Mondragon Unibertsitatea

Faculty of Engineering

Deep Learning

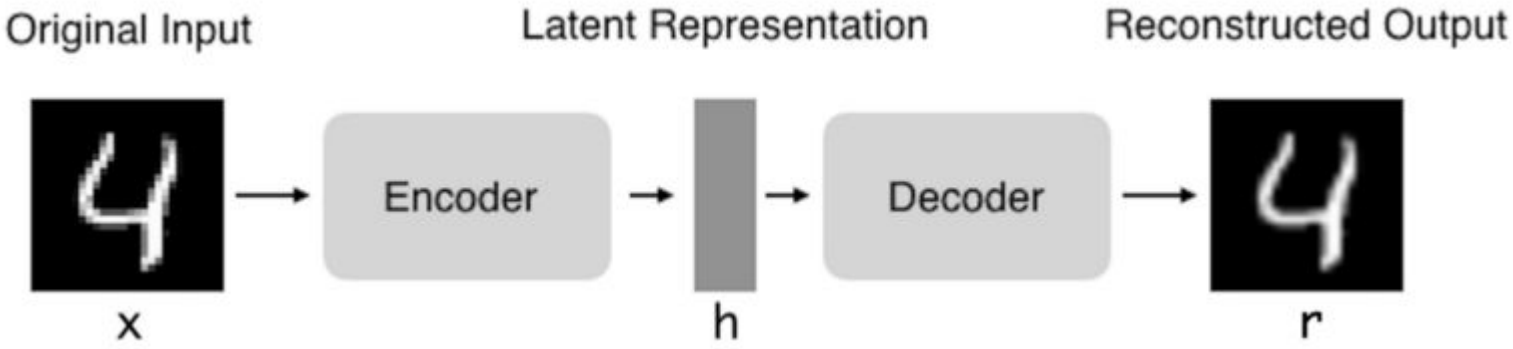
Lecture 3: Unsupervised Deep Learning

In this section:

- Auto-Encoders
  - what are they
  - How do they work
  - Types
- Generative modelling:
  - VAE
  - GAN
- Exercises:
  - MNIST AutoEncoder, latent space clustering
  - LSTM autoencoder Anomaly Detection

Auto-Encoders

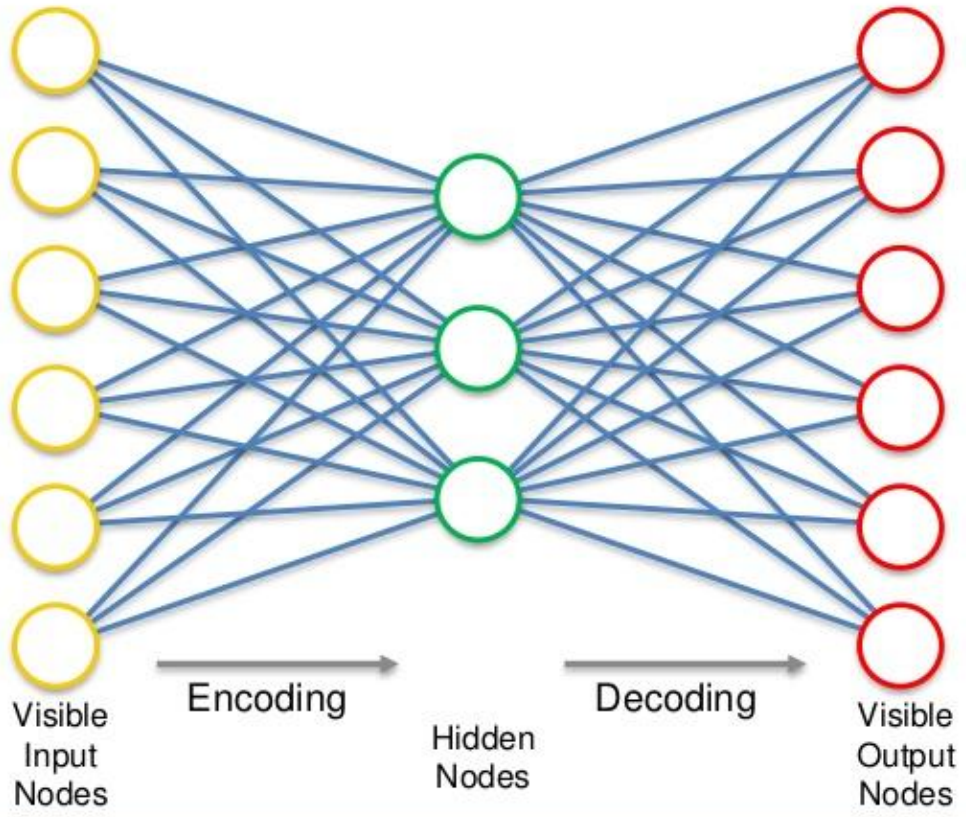
¿que son?


Original Input
Latent Representation
Reconstructed Output
Encoder
Decoder
x
h
r

¿que son?

Mondragon Unibertsitatea
Faculty of Engineering

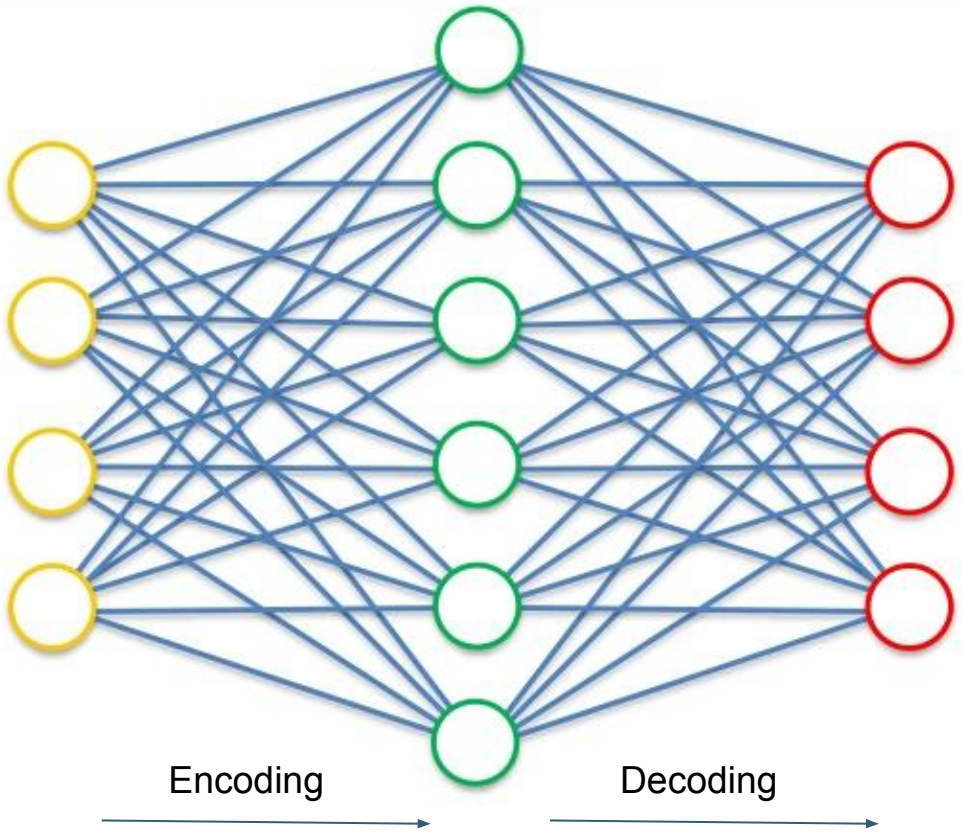
UnderComplete Auto-Encoder



¿que son?

