

Redes Generativas Adversarias (GAN)

Sistemas de Inteligencia Artificial 2025

Generación de datos

Cada imagen 64x64x3
(alto, ancho, canales)

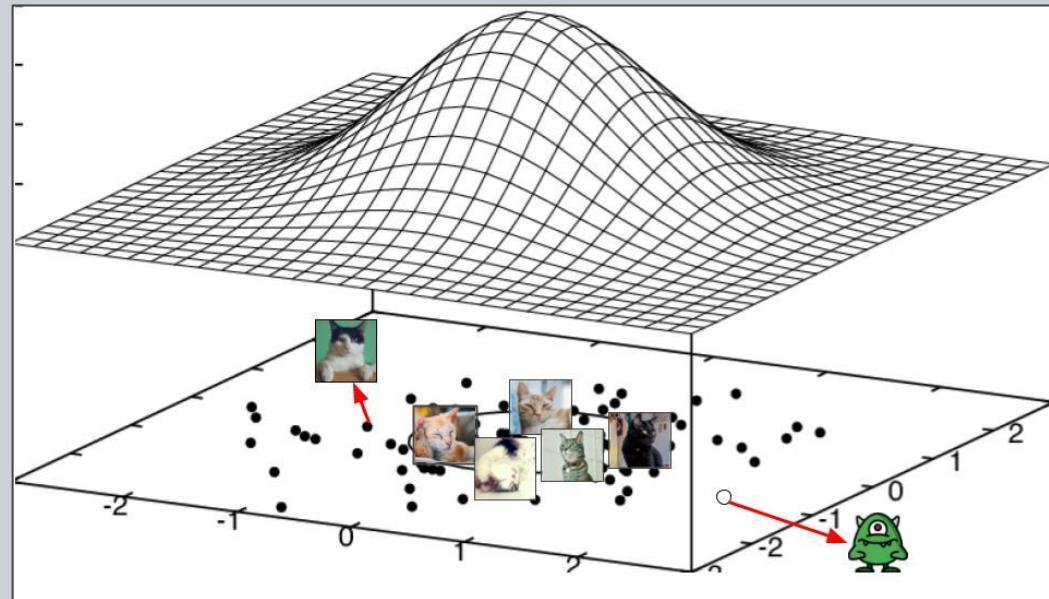


R^{12288}

200
30
58
220
196
•
•
•
•
•

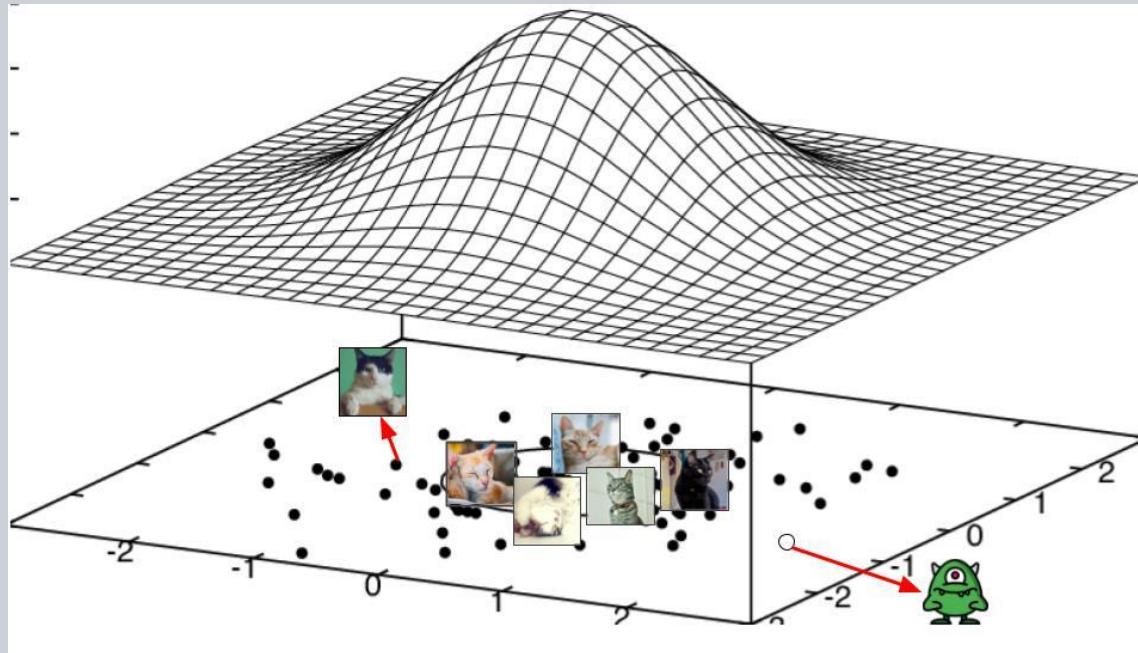


Función de distribución de probabilidad

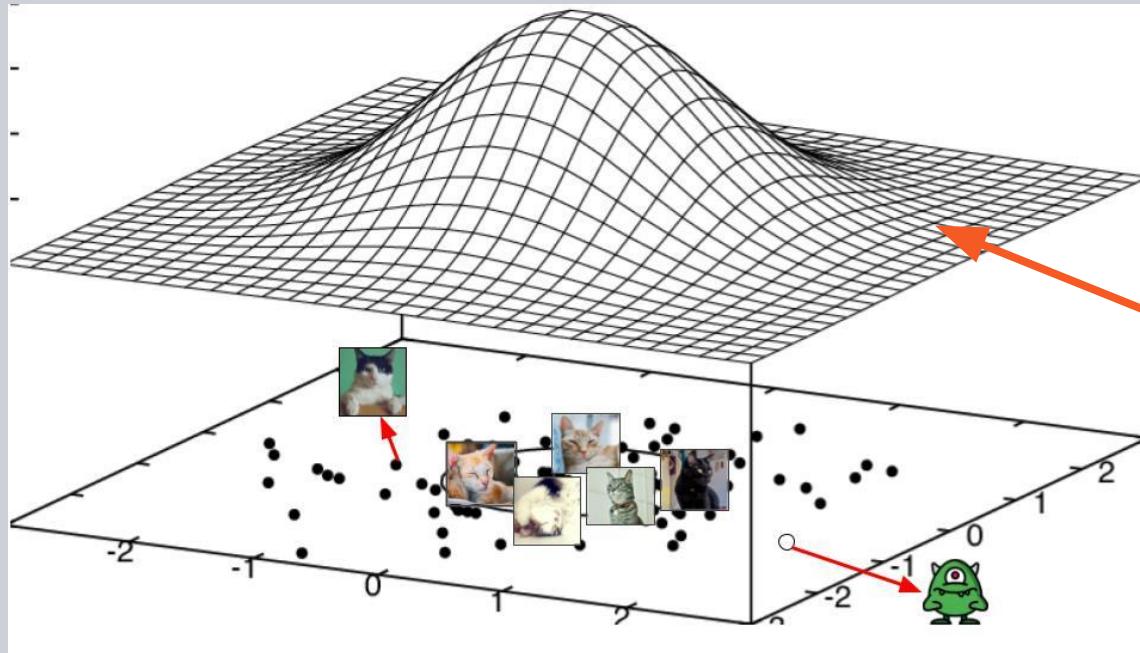


Espacio 12288-dimensional

Sea P_{data} la función de distribución de los datos
Busco generar $x' / x' \sim P_{\text{data}}$



Sea P_{data} la función de distribución de los datos
Busco generar $x' / x' \sim P_{\text{data}}$



¡No conocemos
esta
distribución!

Generative Adversarial Network

Generative Adversarial Nets

Ian J. Goodfellow, Jean Pouget-Abadie,^{*} Mehdi Mirza, Bing Xu, David Warde-Farley,
Sherjil Ozair,[†] Aaron Courville, Yoshua Bengio[†]

Département d'informatique et de recherche opérationnelle
Université de Montréal
Montréal, QC H3C 3J7

Abstract

We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G . The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D , a unique solution exists, with G recovering the training data distribution and D equal to $\frac{1}{2}$ everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

1 Introduction

The promise of deep learning is to discover rich, hierarchical models [2] that represent probability distributions over the kinds of data encountered in artificial intelligence applications, such as natural

Ian J. Goodfellow,
Jean Pouget-Abadie ,
Mehdi Mirza,
Bing Xu,
David Warde-Farley,
Sherjil Ozair,
Aaron Courville,
Yoshua Bengio

