

10

REDES AUTOCODIFICADORAS E REDES ADVERSÁRIAS GENERATIVAS

APRENDIZAGEM PROFUNDA

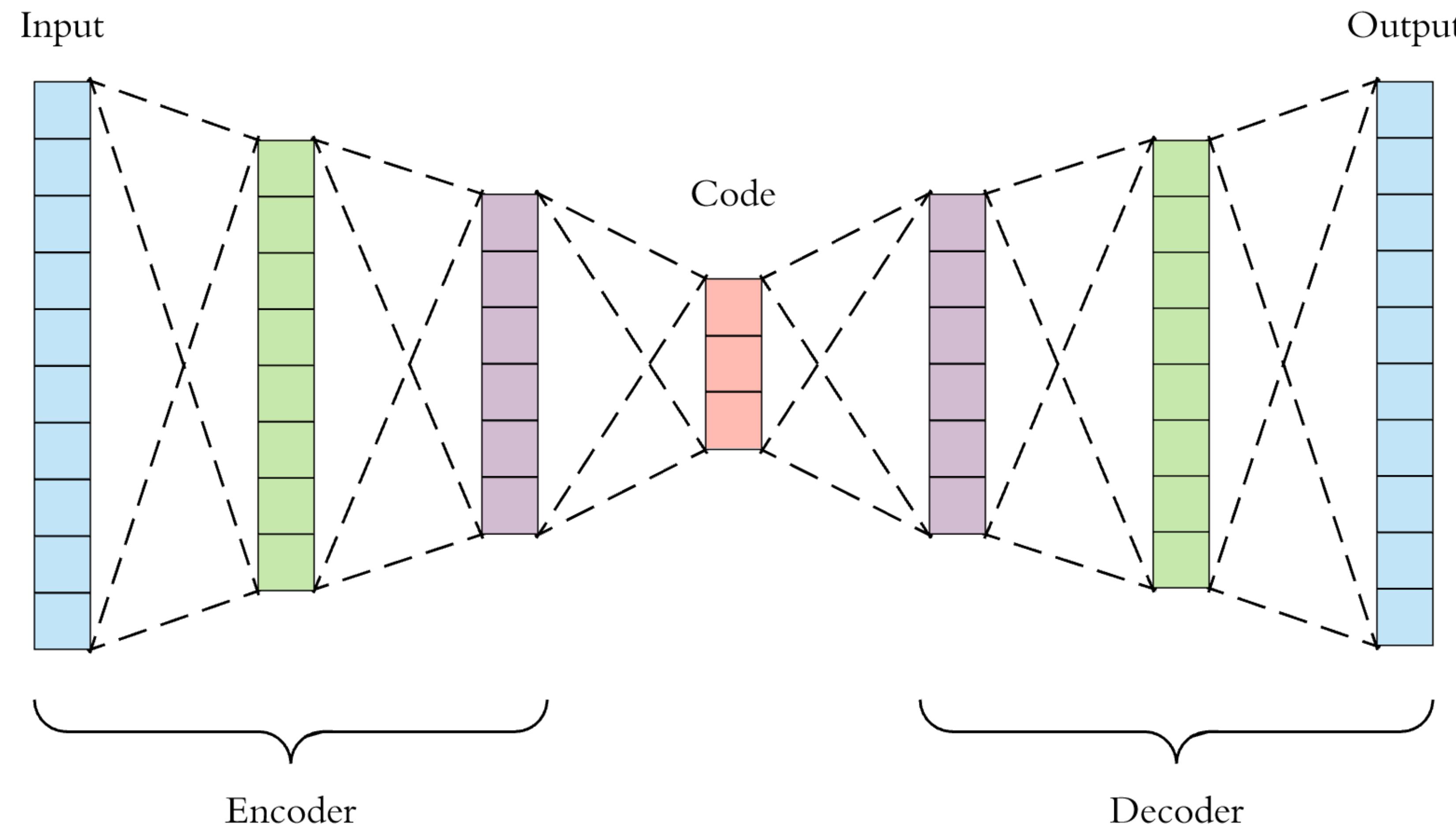
PPGCC – 2023.1

Prof. Saulo Oliveira <saulo.oliveira@ifce.edu.br>

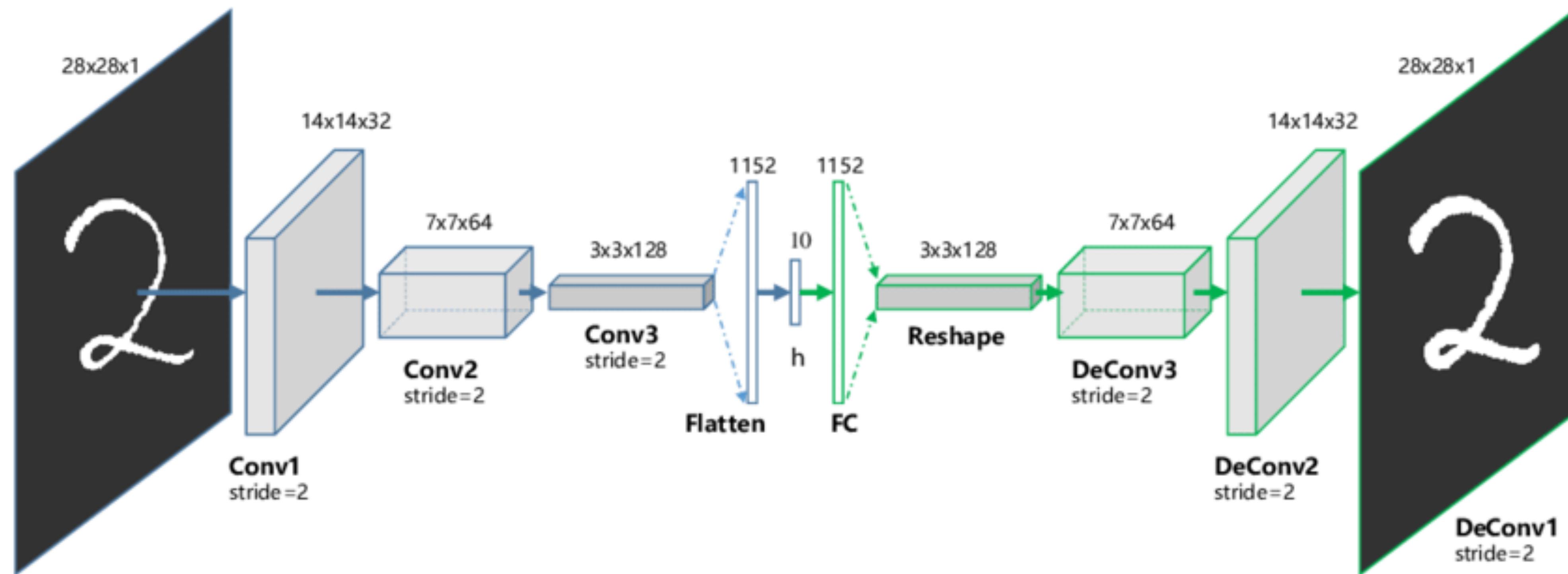


APRENDIZADO NÃO- SUPERVISIONADO

Auto-supervisionado