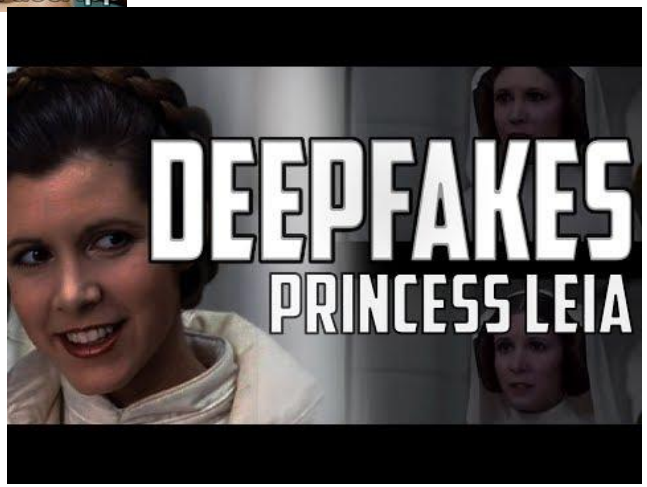
Mondragon Unibertsitatea
Faculty of Engineering

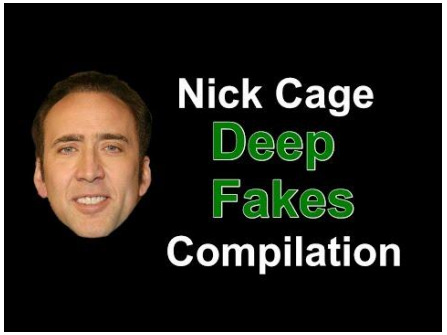
OverComplete Auto-Encoder



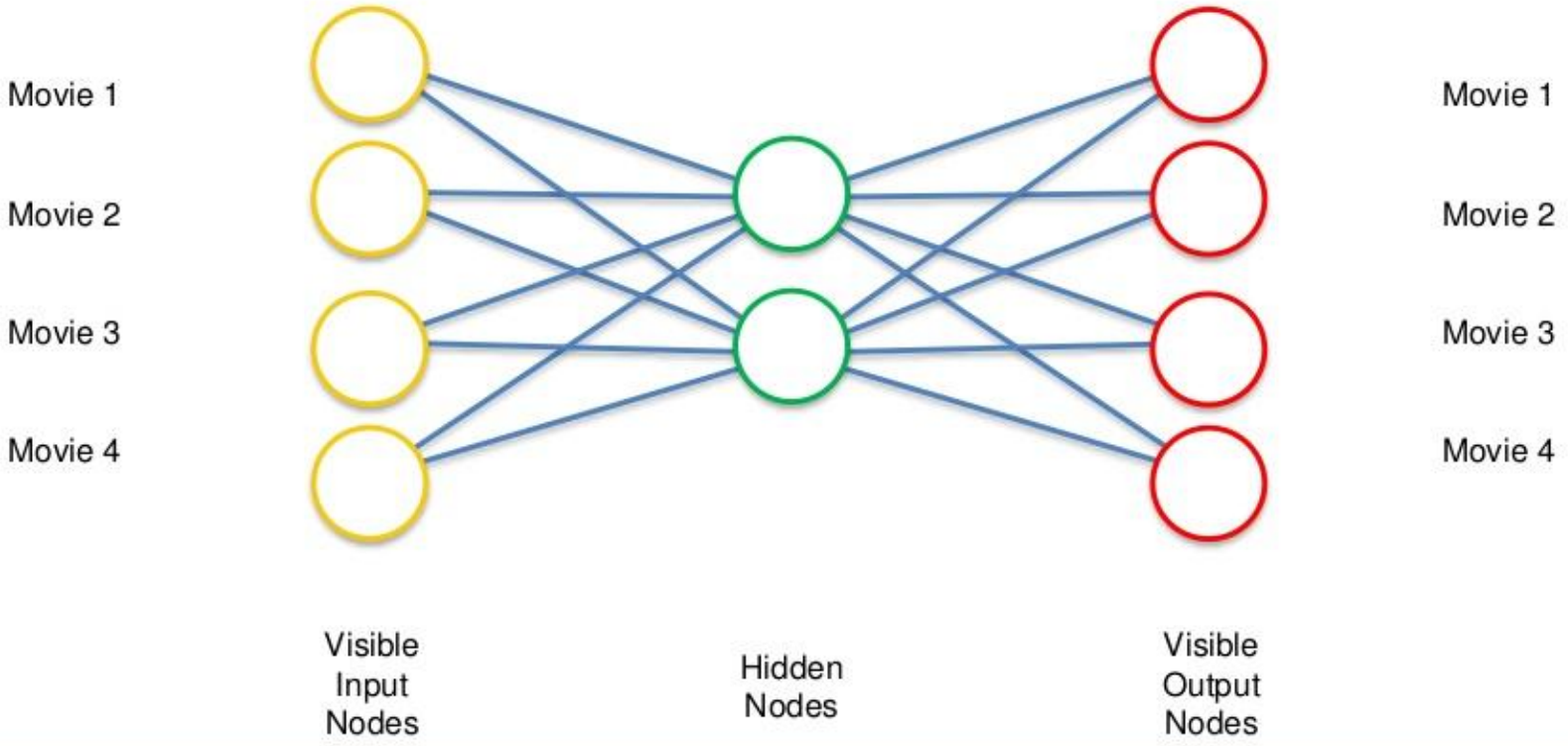
Aplicaciones

1. **Detección de anomalias**
2. **Quitar ruido de medidas**
3. **Reducción de dimensionalidad**
4. **Compresión**
5. **Sistemas de recomendación**
6. Y hacer FAKES