Año: 2014

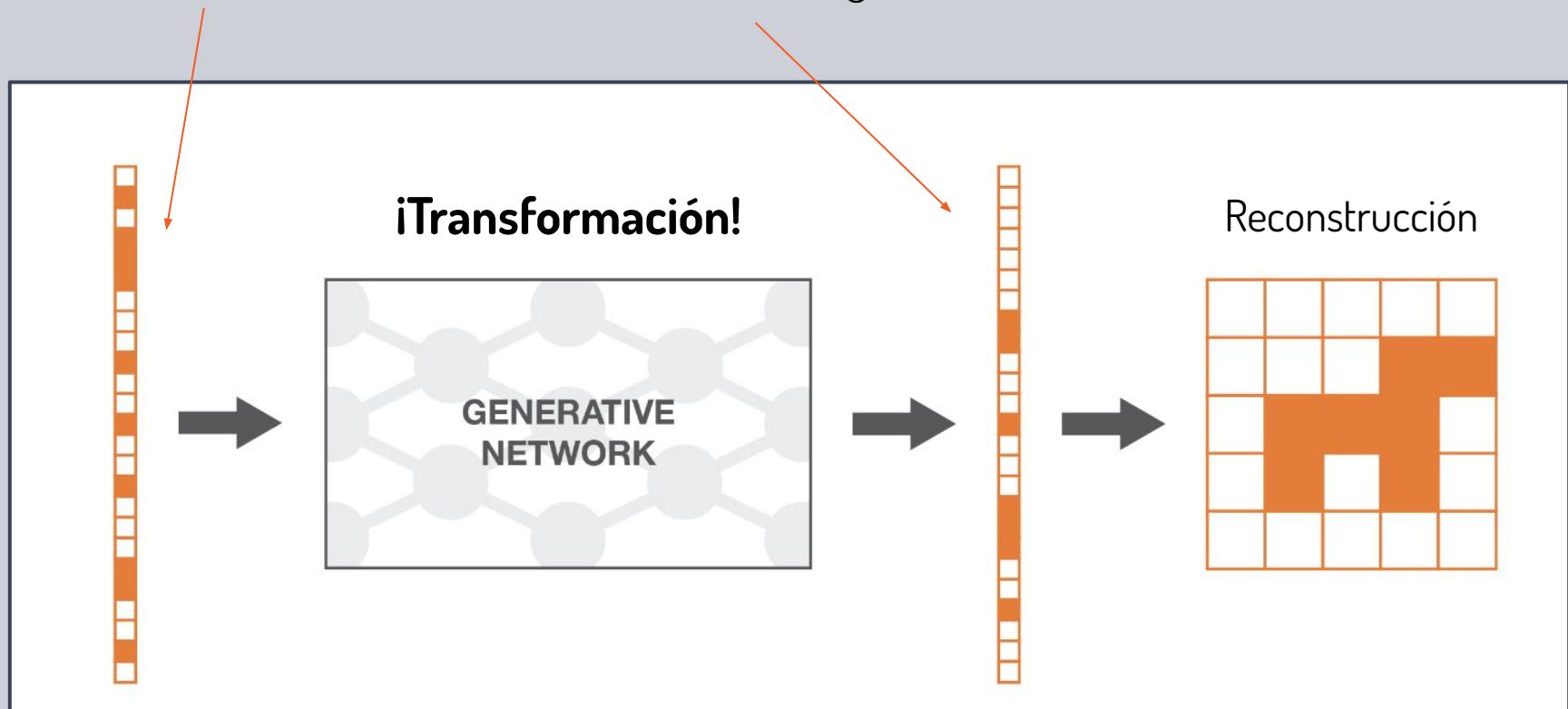


Muestras de distribución uniforme

Muestras de distribución de gatos

Imágenes: 64x64x3

Espacio: 12288-dimensional



Inverse Sampling Theorem

Muestras de distribución uniforme

Muestras
distribuidas

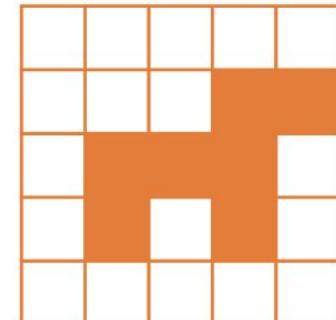
$Y \sim U[0, 1]$, F invertible y $F = CDF(pdf())$,
 $X = F^{-1}(Y)$ tiene $CDF(X) = F$



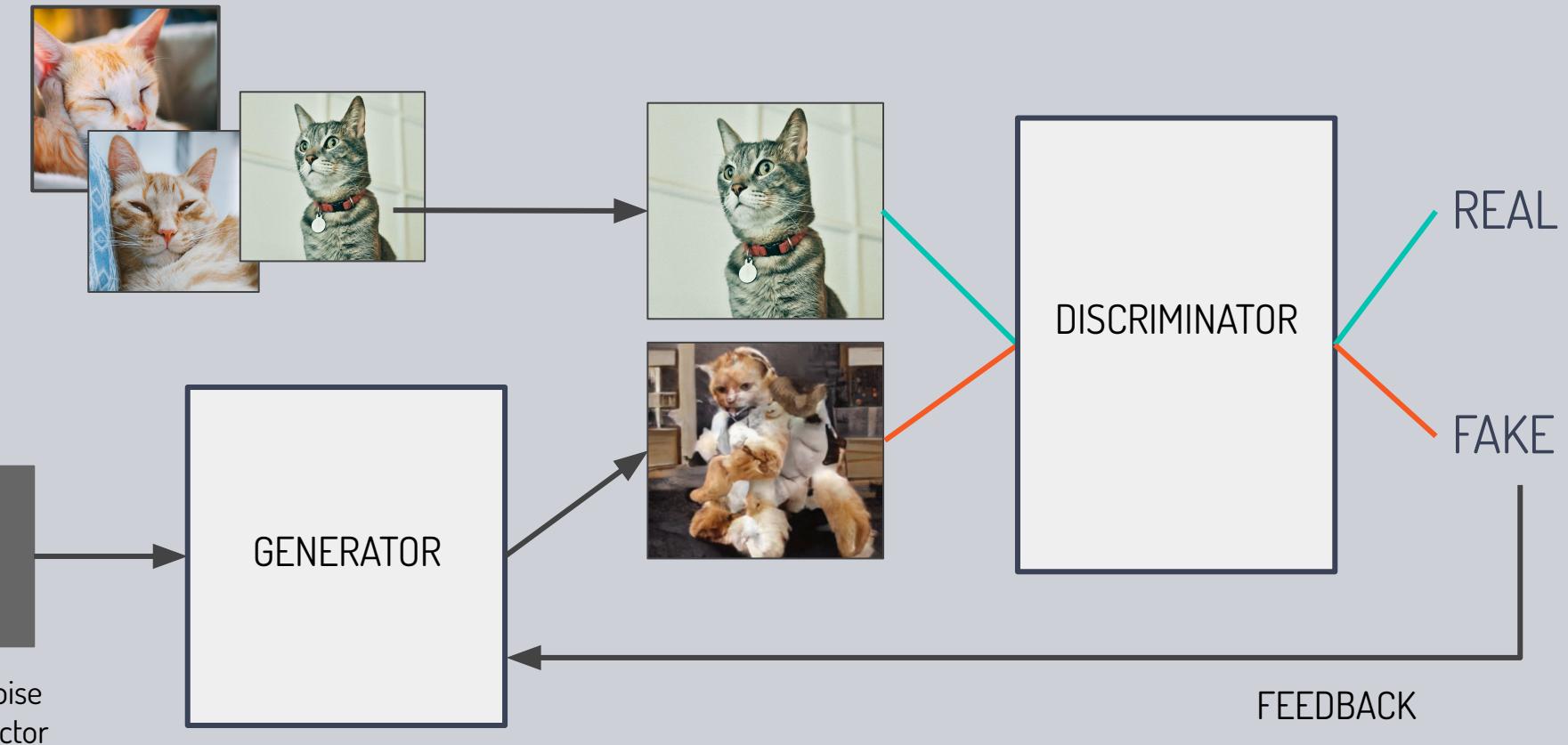
iTransformación!