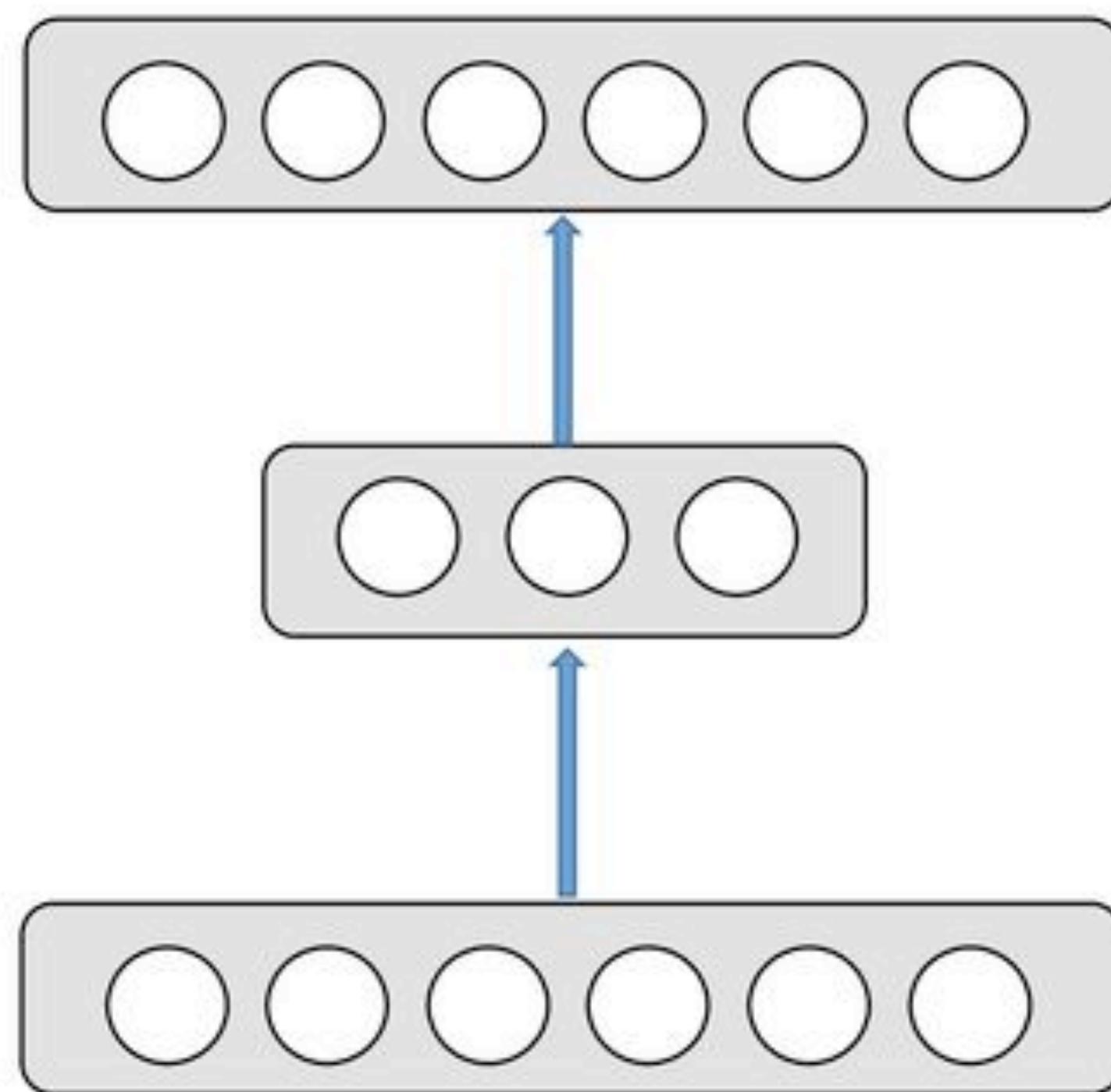


Auto-supervisionado

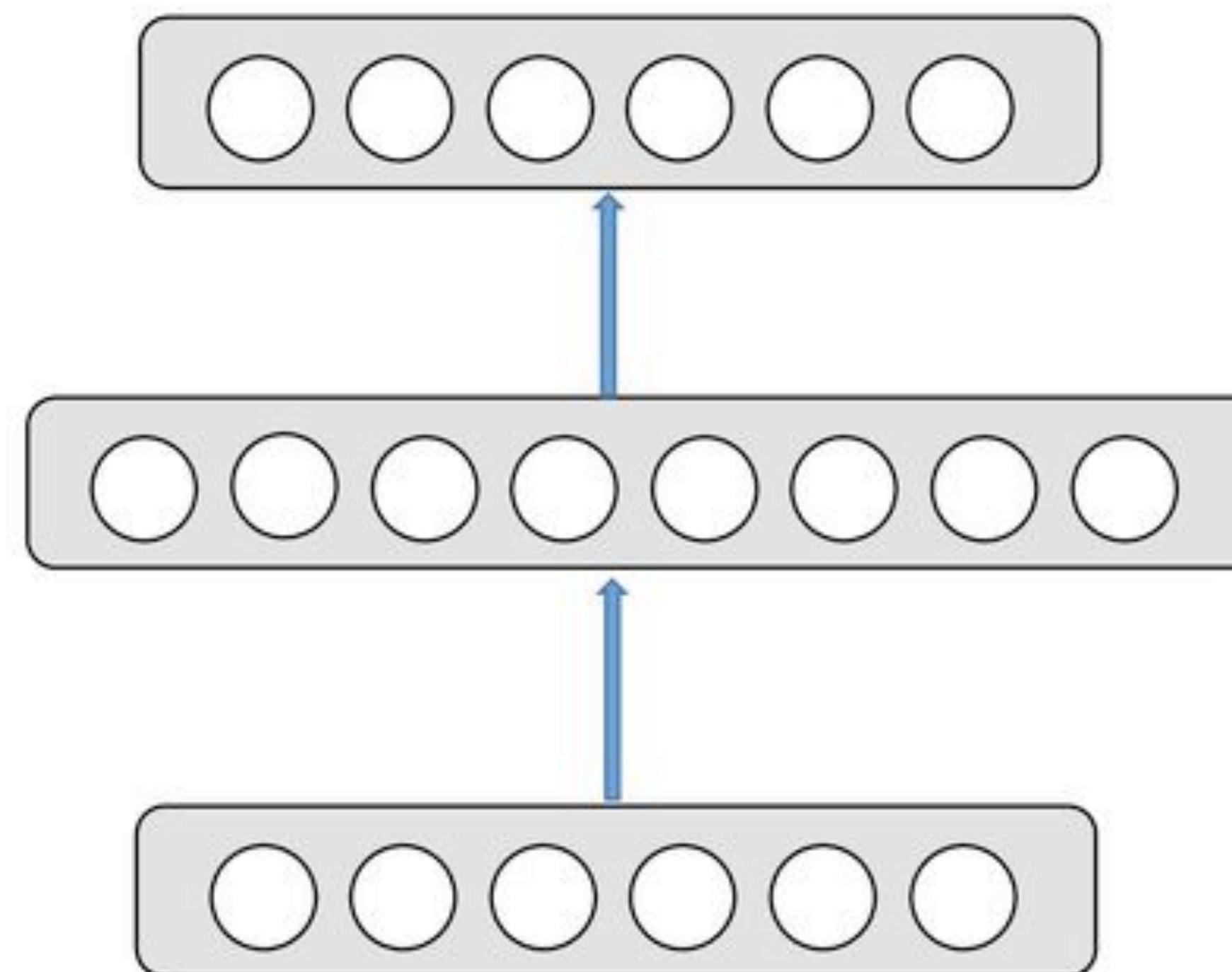


Auto-supervisionado

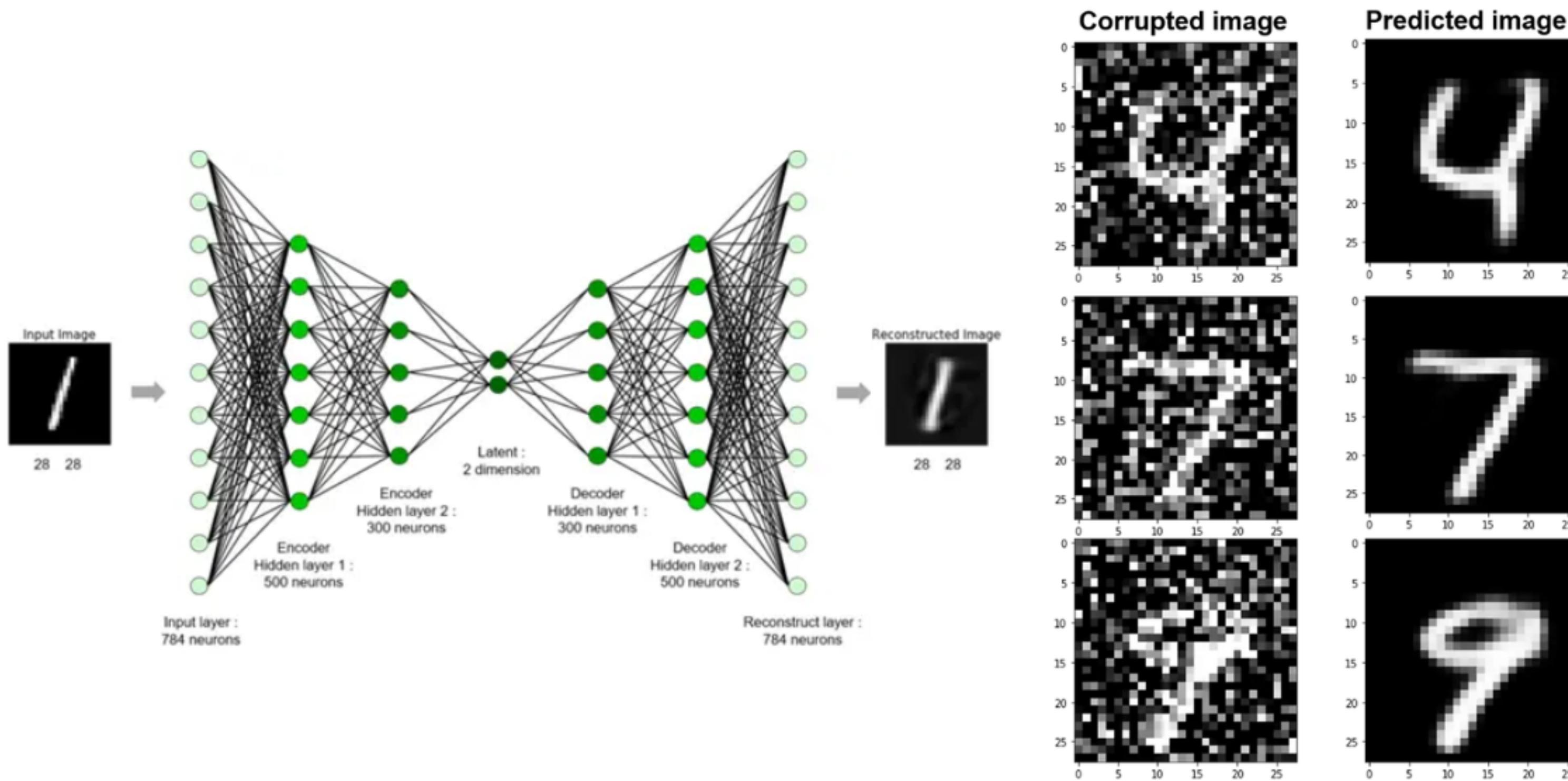
Subcompleto



Sobrecompleto

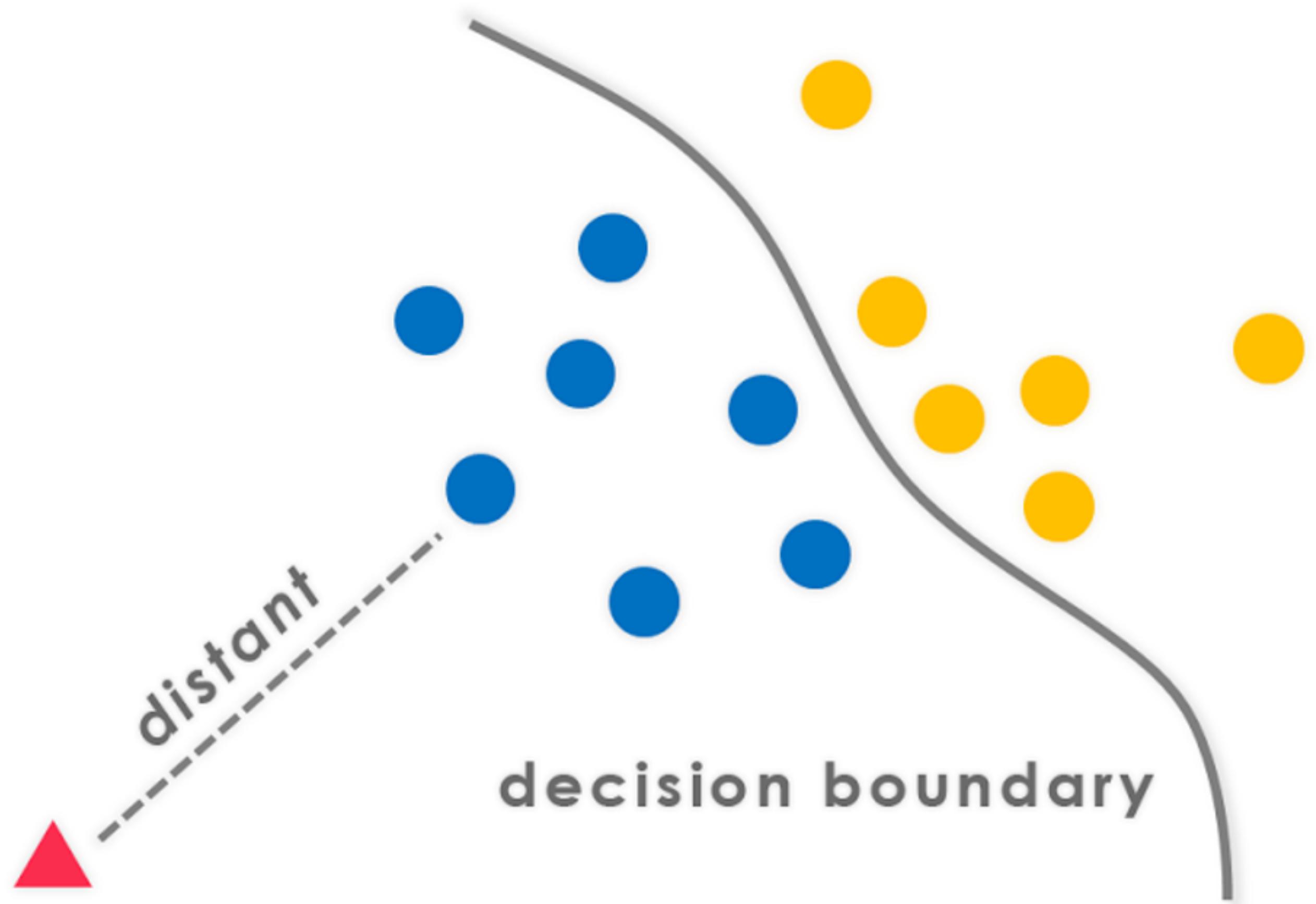


Autocodificador Remove-ruído

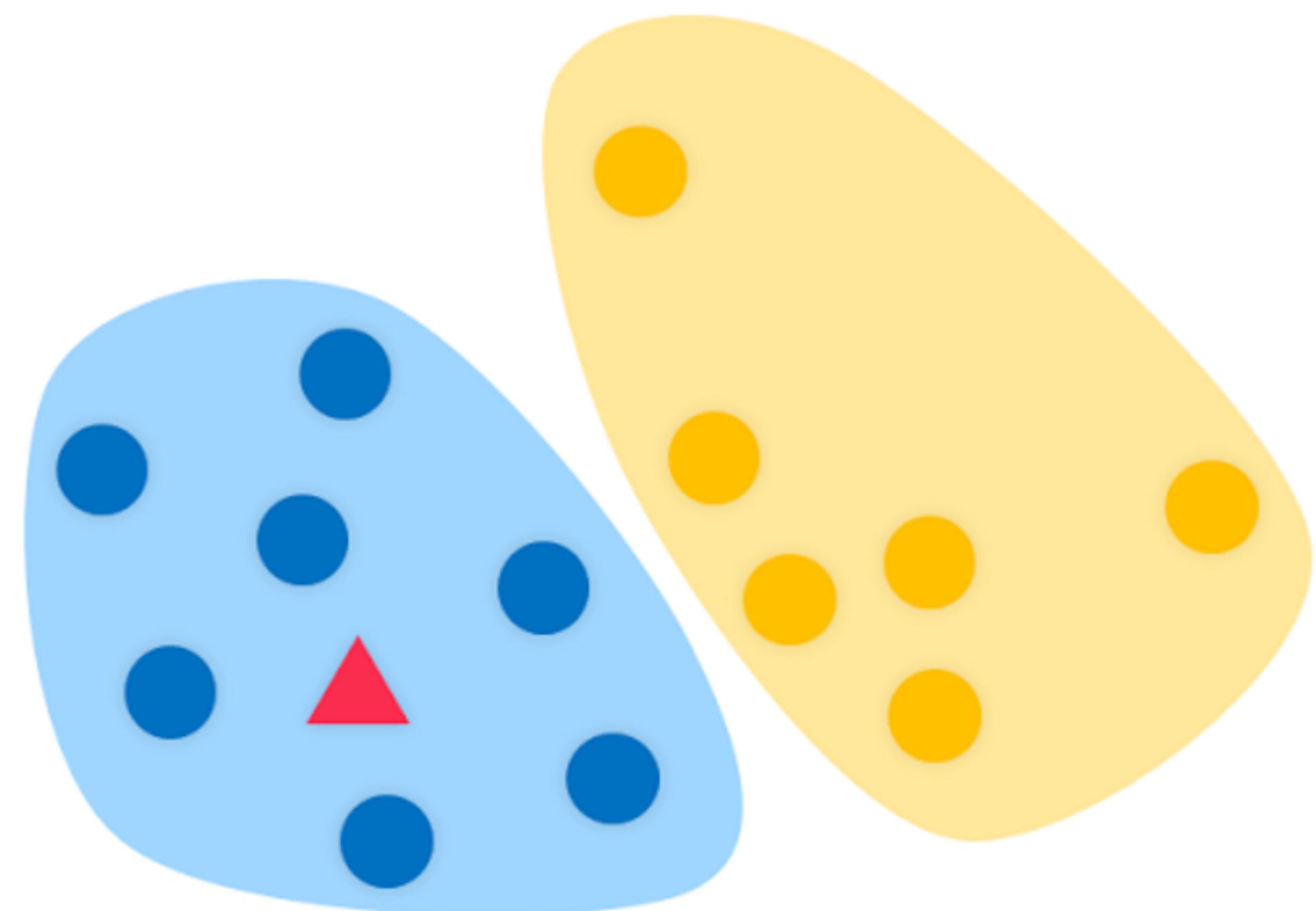