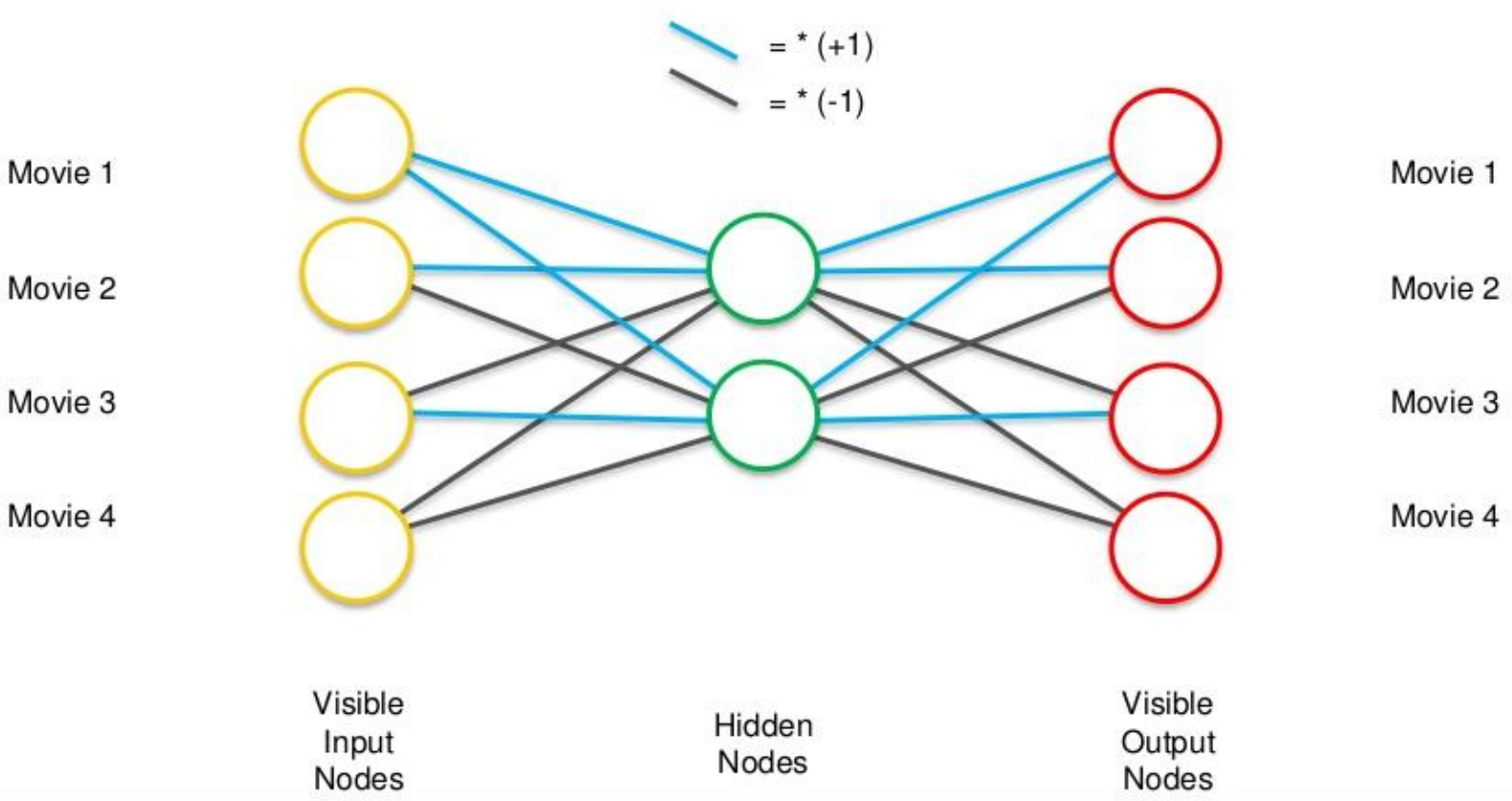




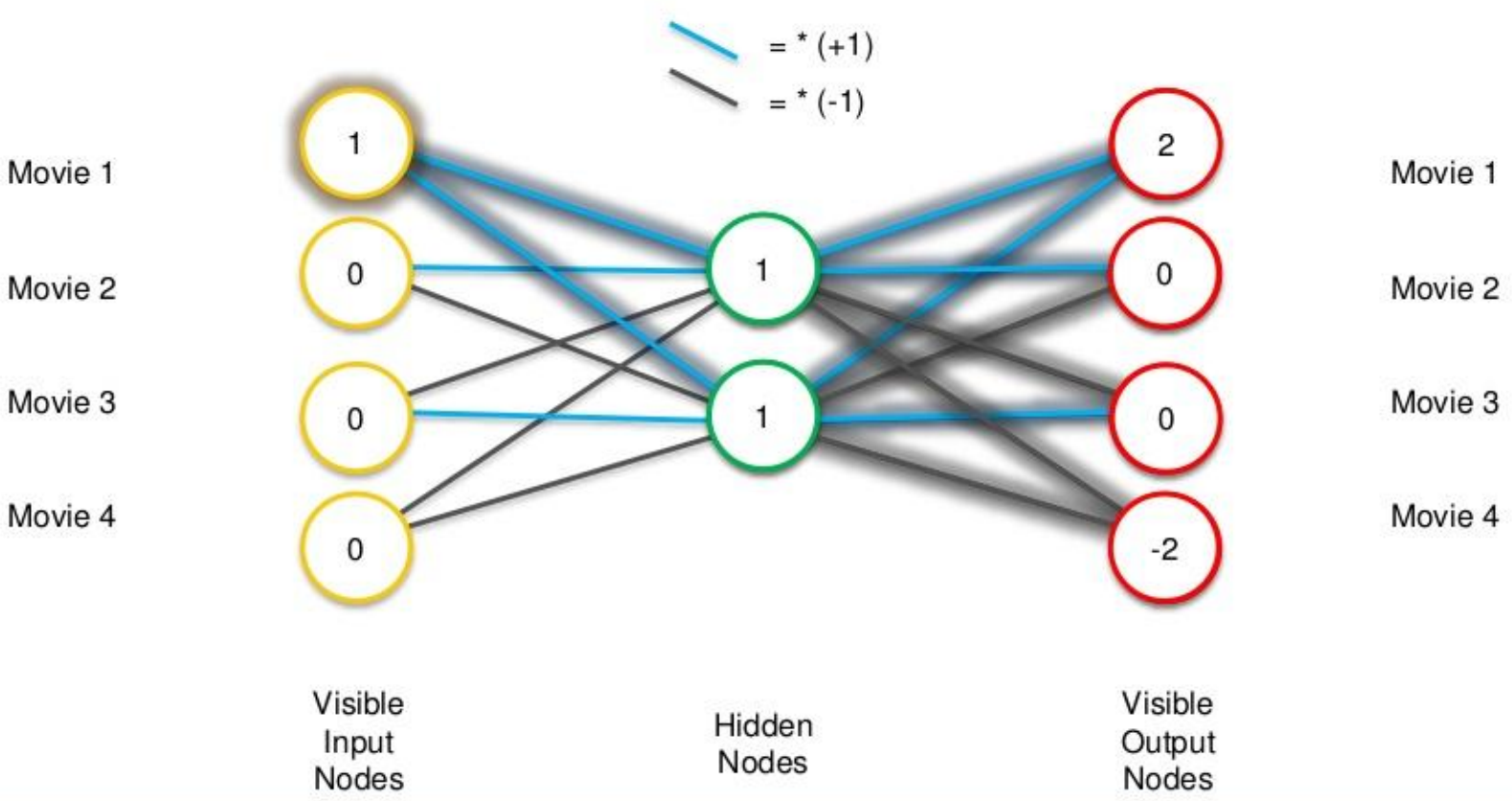
¿que son?


Movie 1
Movie 2
Movie 3
Movie 4
Movie 1
Movie 2
Movie 3
Movie 4
Visible Input Nodes
Hidden Nodes
Visible Output Nodes

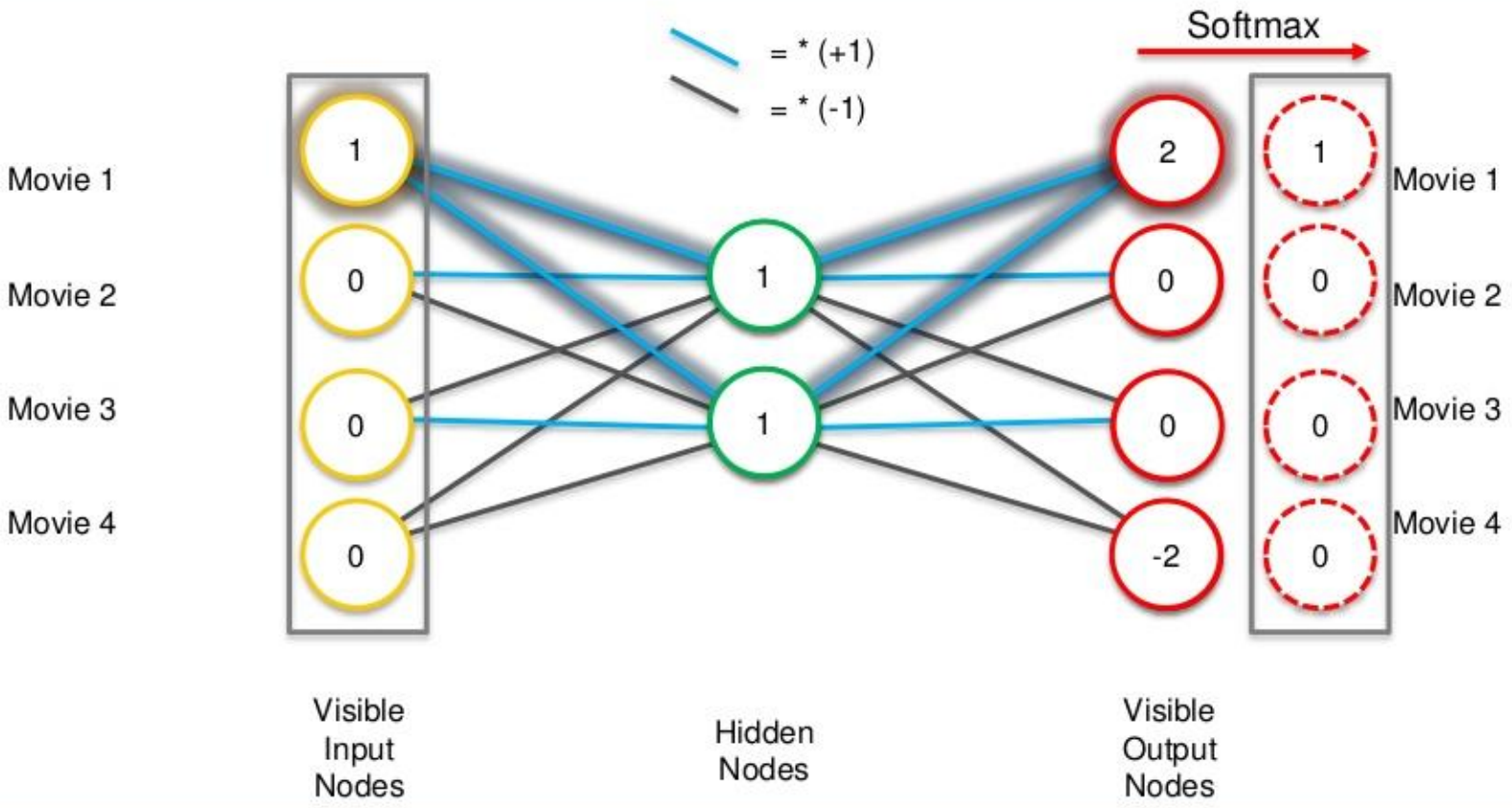
¿que son?


= * (+1)
= * (-1)
Movie 1
Movie 2
Movie 3
Movie 4
Movie 1
Movie 2
Movie 3
Movie 4
Visible Input Nodes
Hidden Nodes
Visible Output Nodes

¿que son?