Reconstrucción

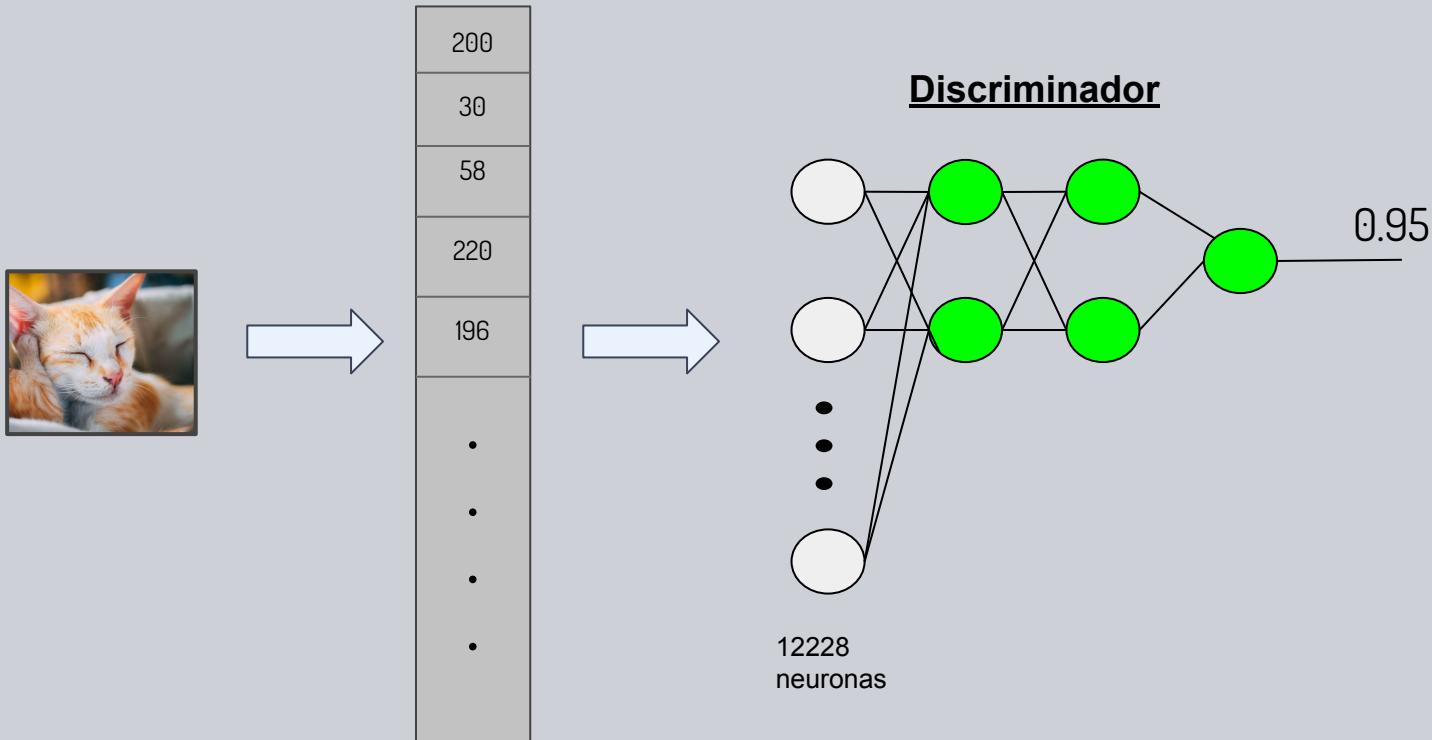


Generative Adversarial Network



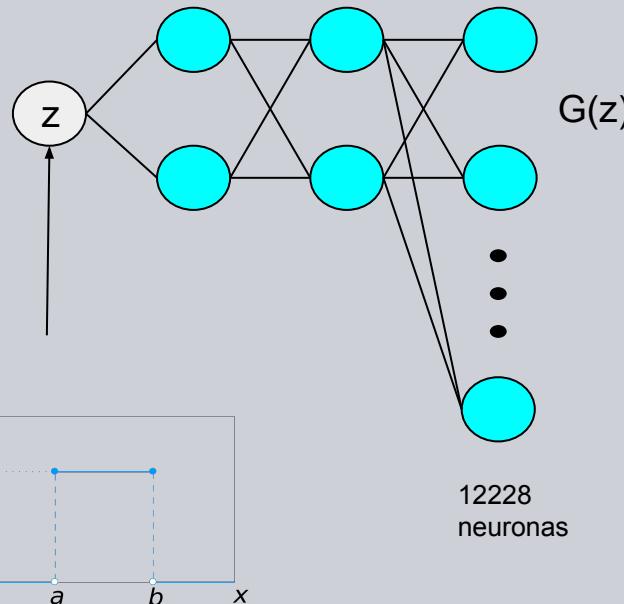
¿Cómo armamos el discriminador?

