\theta}(z))} \quad (2)$$

## Función de coste

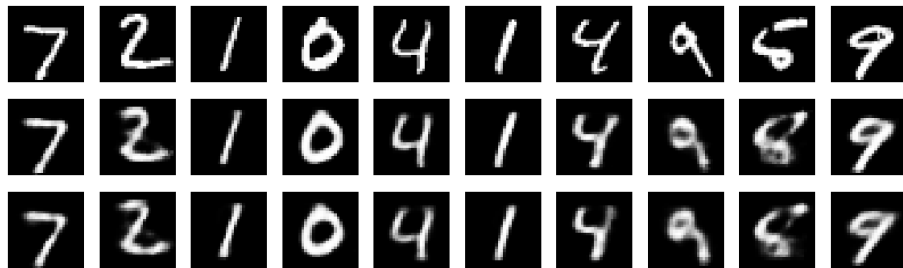
$$\mathcal{L}_{\theta, \phi}(x) = \mathbb{E}_{z \sim q_{\phi}(z|x)} [\log p_{\theta}(x|z)] - D_{\text{KL}}(q_{\phi}(z|x) || p_{\theta}(z))$$

# El reparameterization trick



**Figura:** Doersch, C. (2016). Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*.

# Reconstrucción de imágenes



## Comprobación

Los modelos parecen estar correctamente implementados.



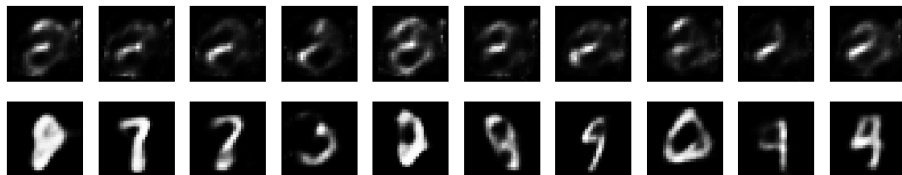
## Generación de imágenes a partir de los centroides



Algo no encaja...

¿No es un poco raro que los resultados sean tan parecidos cuando los espacios latentes son diferentes?

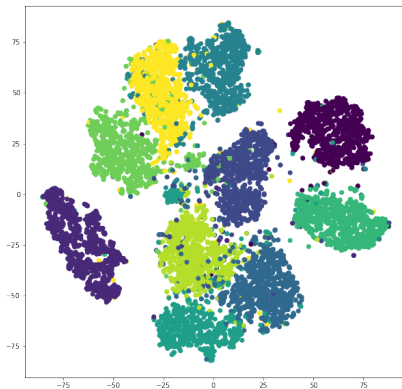
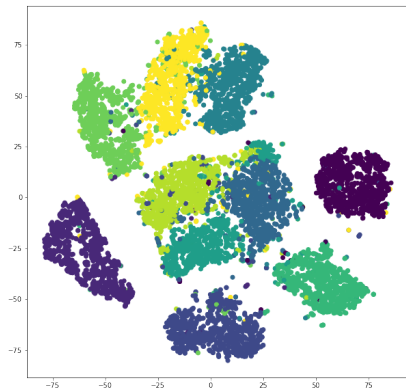
# Generación de imágenes a partir de la distribución a priori



## Resultado

- Las imágenes generadas por el auto-encoder estándar no parecen dígitos.
- ¿Por qué los dígitos generados por el VAE son tan primitivos?

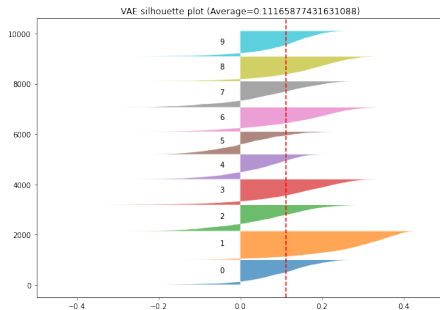
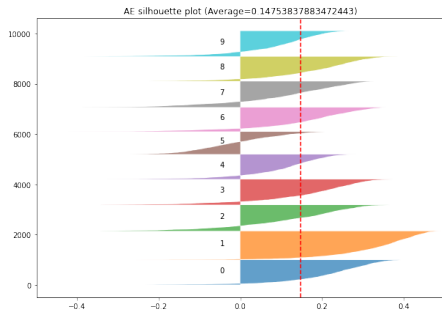
# Visualización del espacio latente



## Resultado

No parece haber una gran diferencia...

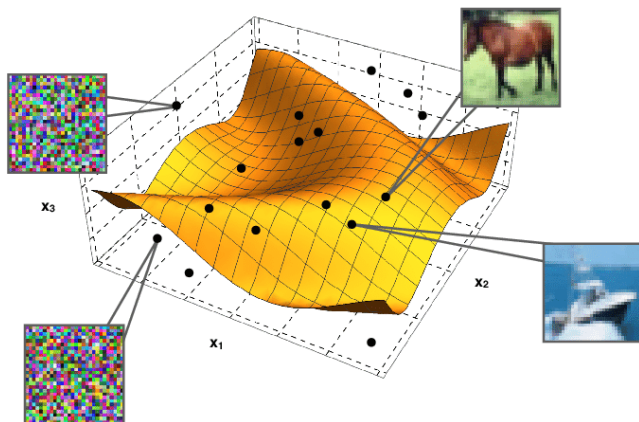
# Separación de las clases



## Resultado

El auto-encoder variacional no separa las clases mejor que un autoencoder estándar.

# Conclusiones: Generacion de imágenes



**Figura:** Goldt, S., Mézard, M., Krzakala, F., & Zdeborová, L. (2020). Modeling the influence of data structure on learning in neural networks: The hidden manifold model. *Physical Review X*, 10(4), 041044.

# Conclusiones: Separación de las clases

## El VAE no separa mejor las clases

Los resultados obtenidos usando el algoritmo SILHOUETTE muestran que el auto-encoder variacional no separa las clases mejor que el auto-encoder estándar.

## Siguientes pasos

- Estudio de las variantes del VAE
- Estudio en detalle de dichas propiedades desde el punto de vista del *Representation Learning*.