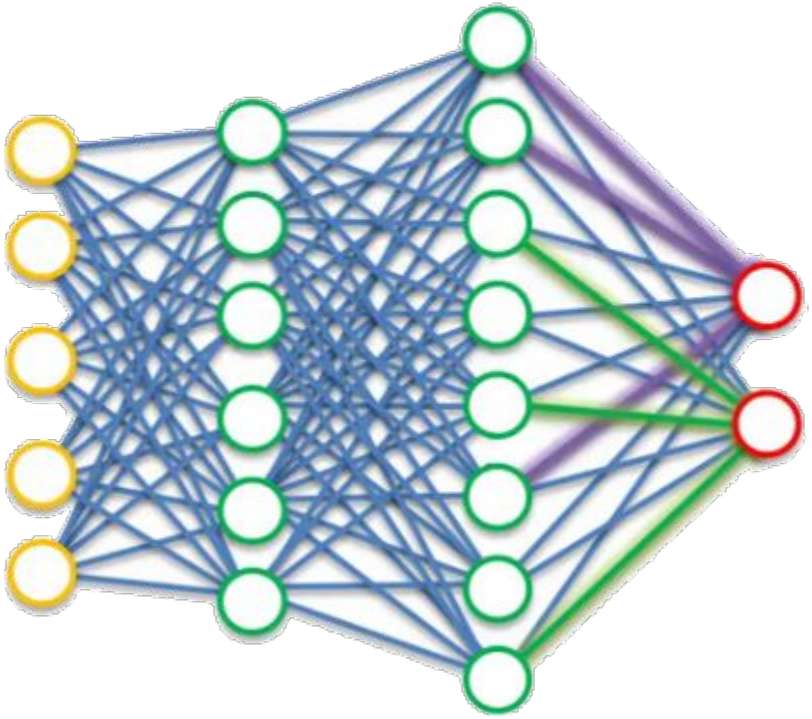

= * (+1)
= * (-1)
Movie 1
Movie 2
Movie 3
Movie 4
1
0
0
0
1
1
2
0
0
-2
Movie 1
Movie 2
Movie 3
Movie 4
Visible Input Nodes
Hidden Nodes
Visible Output Nodes

¿que son?


= * (+1)
= * (-1)
Softmax
Movie 1
Movie 2
Movie 3
Movie 4
1
0
0
0
1
1
2
0
0
-2
1
0
0
0
Movie 1
Movie 2
Movie 3
Movie 4
Visible Input Nodes
Hidden Nodes
Visible Output Nodes