Generative Adversarial Network

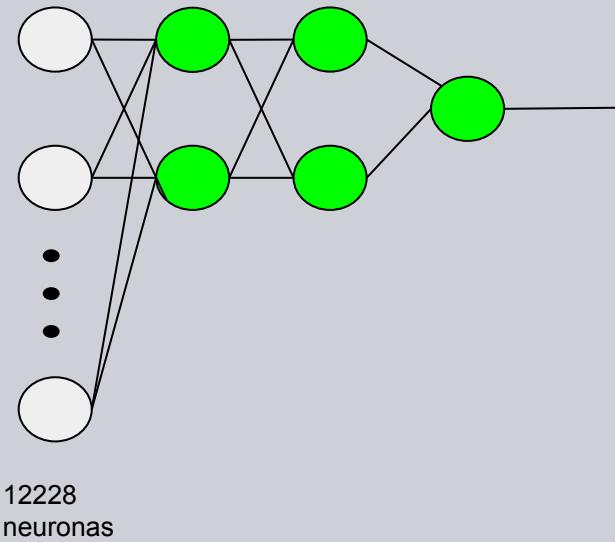


¿Cómo armamos el generador?

Generador



Discriminador

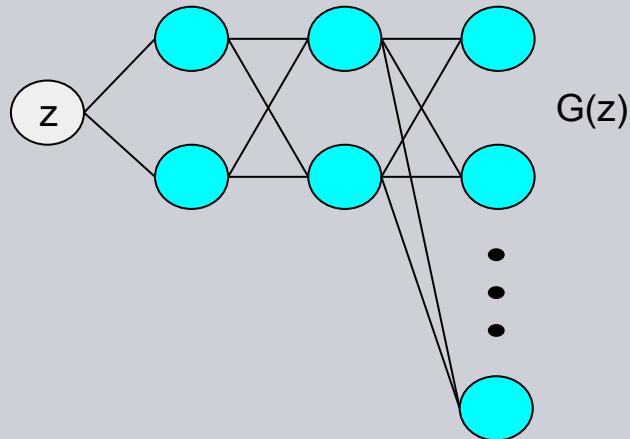


Conjunto de Datos

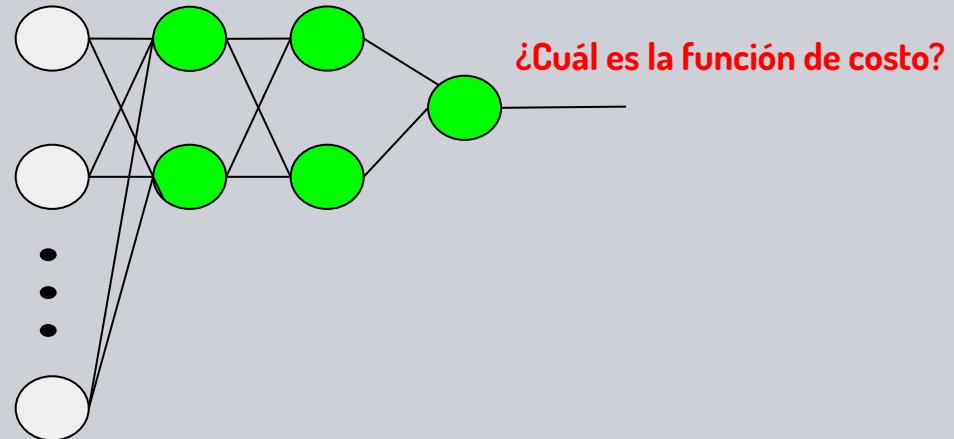
p_{data} (?)



Generador



Discriminador



Binary Cross Entropy

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N \{(y_i) \log p(y_i) + (1 - y_i) \log(1 - p(y_i))\}$$

$p(y_i)$ es la probabilidad de que el discriminador diga que la etiqueta es y_i .

FUNCIÓN DE COSTO

Binary Cross Entropy

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N \{(y_i) \log p(y_i) + (1 - y_i) \log(1 - p(y_i))\}$$

$$L_D = H(y, D(x)) = -\frac{1}{N} \sum_{i=1}^N \left[(y_i) \log(D(x_i)) + (1 - y_i) \log(1 - D(x_i)) \right]$$

Dato “real”



1

Quiero
que sea
lo más
cercano
a log 1

(1 - 1)



FUNCIÓN DE COSTO

Binary Cross Entropy

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N \{(y_i) \log p(y_i) + (1 - y_i) \log(1 - p(y_i))\}$$

$$L_D = H(y, D(x)) = -\frac{1}{N} \sum_{i=1}^N \left[(y_i) \log(D(x_i)) + (1 - y_i) \log(1 - D(x_i)) \right]$$

Dato
generado



0



(1 - 0)