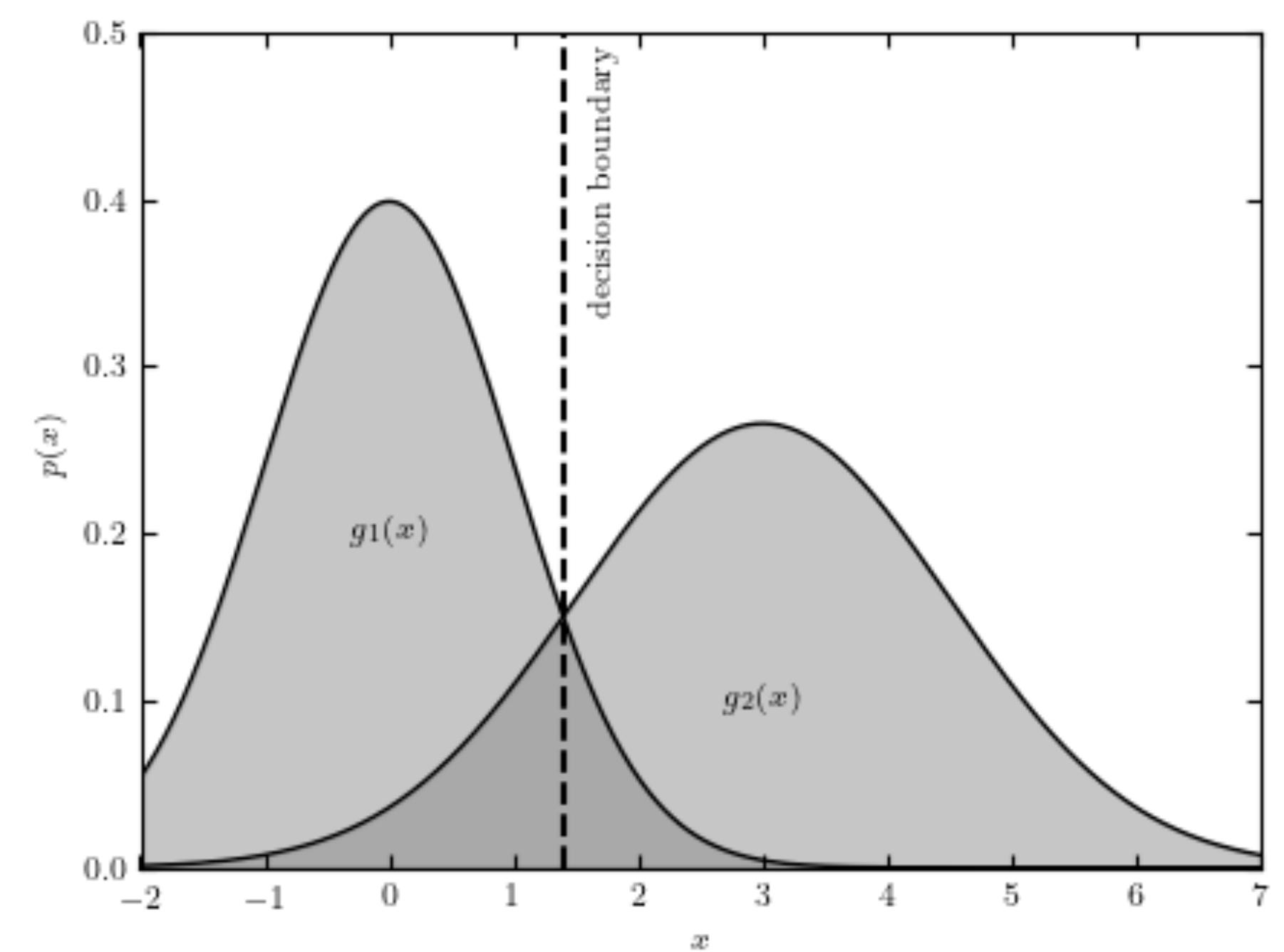
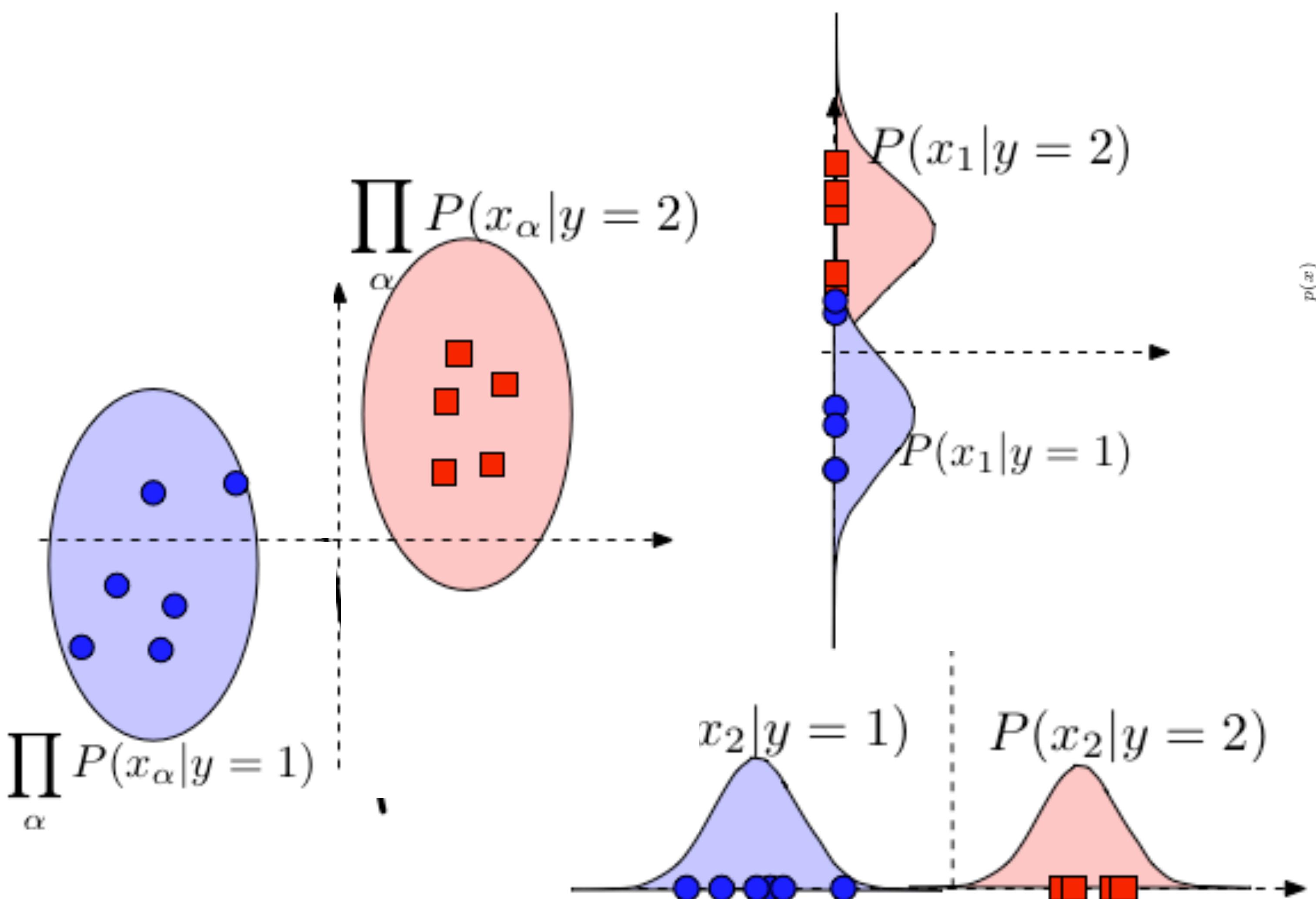
MODELO DISCRIMINATIVO VS GENERATIVO

Discriminative



Generative





$$g(\mathbf{x}|y=1) = g(\mathbf{x}|y=2)$$

$$g(\mathbf{x}|y=1) - g(\mathbf{x}|y=2) = 0$$

Probabilidade condicional

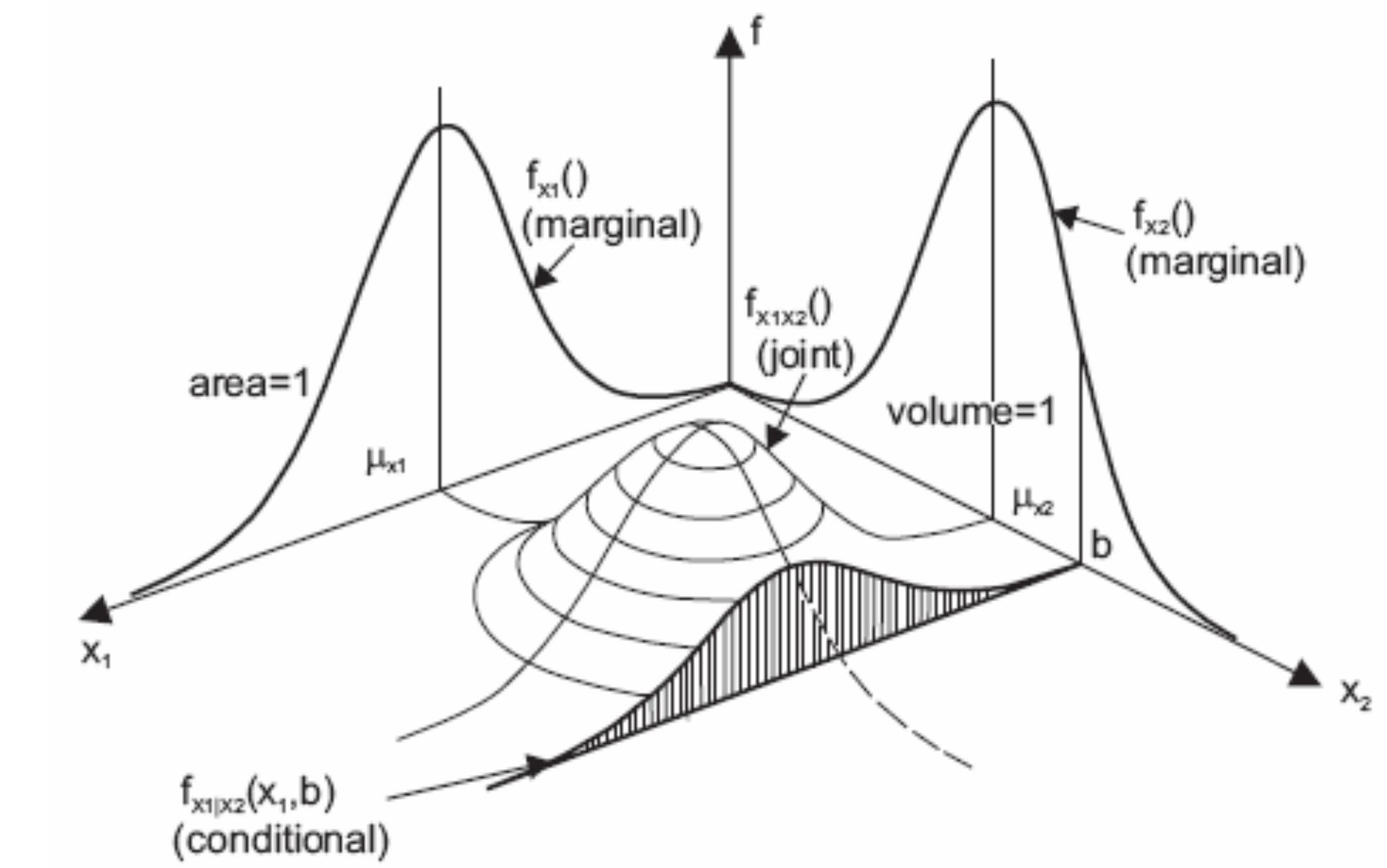