SoftMax

- Hace que las salidas de un vector sumen en total 1

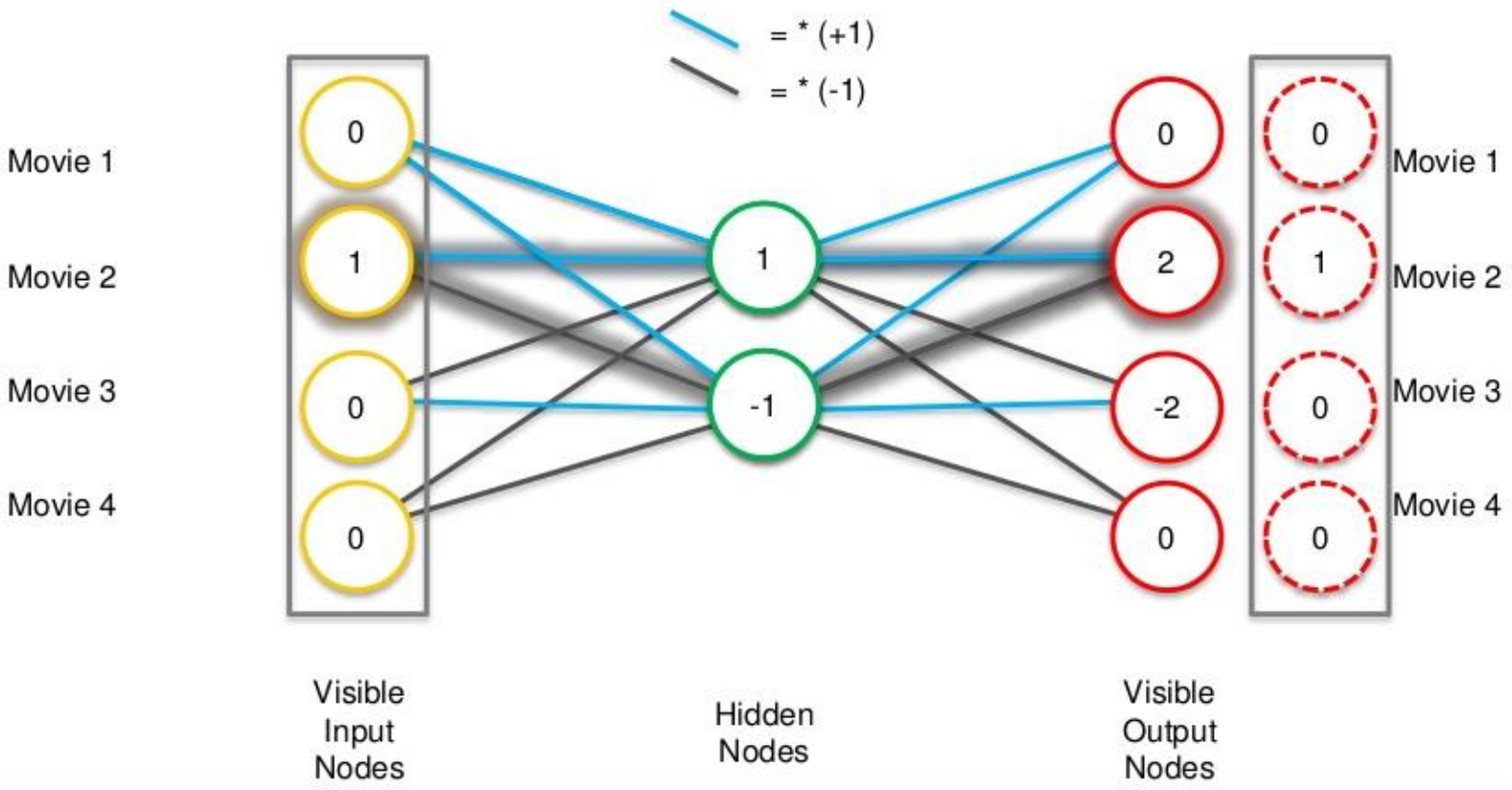


$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

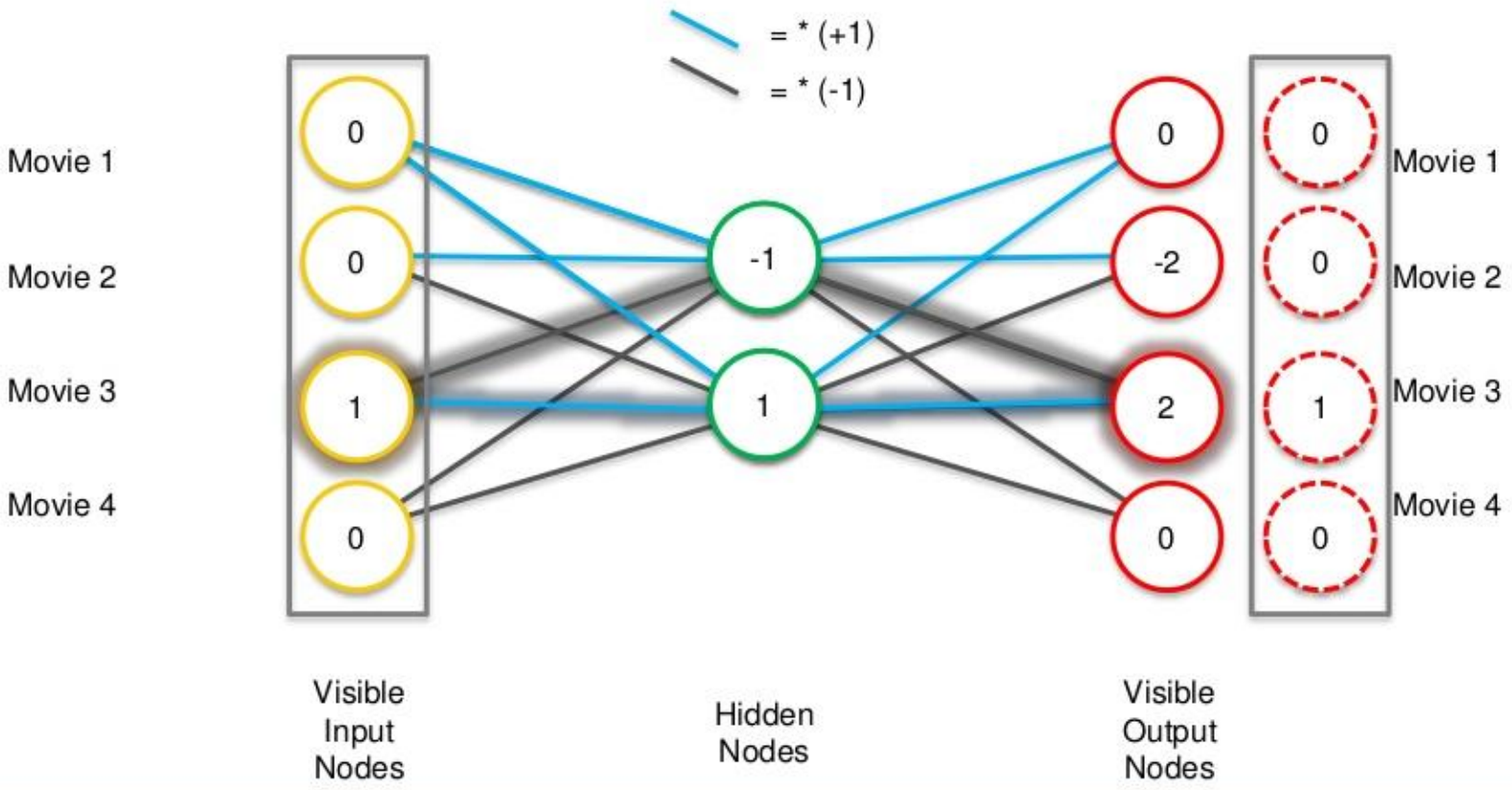
Dog → z_1 → 0.95

Cat → z_2 → 0.05

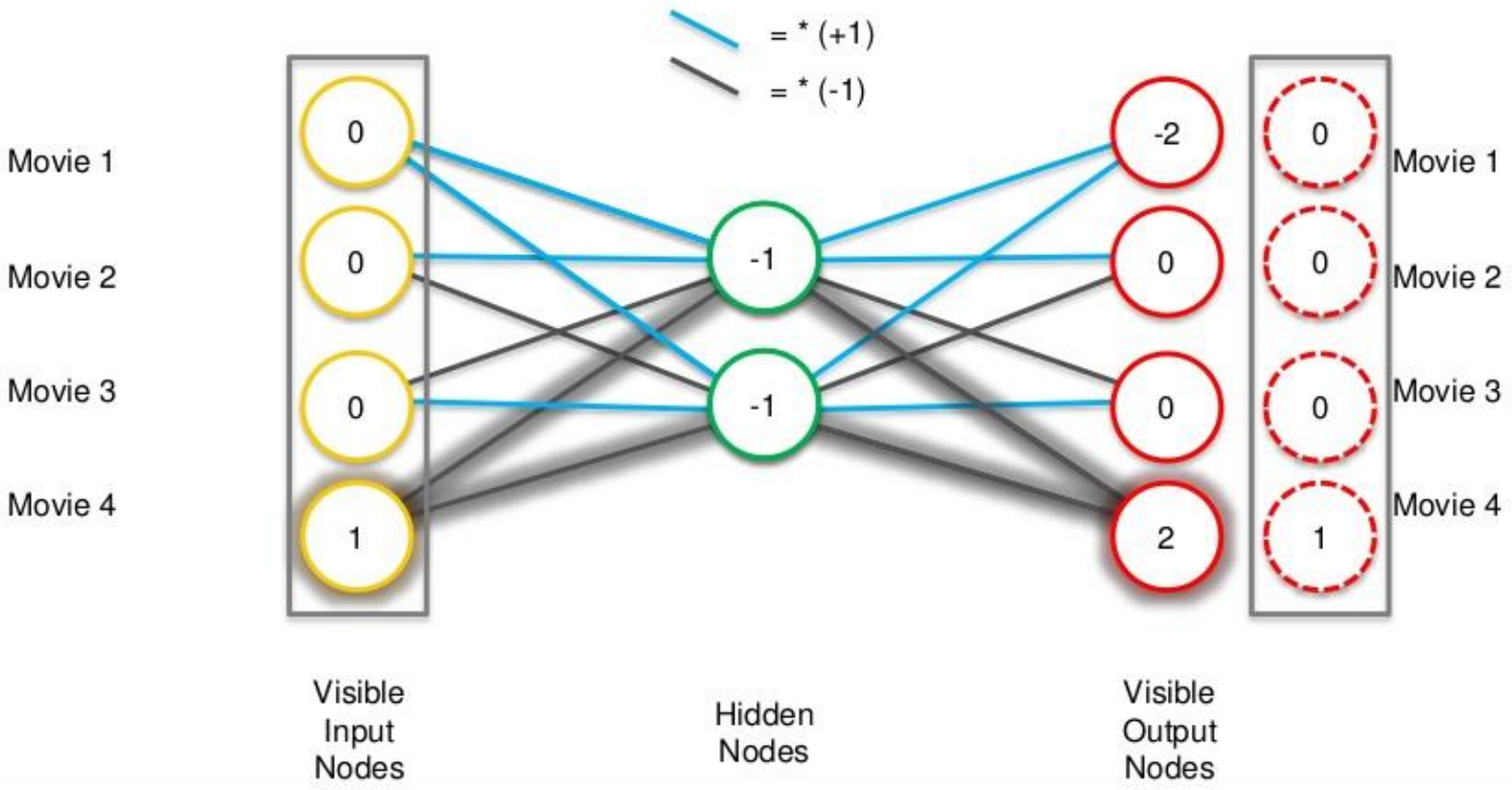
¿que son?


= * (+1)
= * (-1)
Movie 1
Movie 2
Movie 3
Movie 4
0
1
0
0
1
-1
0
2
-2
0
0
1
0
0
Movie 1
Movie 2
Movie 3
Movie 4
Visible Input Nodes
Hidden Nodes
Visible Output Nodes

¿que son?


= * (+1)
= * (-1)
Movie 1
Movie 2
Movie 3
Movie 4
0
0
1
0
-1
1
0
-2
2
0
0
0
1
0
Movie 1
Movie 2
Movie 3
Movie 4
Visible Input Nodes
Hidden Nodes
Visible Output Nodes

¿que son?


= * (+1)
= * (-1)
Movie 1
Movie 2
Movie 3
Movie 4
0
0
0
1
-1
-1
-2
0
0
2
0
0
0
1
Movie 1
Movie 2
Movie 3
Movie 4
Visible Input Nodes
Hidden Nodes
Visible Output Nodes

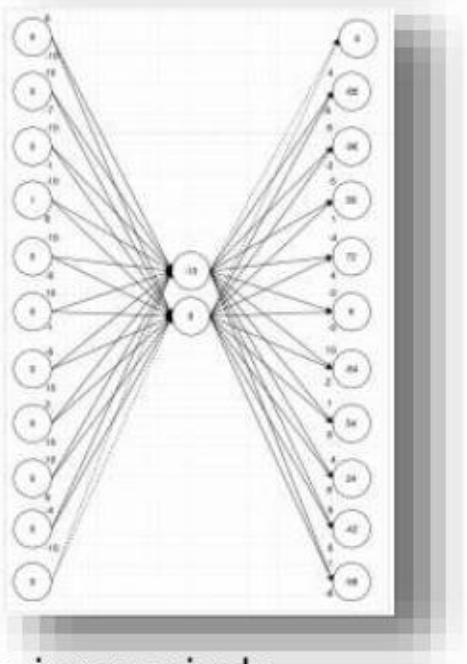
Aditional Reading

Mondragon Unibertsitatea
Faculty of Engineering

Additional Reading:

Neural Networks Are Impressively Good At Compression

By Malte Skarupke (2016)



Link:

https://probablydance.com/2016/04/30/neural-networks-are-impressively-good-at-compression/

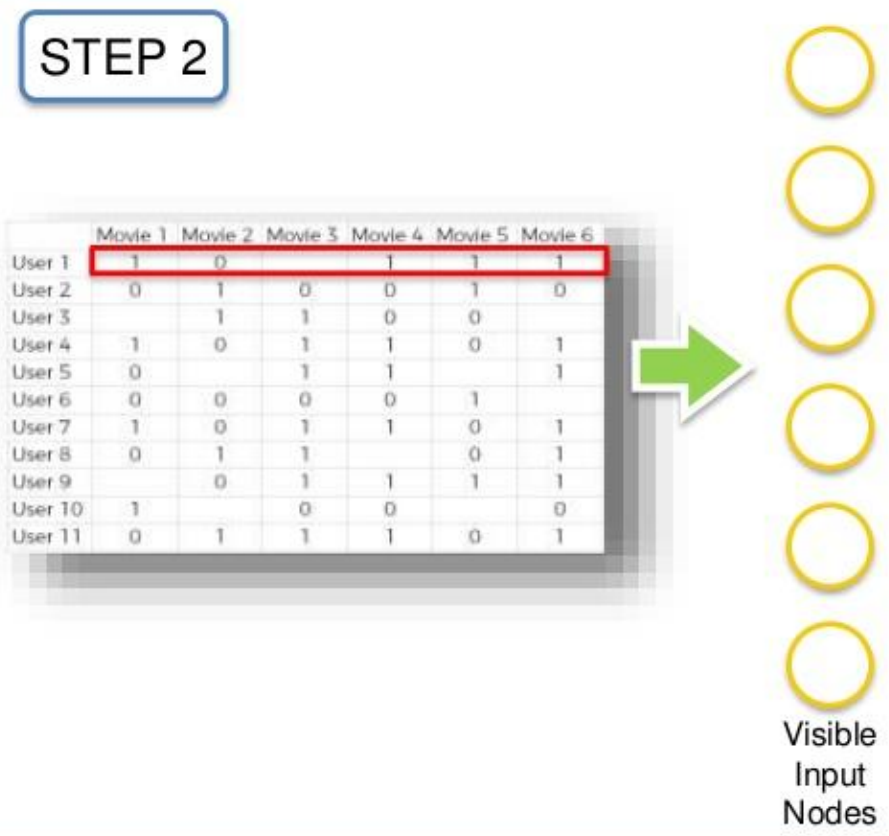
¿Como Funcionan?