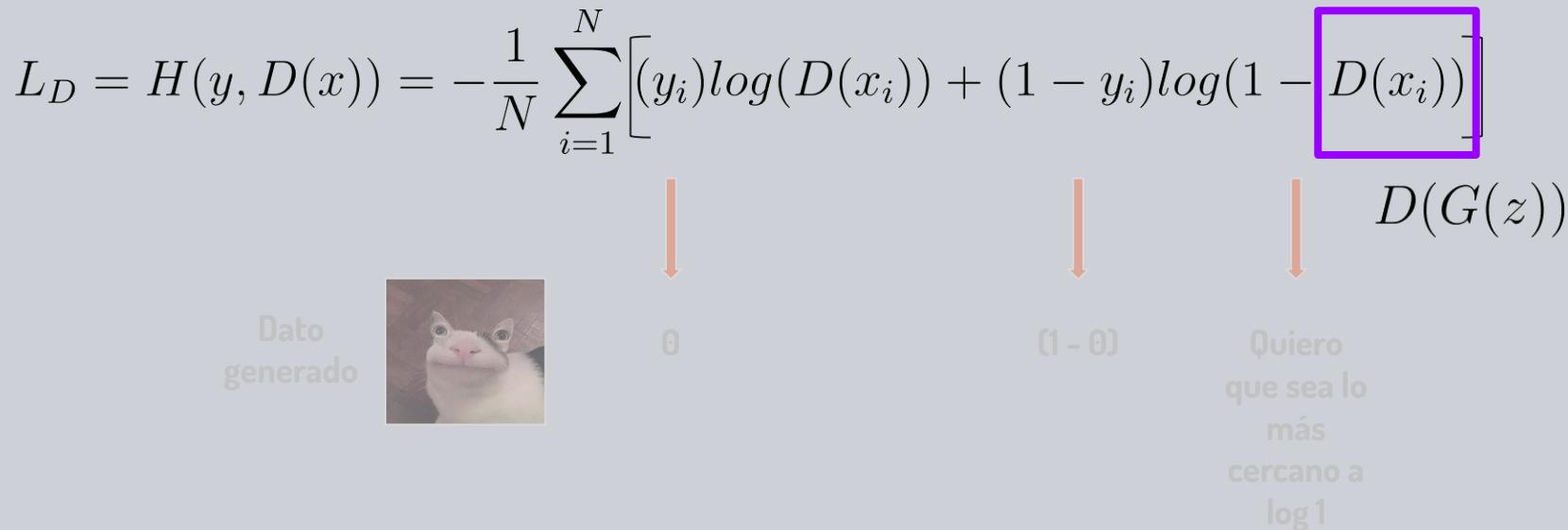
Quiero
que sea lo
más
cercano a
 $\log 1$



FUNCIÓN DE COSTO

Binary Cross Entropy

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N \{(y_i) \log p(y_i) + (1 - y_i) \log(1 - p(y_i))\}$$



JUEGO MINIMAX (dinámica adversarial)

2 jugadores

D busca maximizar y **G** busca minimizar.



$$\min_G \max_D V(D, G) = \min_G \max_D \left[\mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log (1 - D(G(z)))] \right]$$

Punto de convergencia: la salida del discriminador es 0.5 para cualquier dato de entrada (real o generado).