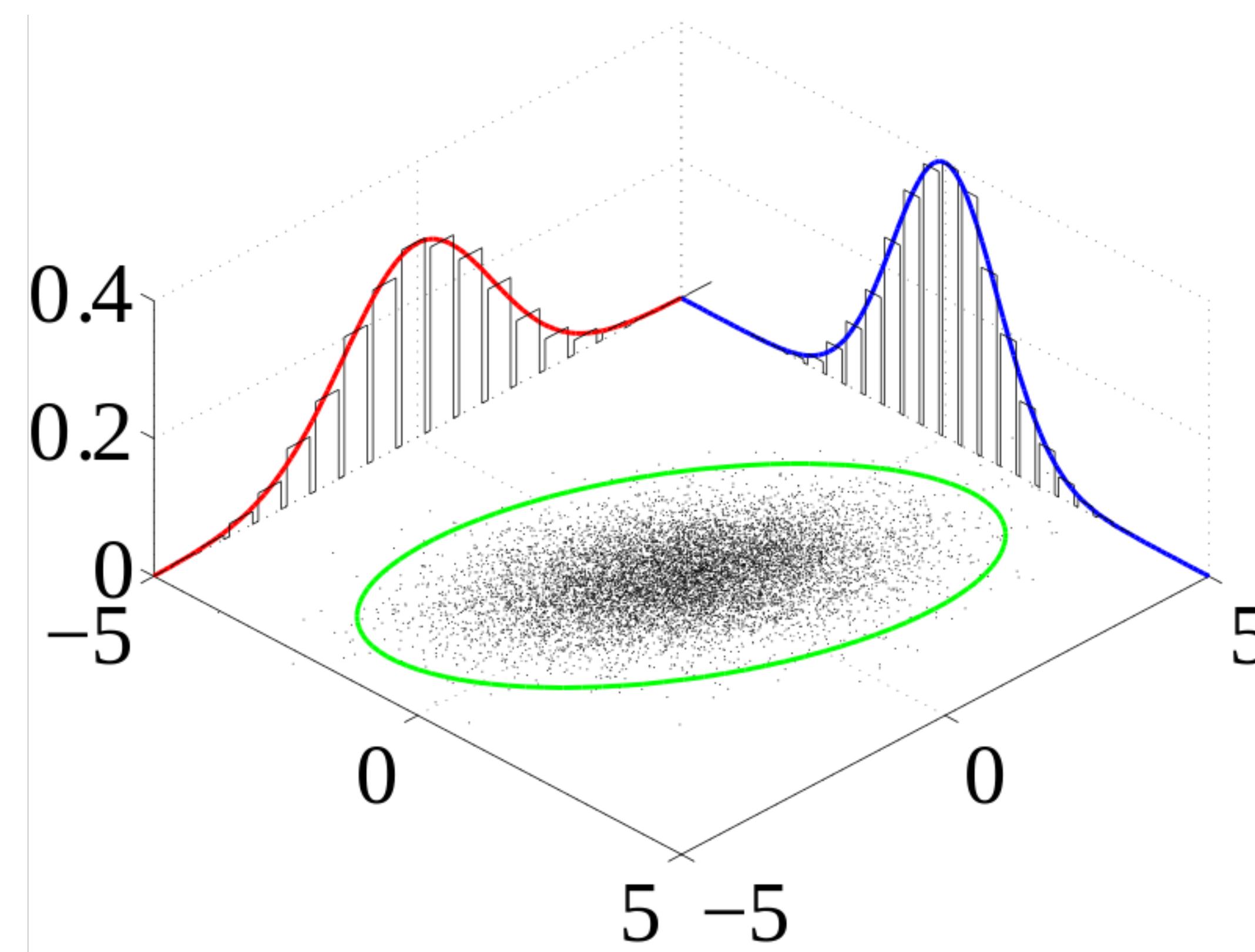
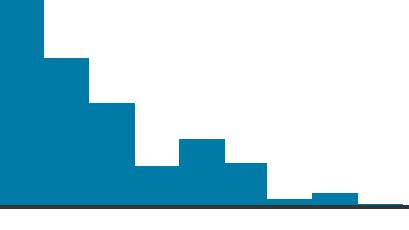
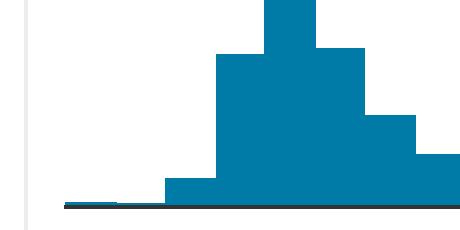
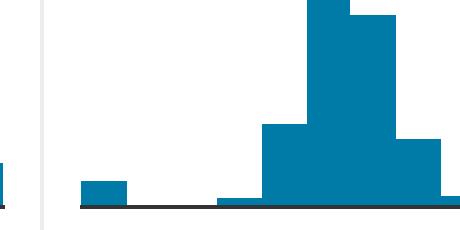
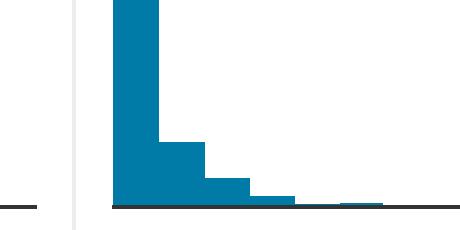
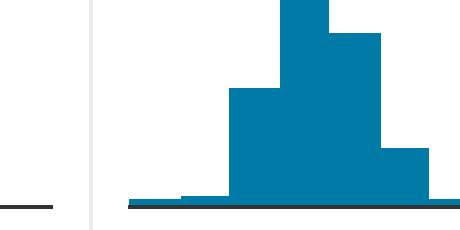
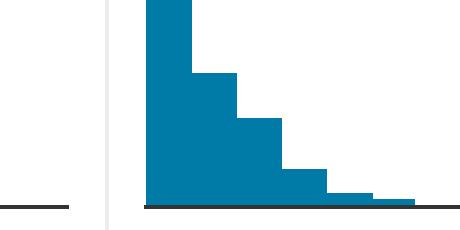
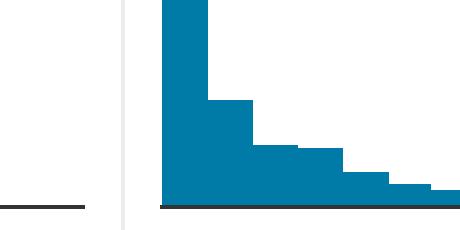