Mondragon Unibertsitatea
Faculty of Engineering

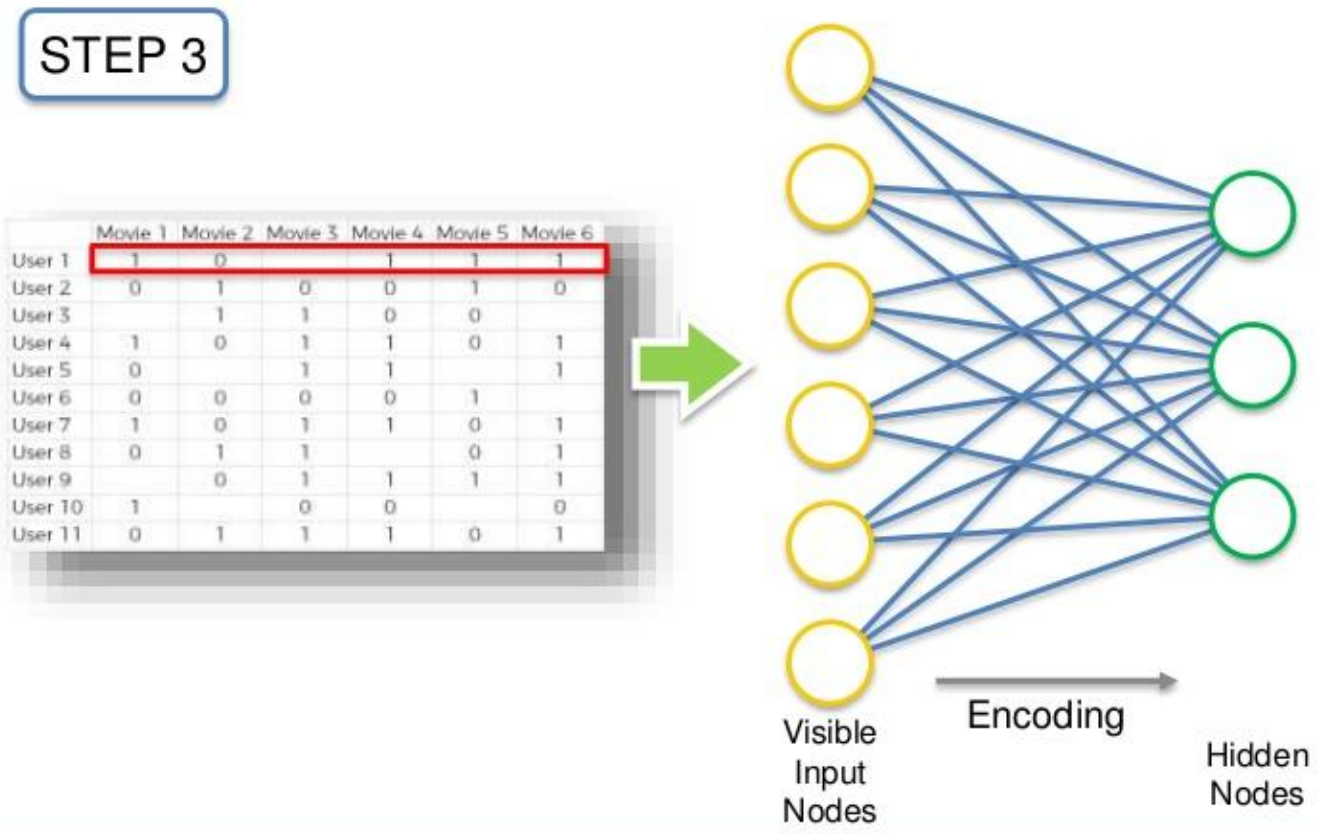
STEP 1

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie 6
User 1	1	0		1	1	1
User 2	0	1	0	0	1	0
User 3		1	1	0	0	
User 4	1	0	1	1	0	1
User 5	0		1	1		1
User 6	0	0	0	0	1	
User 7	1	0	1	1	0	1
User 8	0	1	1		0	1
User 9		0	1	1	1	1
User 10	1		0	0		0
User 11	0	1	1	1	0	1

¿Como Funcionan?



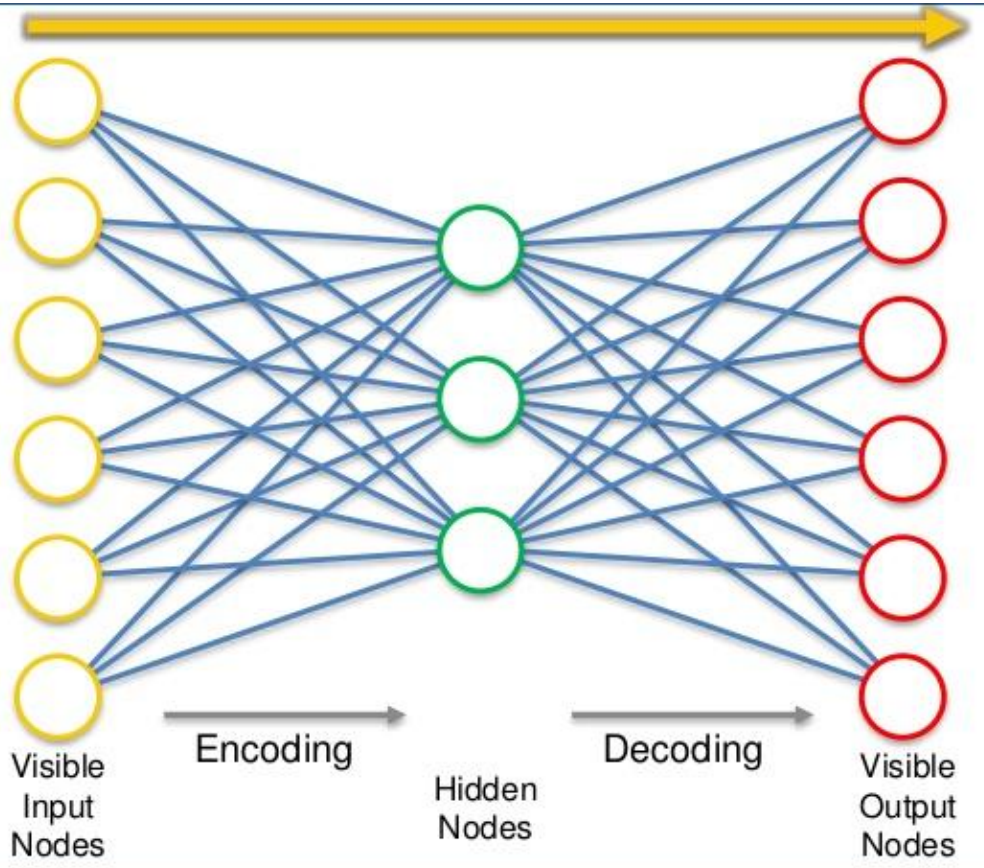
¿Como Funcionan?