..... D

— $P_g(x)$

..... $P_{data}(x)$

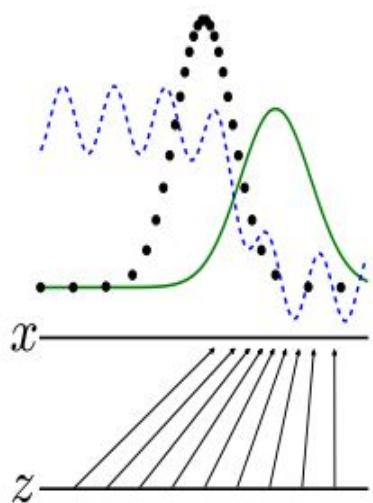
D = Discriminator

G = Generator

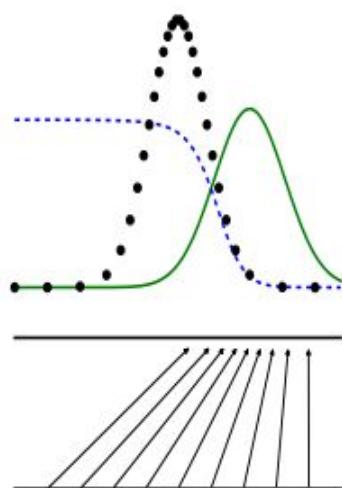
$P_z(z)$ = Input noise distribution

$P_{data}(x)$ = Original data distribution

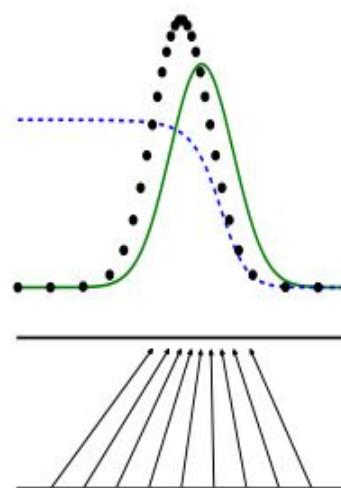
$P_g(x)$ = Generated distribution



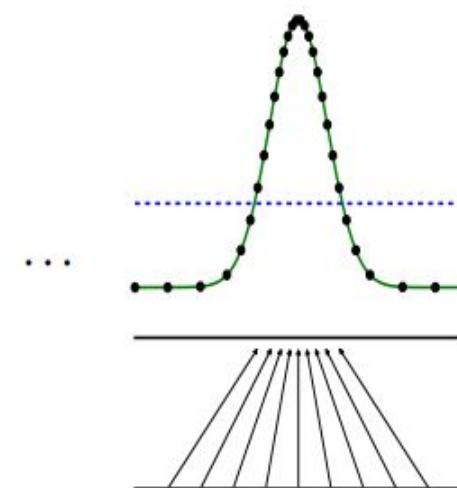
(a)



(b)



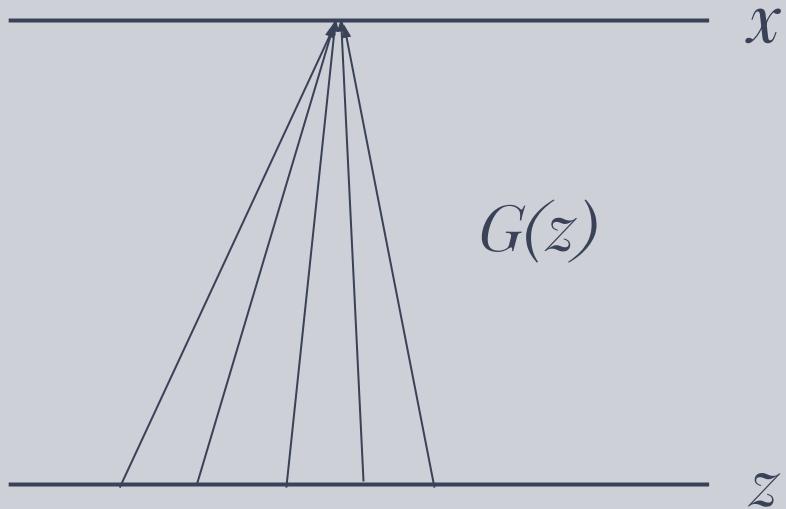
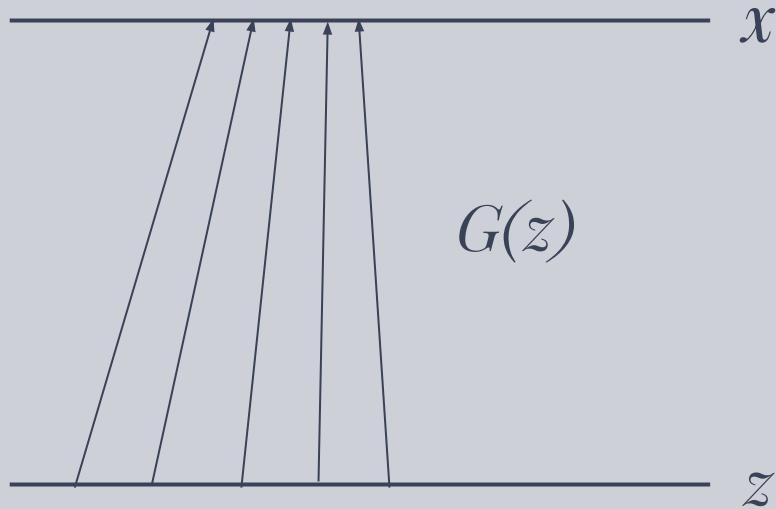
(c)



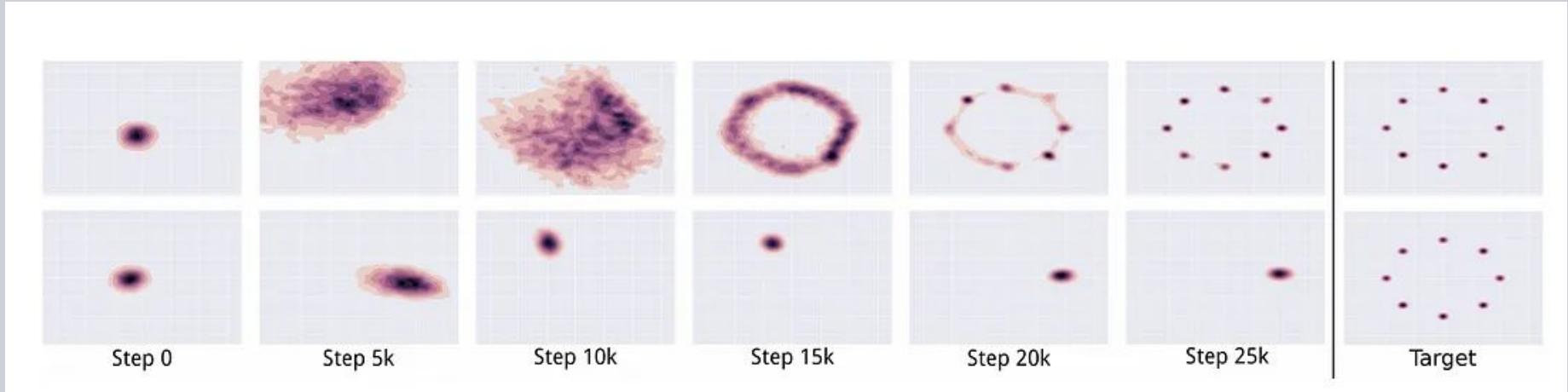
(d)

Demo práctica

Colapso Modal



Colapso Modal



Fuente: <https://arxiv.org/pdf/1611.02163.pdf>

¿Por qué $\frac{1}{2}$?