Figure 2.13-Joint and marginal probability density function (PDF), from Melchers (1999).

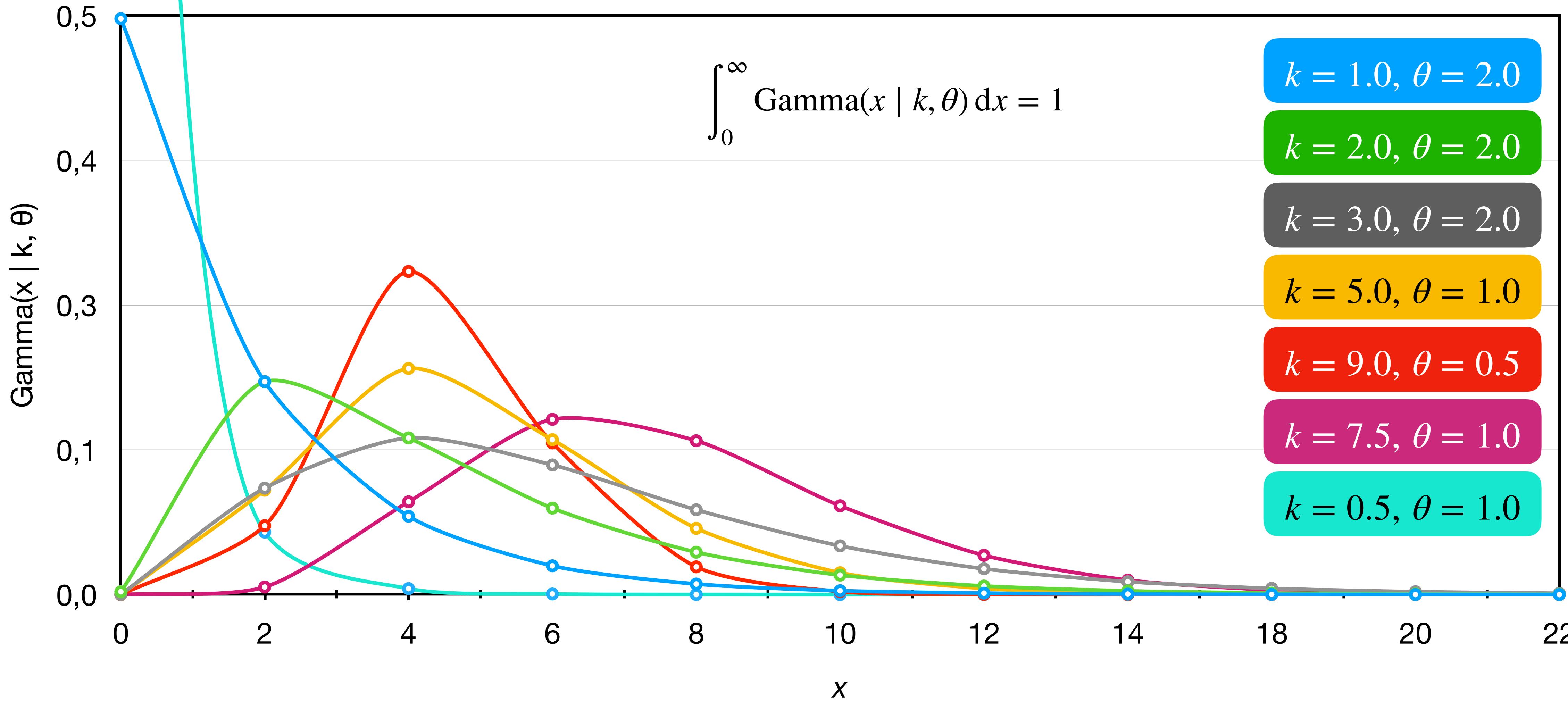
DISTRIBUIÇÃO DOS ATRIBUTOS

[Detail](#) [Compact](#) [Column](#)9 of 9 columns **About this file**

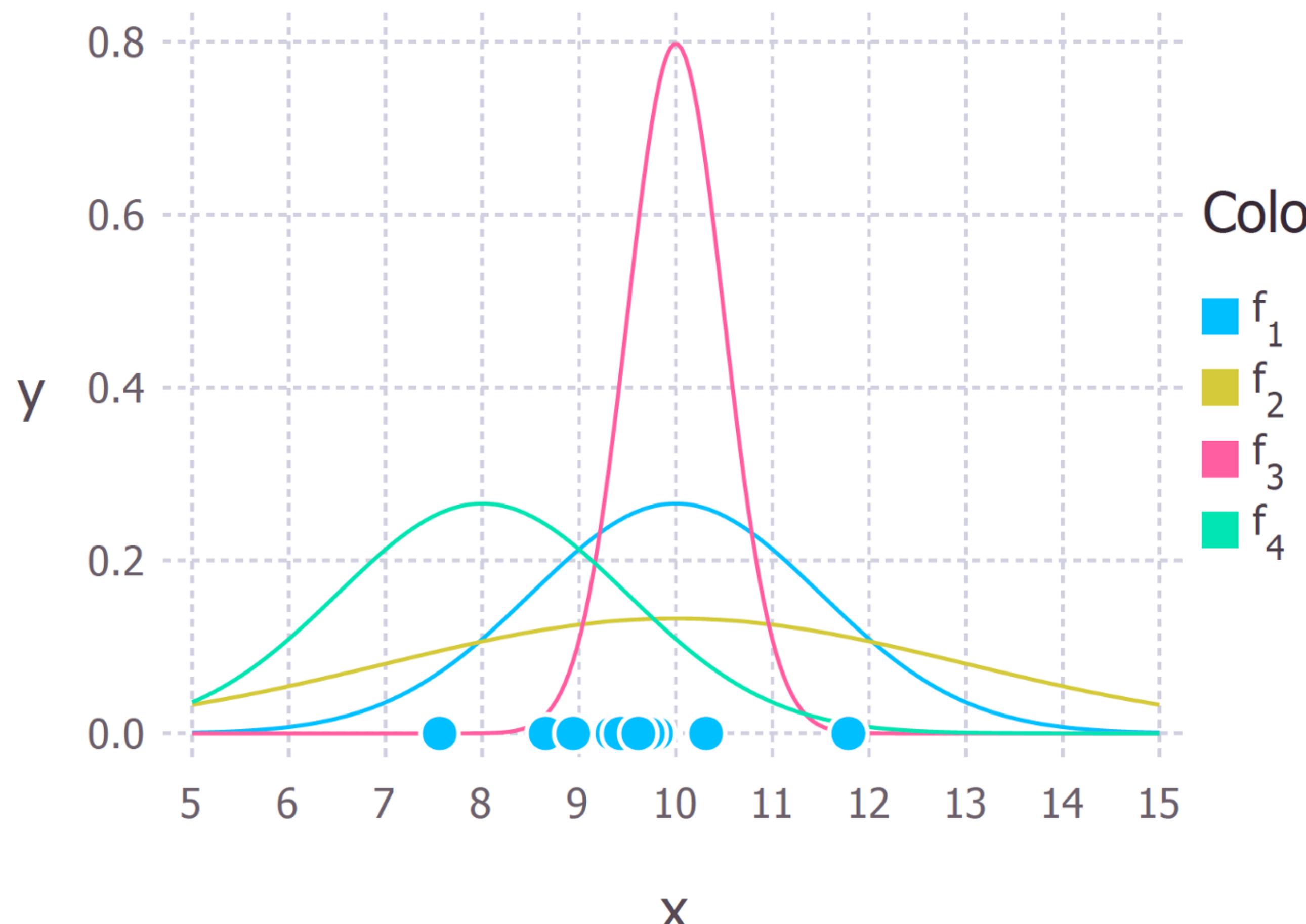
The datasets consist of several medical predictor (independent) variables and one target (dependent) variable, Outcome. Independent variables include the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

# Pregnancies	# Glucose	# BloodPressure	# SkinThickness	# Insulin	# BMI	# DiabetesPedigree...	# Age	# Outcome
Number of times pregnant	Plasma glucose concentration a 2 hours in an oral glucose tolerance test	Diastolic blood pressure (mm Hg)	Triceps skin fold thickness (mm)	2-Hour serum insulin (mu U/ml)	Body mass index (weight in kg/(height in m)^2)	Diabetes pedigree function	Age (years)	Class variable (0 or 1) 268 of 768 are 1, the others are 0