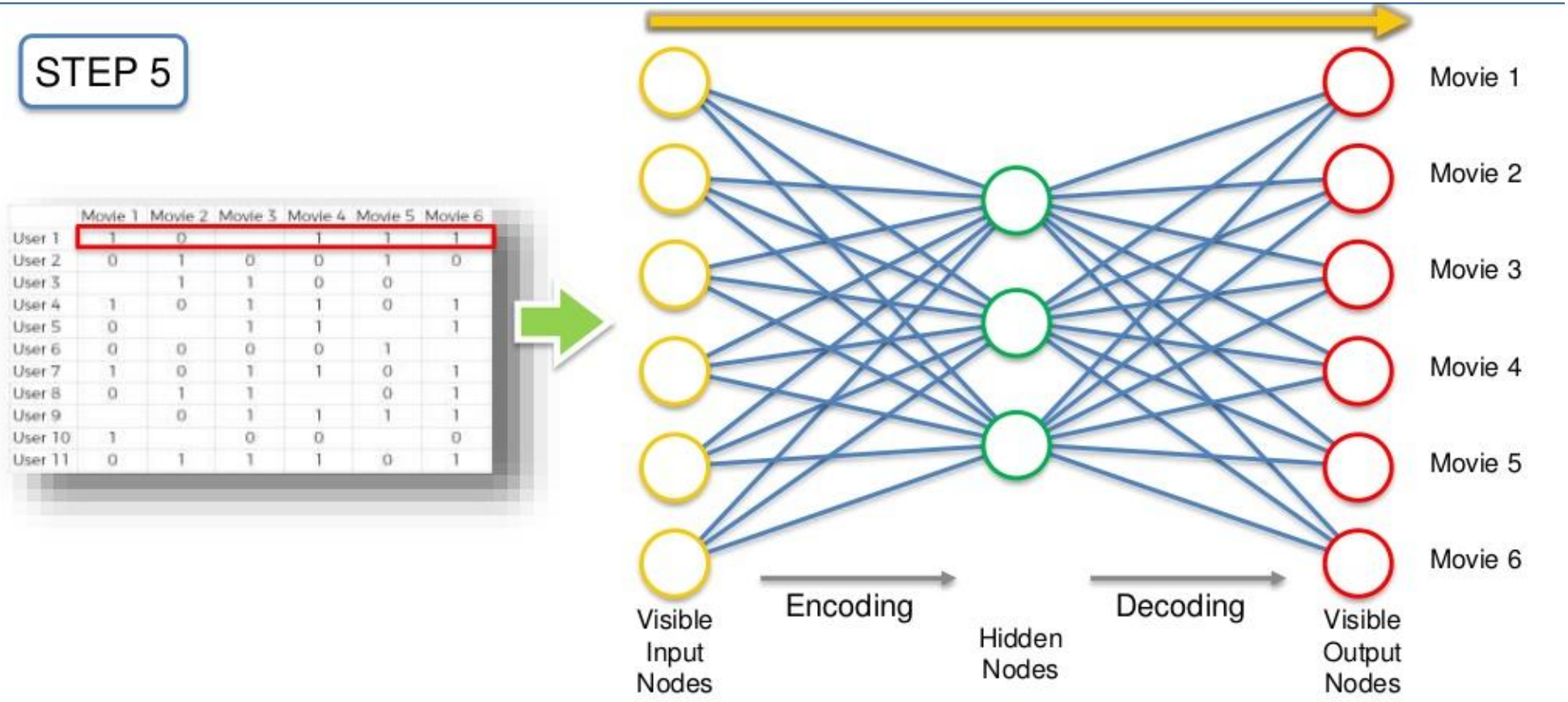
¿Como Funcionan?

STEP 4

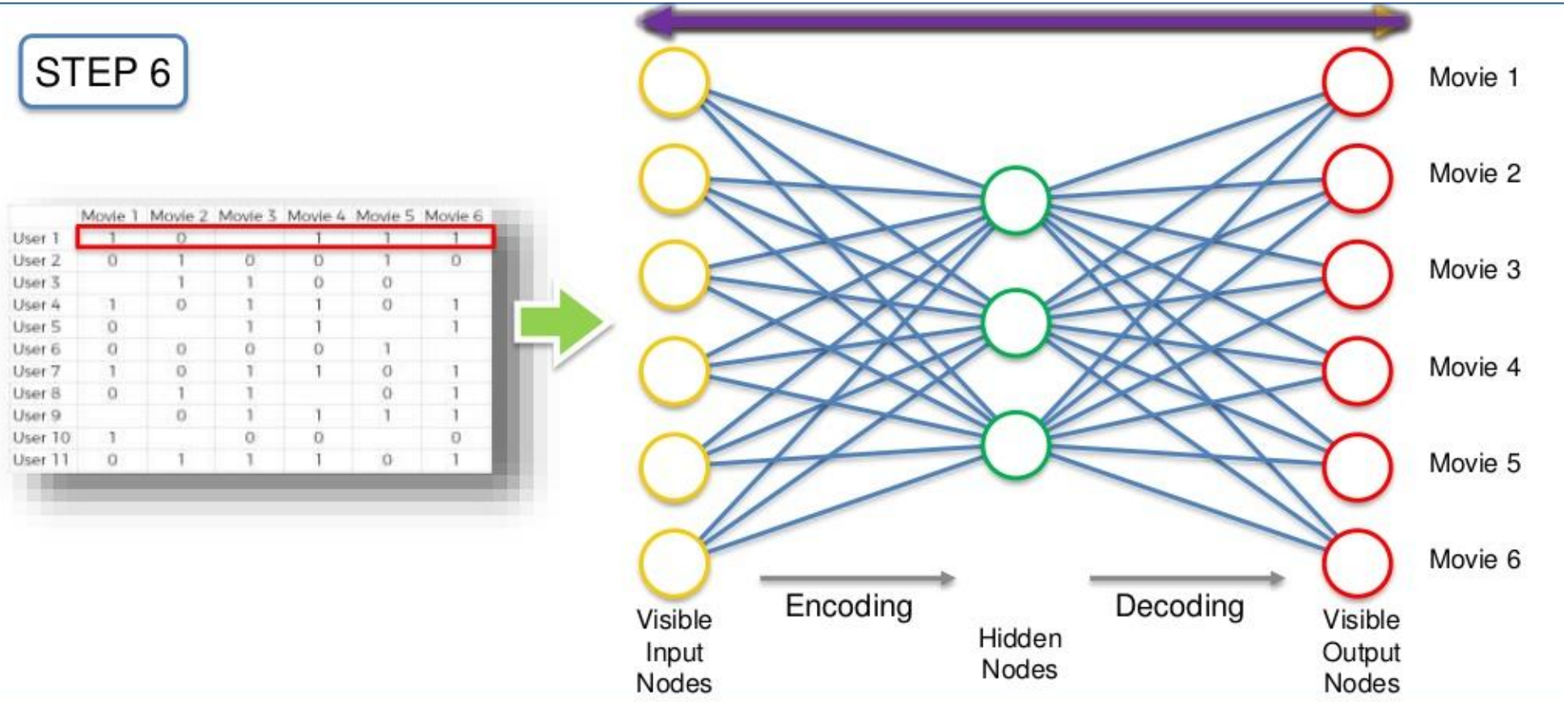
	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie 6
User 1	1	0		1	1	1
User 2	0	1	0	0	1	0
User 3		1	1	0	0	
User 4	1	0	1	1	0	1
User 5	0		1	1		1
User 6	0	0	0	0	1	
User 7	1	0	1	1	0	1
User 8	0	1	1		0	1
User 9		0	1	1	1	1
User 10	1		0	0		0
User 11	0	1	1	1	0	1



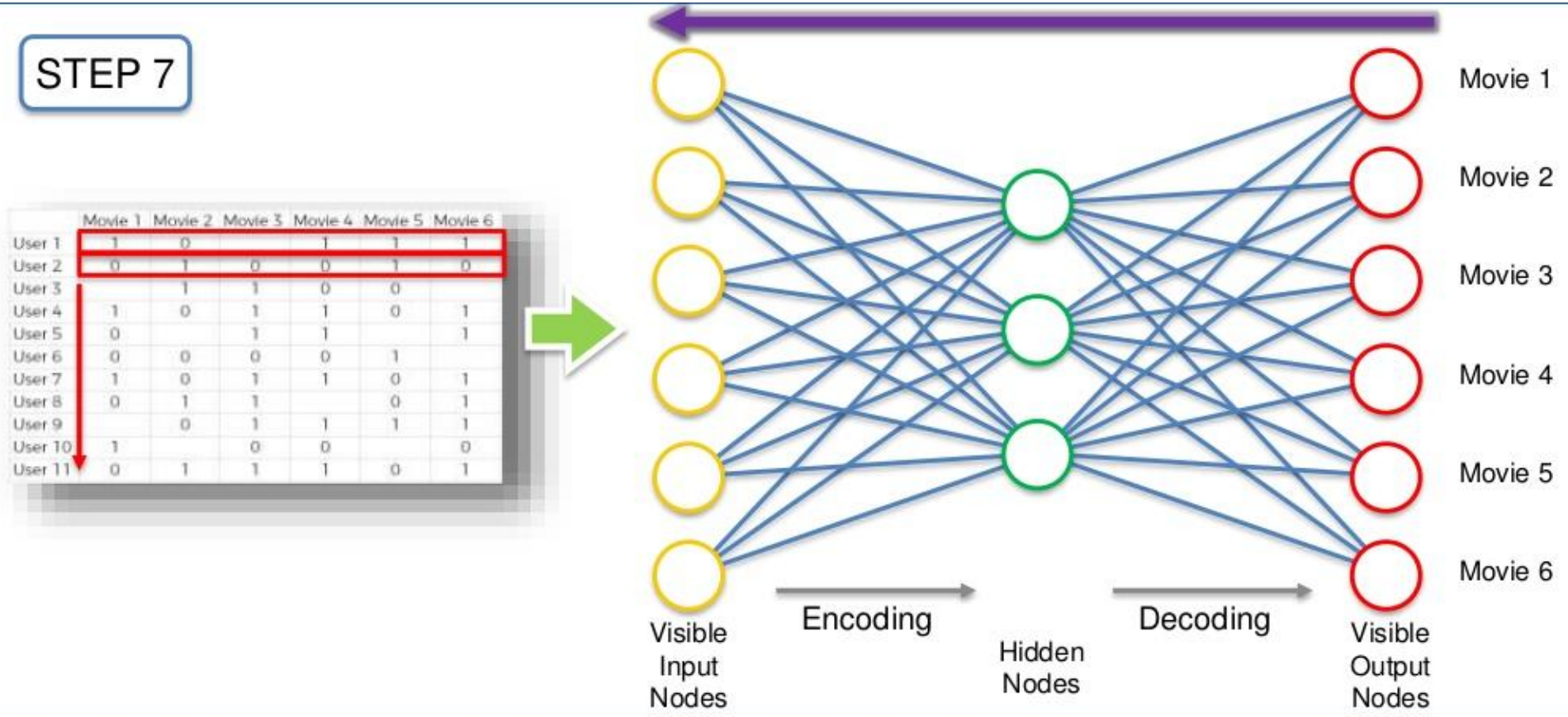
¿Como Funcionan?