FUNCIÓN OBJETIVO

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))].$$

¿Por qué $\frac{1}{2}$?

$$\min_G \max_D V(D, G)$$

¿Cuál es el D^* que maximiza?

$$D^* = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

¿Por qué $\frac{1}{2}$?

$$\min_G \max_D V(D, G)$$

$$D^* = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

¿Qué pasa si reemplazo?

$$V(G, D^*) = -\log 4 + 2DJS(p_{data} \| p_g)$$

¿Por qué $\frac{1}{2}$?

$$D^* = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

$$\min_G \max_D V(D, G)$$

¿Cuándo es mínima esta expresión? (V)

$$V(G, D^*) = -\log 4 + 2DJS(p_{data} | p_g)$$

¿Por qué $\frac{1}{2}$?

$$D^* = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

$$\min_G \max_D V(D, G)$$

¿Cuándo es mínima esta expresión? (V)

$$V(G, D^*) = -\log 4 + 2DJS(p_{data} | p_g)$$

Constante

Cuando las distribuciones son iguales

¿Por qué $\frac{1}{2}$?

$$D^* = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

$$\min_G \max_D V(D, G)$$

D* = 0.5

¿Cuándo es mínima esta expresión? (V)

$$V(G, D^*) = -\log 4 + 2DJS(p_{data} | p_g)$$

Constante

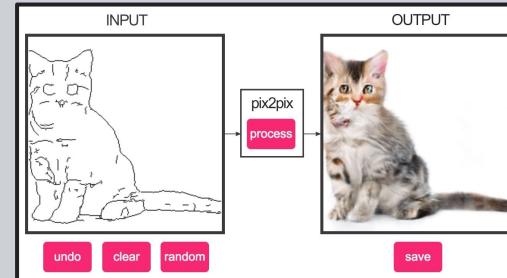
Cuando las distribuciones son iguales

RESUMEN

- El enfoque se basa en Teoría de Juegos: se modela como un problema con dos jugadores, minimax.
- Permite entrenar la red iniciando con muestras de una distribución simple (ej: uniforme).
- Para la mejora de generación de nuevos datos no se realiza comparación directa con los datos originales.
- Difícil de entrenar y esta es la **MAYOR** desventaja. Requiere atención manual para analizar las salidas de las imágenes.

VARIANTES: <https://github.com/hindupuravinash/the-gan-zoo>

- Conditional GAN
- Deep Convolutional GAN (DCGAN)
- Progressive Growing GAN (PGGAN)
- StyleGan
- ...



APLICACIONES

- Traducción texto a imagen (ej: DCGAN)
(Lab. <https://creator.nightcafe.studio/>)
- Traducción bordes a imagen (Pix2Pix)
- Generación de imágenes (especialmente importante en el área médica)
- ...

Caso de Aplicación

Proyecto final de Grado: Use of generative adversarial networks for the creation and manipulation of facial images in the context of studying false memories and its effects on wrongful conviction cases: implementation of StyleGAN's generative image modeling and style mixing properties to design an interface for experimentation purposes



Resumen

[LINK](#)

Proyecto final de Grado: Controlling face's frame generation in StyleGAN's latent space operations: modifying faces to deceive our memory



Resumen

"Innocence Project is a non-profitable organization that works in reducing wrongful convictions. In collaboration with El Laboratorio de Sueño y Memoria from Instituto Tecnológico de Buenos Aires (ITBA), they are studying human memory in the context of face identification. They have a strong hypothesis stating that human memory heavily relies in face's frame to recognize faces. If this is proved, it could mean that face recognition in police li-

[LINK](#)

Lección Importante!

GAN modela un formato o estrategia de entrenamiento no es una arquitectura de red neuronal

Esquemas que pueden reemplazar a GAN:

- Modelos de difusión - comúnmente utilizados para imágenes
- Modelos autorregresivos (próxima clase!) - comúnmente utilizados para texto.

The End

GAN Lab: <https://poloclub.github.io/ganlab/>

Paper original Generative Adversarial Networks: <https://arxiv.org/pdf/1406.2661.pdf>

Sobre Binary Cross Entropy:

<https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a>

Optimización y Convergencia de la red GAN:

<https://arxiv.org/pdf/1406.2661.pdf>

https://www.youtube.com/watch?v=Gib_kiXgnvA (favorito personal)

<https://srome.github.io/An-Annotated-Proof-of-Generative-Adversarial-Networks-with-Implementation-Notes/>

<https://jonathan-hui.medium.com/proof-gan-optimal-point-658116a236fb>

<https://arxiv.org/pdf/1705.07215.pdf>

Sobre el origen de la idea:

<https://www.technologyreview.com/2018/02/21/145289/the-ganfather-the-man-whos-given-machines-the-gift-of-imagination/>

<https://www.deeplearning.ai/the-batch/ian-goodfellow-a-man-a-plan-a-gan/>