 17	 199	 122	 99	 846	 67.1	 2.42	 81	 1
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0

A Distribuição Gamma



Máxima verossimilhança



- Achar o conjunto de parâmetros que maximize a função de verossimilhança, i.e.,

$$\theta^* = \arg \max_{\theta \in \Theta} \mathcal{L}(\mathbf{X}, \theta)$$

- Achar o conjunto de parâmetros mais prováveis segundo os dados observados, i.e.,

$$\mathcal{L}(\mathbf{X}, \theta) = \prod_{i=1}^N f_n(\mathbf{x}_i | \theta)$$

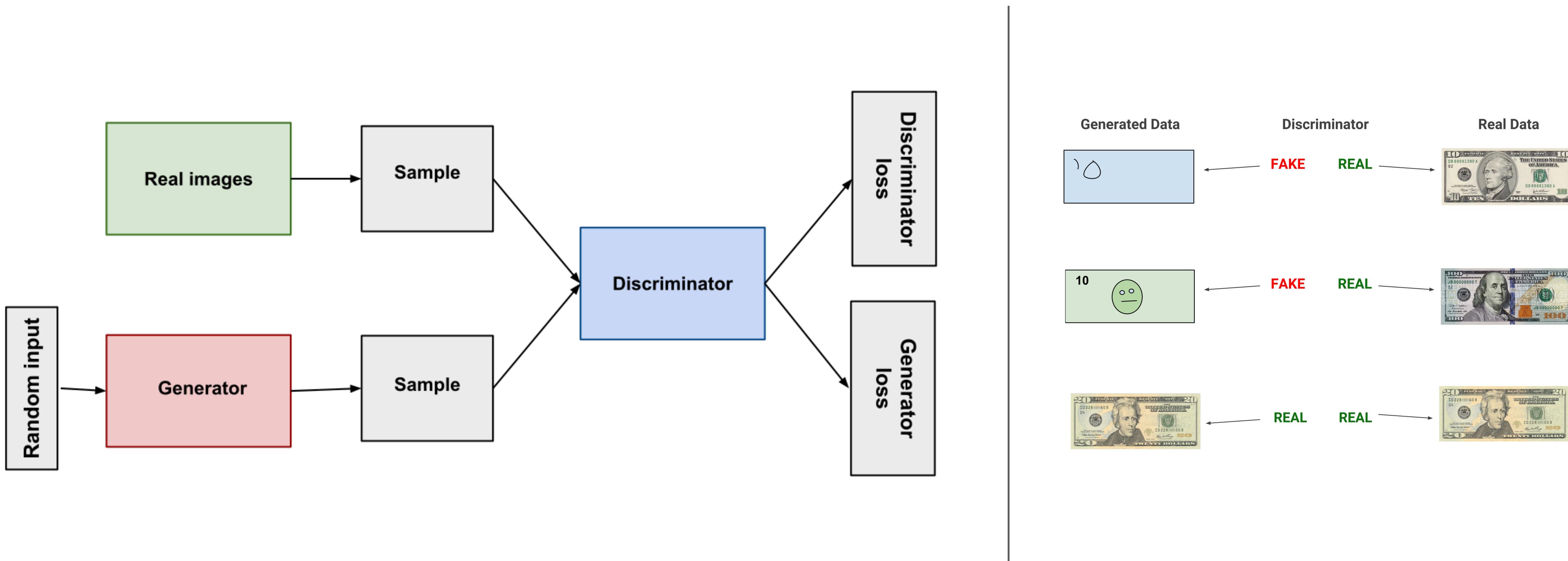
- Na prática, muitas vezes é conveniente trabalhar com o logaritmo natural da função de verossimilhança, i.e.,

$$\ell(\mathbf{X}, \theta) = \log \mathcal{L}(\mathbf{X}, \theta)$$

$$\ell(\mathbf{X}, \theta) = \frac{1}{N} \sum_{i=1}^N \log f_n(\mathbf{x}_i | \theta)$$

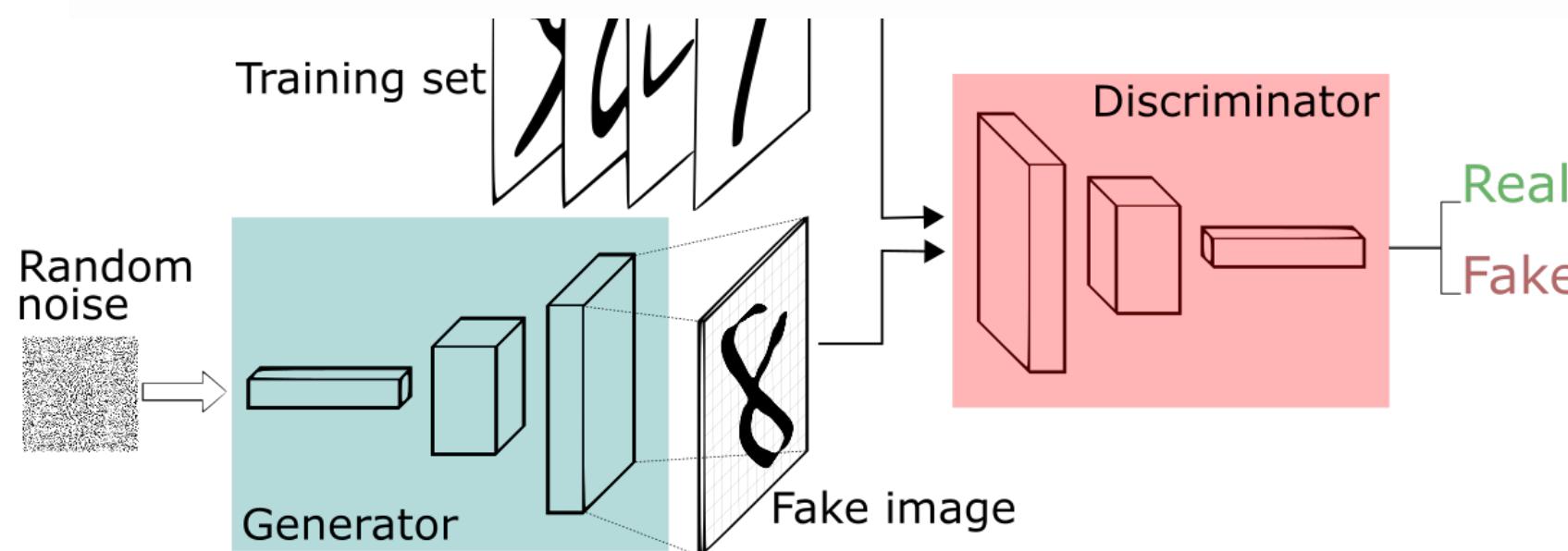
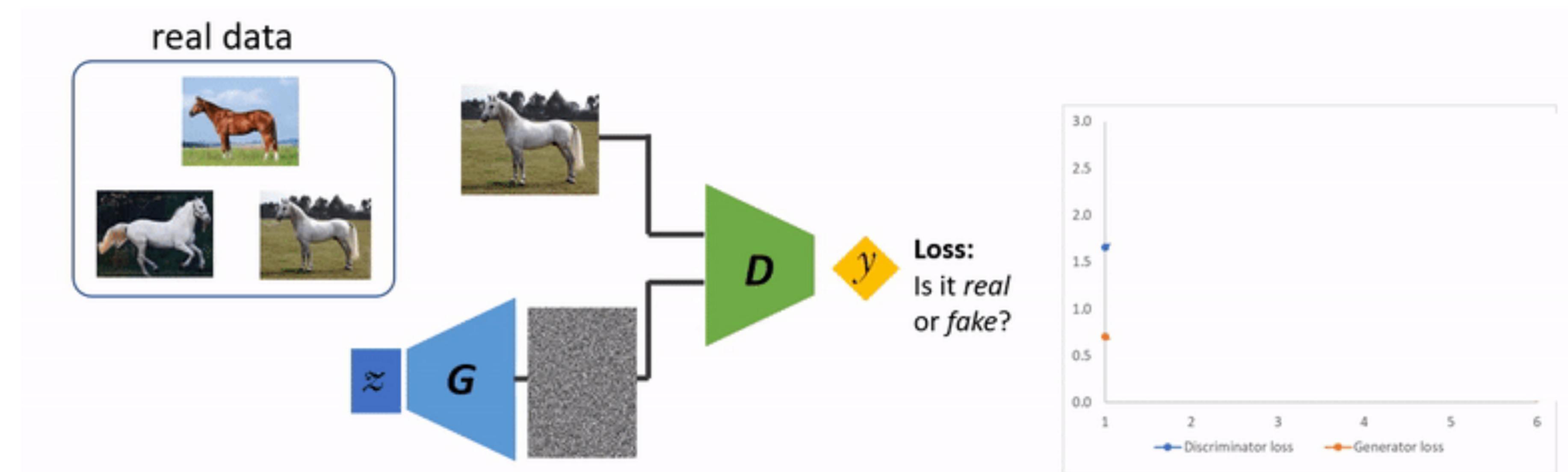
REDES ADVERSÁRIAS GENERATIVAS

VISÃO GERAL DO MODELO



https://developers.google.com/machine-learning/gan/gan_structure

FUNCIONANDO NA PRÁTICA



[https://medium.com/@jamaltoutouh/
lipizzaner-a-framework-for-co-evolutionary-
distributed-gan-training-7a725fb17e49](https://medium.com/@jamaltoutouh/lipizzaner-a-framework-for-co-evolutionary-distributed-gan-training-7a725fb17e49)

TREINAMENTO DE UMA GAN

TREINAMENTO DE UMA GAN

Como uma GAN contém duas redes treinadas separadamente, seu algoritmo de treinamento deve abordar duas complicações:

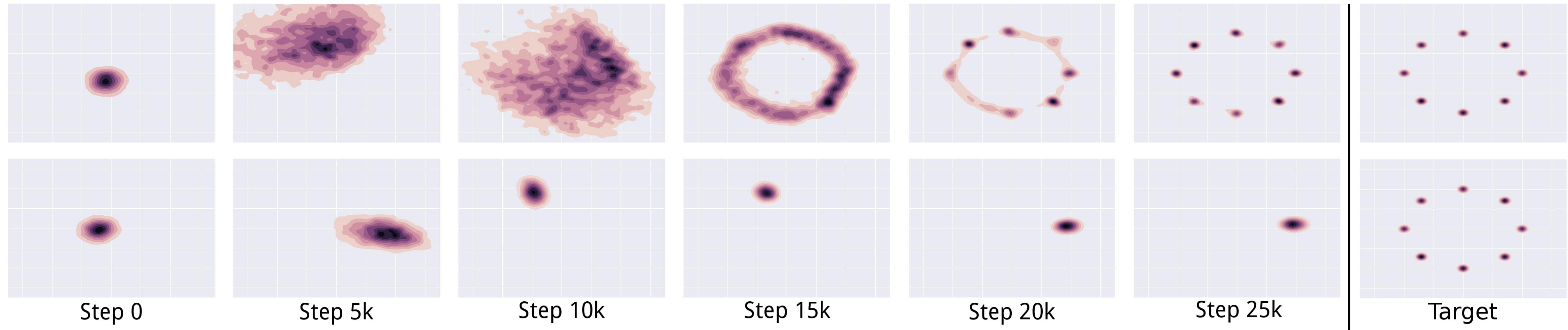
- As GANs devem conciliar dois tipos diferentes de treinamento (gerador e discriminador);
- A convergência GAN é difícil de identificar.

Treinamento alternado

O gerador e o discriminador possuem processos de treinamento diferentes. Então, como treinamos o GAN como um todo? O treinamento GAN prossegue em períodos alternados:

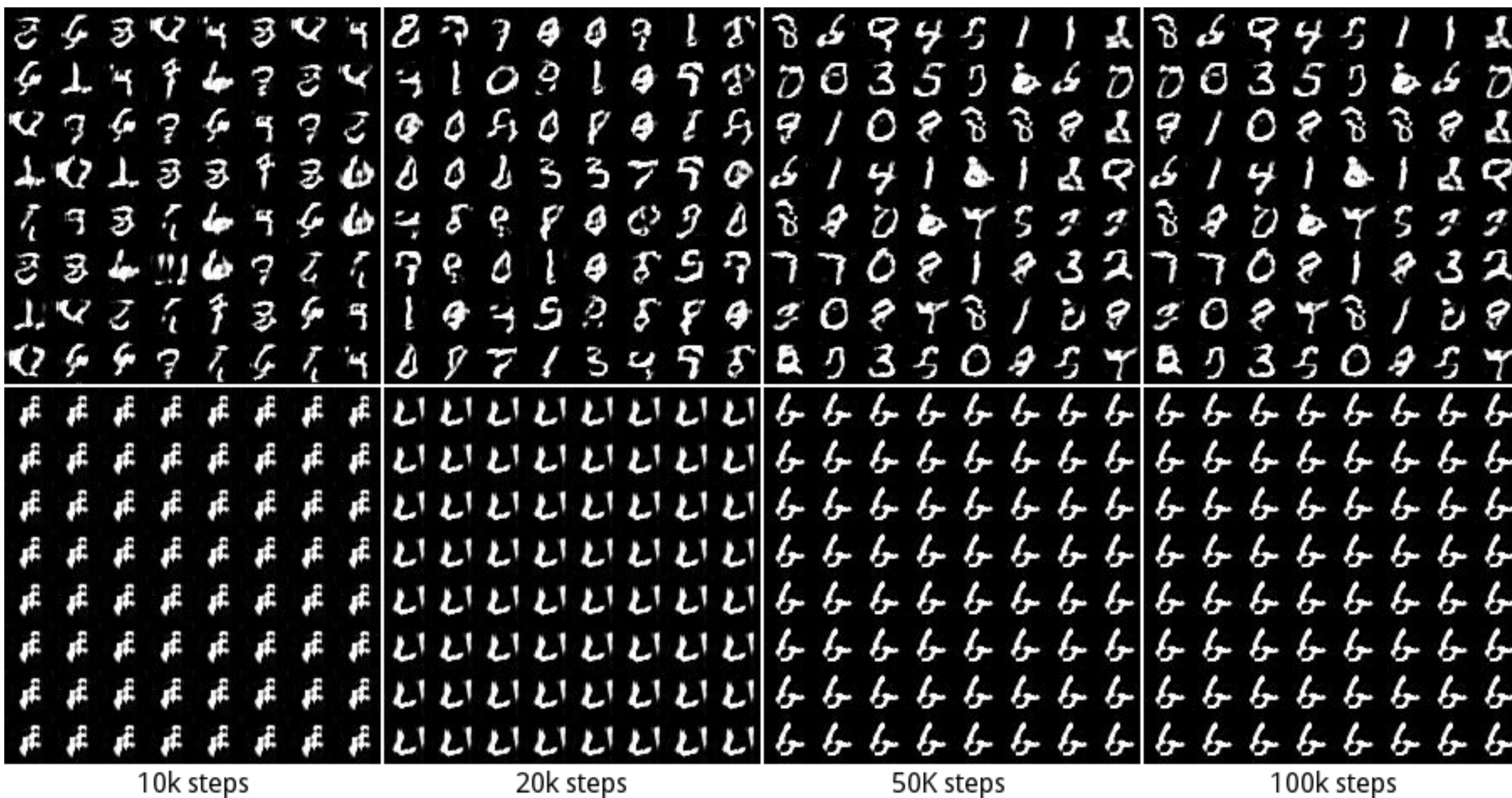
1. O discriminador treina para uma ou mais épocas.
2. O gerador treina por uma ou mais épocas.
3. Repita os passos 1 e 2 para continuar a treinar as redes geradora e discriminadora;
4. Pare quando o discriminador tiver aproximadamente 50% de acurácia.

CONVERGÊNCIA...KKKK



Modo Colapso

CONVERGÊNCIA...KKKK

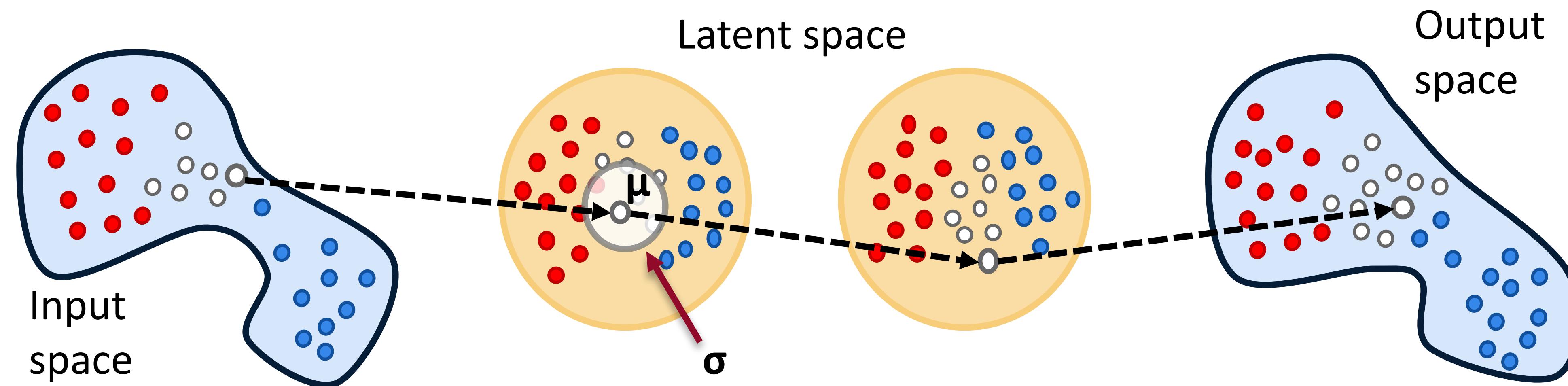
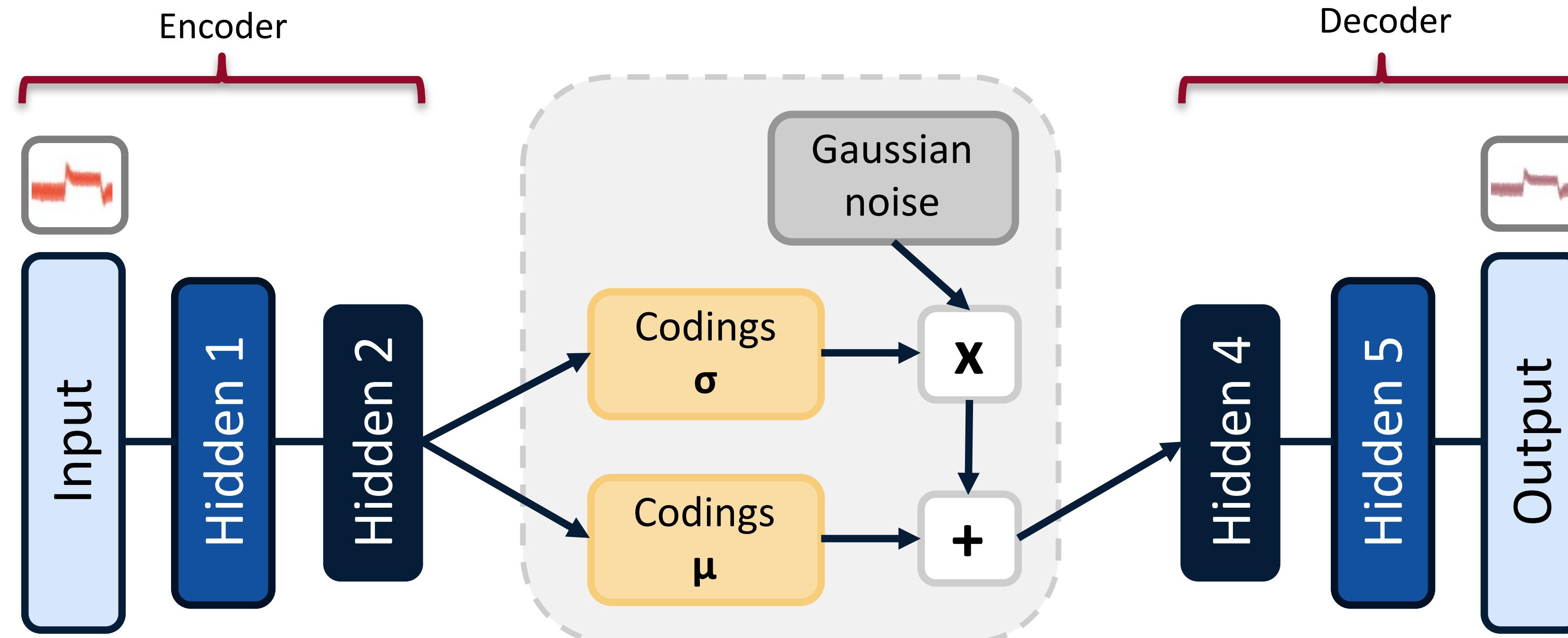


AUTOCODIFICADORES VARIACIONAIS

Por que variacional?

- Um dos principais problemas de estatísticas modernas e aprendizado de máquina é aproximar distribuições de probabilidade difíceis de calcular. Esse problema é especialmente importante na modelagem probabilística, que enquadra toda a inferência sobre quantidades desconhecidas como um cálculo sobre uma distribuição condicional.
- Inferência variacional é um método que se aproxima das distribuições de probabilidade por meio da otimização e tem sido usado em uma miríade de aplicativos no aprendizado de máquina e tende a ser mais rápido do que os métodos mais tradicionais, como a amostragem de Monte Carlo da cadeia Markov.
- Para entradas válidas, o AE é capaz de comprimi-las, eliminando essencialmente a redundância (codificador). No entanto, devido ao espaço latente não regularizado do AE, **o decodificador não pode ser usado para gerar dados de entrada válidos de vetores amostrados de o espaço latente.**

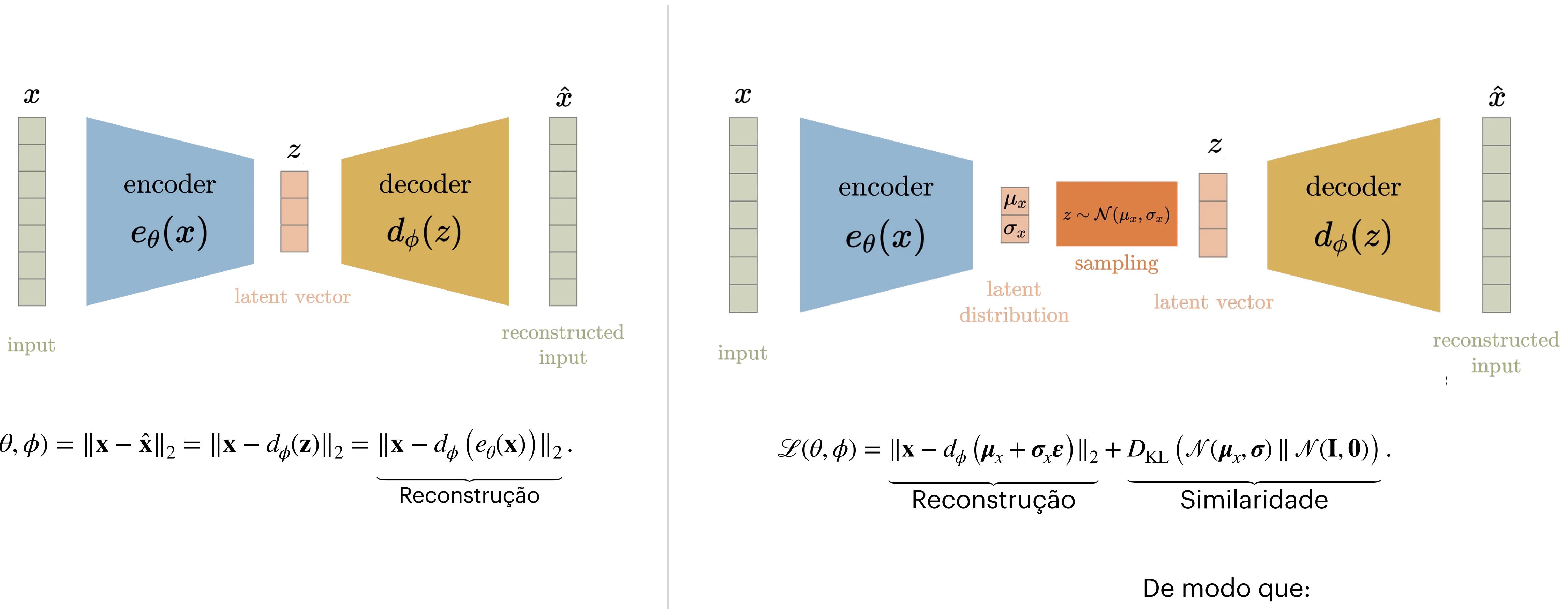
<https://learn.microsoft.com/pt-br/events/neural-information-processing-systems-conference-nips-2016/variational-inference-foundations-modern-methods>



Principais diferenças

AUTOCOFICADORA	AUTOCODIFICADORA VARACIONAL
Usado para gerar uma transformação comprimida (geralmente) de entrada em um espaço latente.	Impõe condições na variável latente para ter norma unitária.
A variável latente não está regularizada .	Espaço latente regularizado .
Escolher uma variável latente aleatória irá gerar uma saída de lixo .	Um valor aleatório da variável latente gera uma saída significativa no decodificador.
A variável latente tem descontinuidade .	A variável latente é suave e contínua .
Variáveis latentes são valores determinísticos .	A entrada do decodificador é estocástica e é amostrada de uma distribuição Normal com média e variância da saída do codificador.
O espaço latente carence da capacidade generativa.	O espaço latente possui capacidades generativas.
	A variável latente na forma comprimida é média e variância.

Principais diferenças



De modo que:

- $\mu_x, \sigma_x = e_\theta(x)$.
- $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

REFERÊNCIAS

- Aqeel Anwar. **Difference between AutoEncoder (AE) and Variational AutoEncoder (VAE)**. <https://towardsdatascience.com/difference-between-autoencoder-ae-and-variational-autoencoder-vae-ed7be1c038f2>. 2021., Acessado em 20 maio de 2023.
- BISHOP, Christopher M.; NASRABADI, Nasser M. **Pattern recognition and machine learning**. New York: springer, 2006.
- Deep Learning Book. **A Matemática dos Variational Autoencoders (VAEs)**. <https://www.deeplearningbook.com.br/a-matematica-dos-variational-autoencoders-vaes/>. 2022., Acessado em 20 maio de 2023.
- Fernando J. Von Zuben. **Notas de Aula: Redes Neurais Artificiais**. DCA/FEEC/Unicamp, 2014. Disponível em: https://www.dca.fee.unicamp.br/~vonzuben/ia013_1s14/notas_de_aula/topico2_IA013_1s2014_Parte2.pdf. Acessado em: 14 de maio de 2022.
- Google. **Overview of GAN Structure**. https://developers.google.com/machine-learning/gan/gan_structure. 2021, Accessed on Mar 2021.
- Richard O. Duda, Peter E. Hart, David G. Stork. **Pattern Classification**. John Wiley & Sons, 2012.
- Thalles Silva. **A Short Introduction to Generative Adversarial Networks**. <https://sthalles.github.io/intro-to-gans/>. 2017, Accessed on Mar 2021.