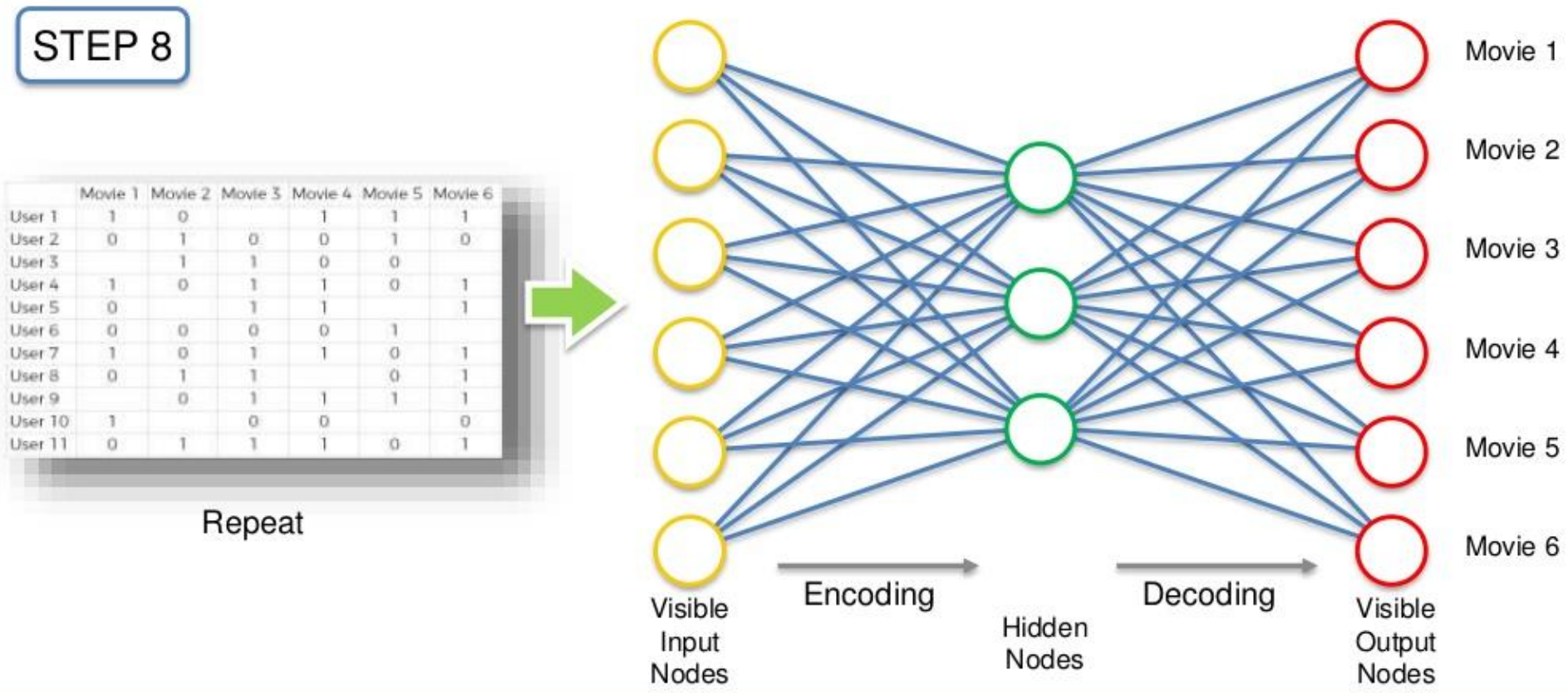
¿Como Funcionan?



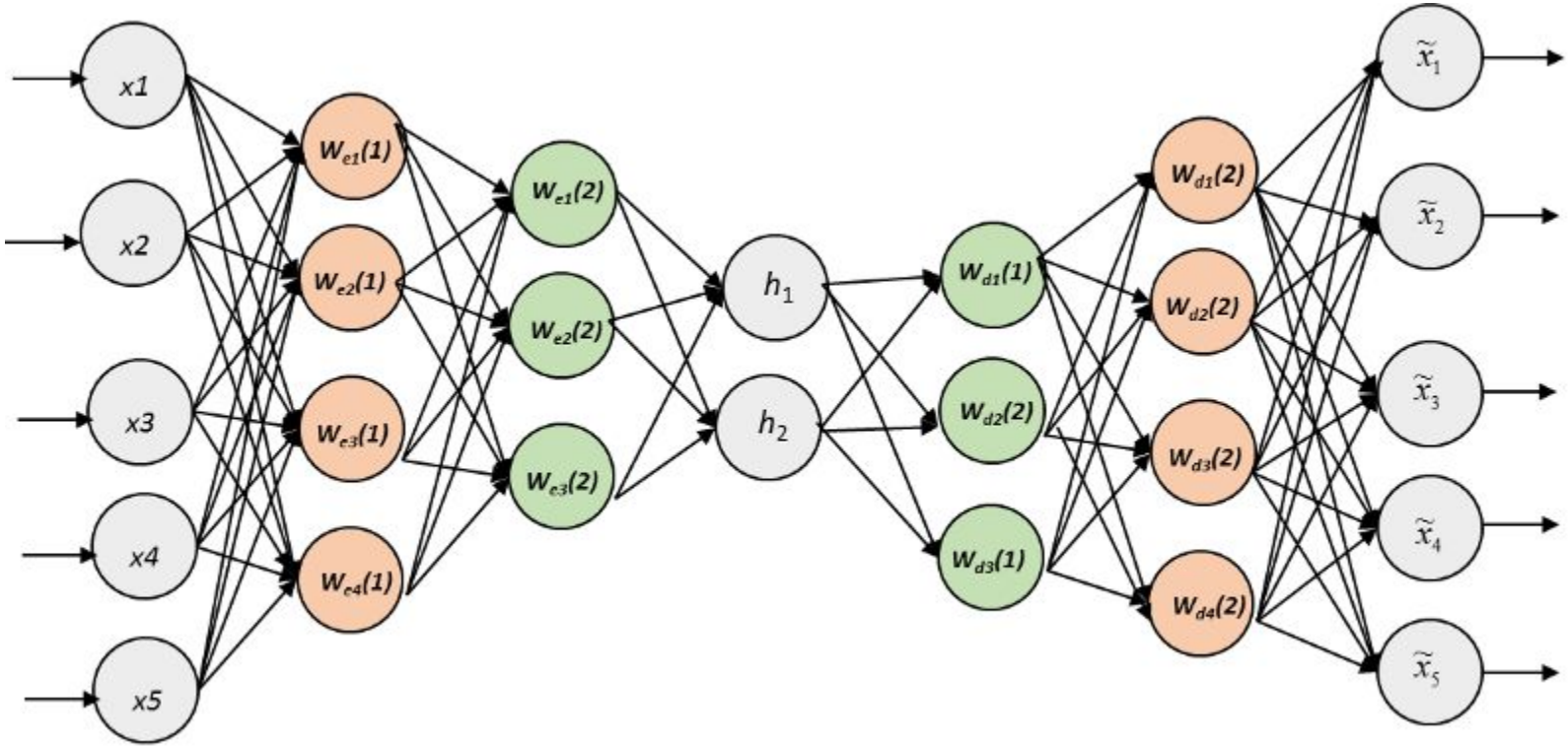
¿Como Funcionan?



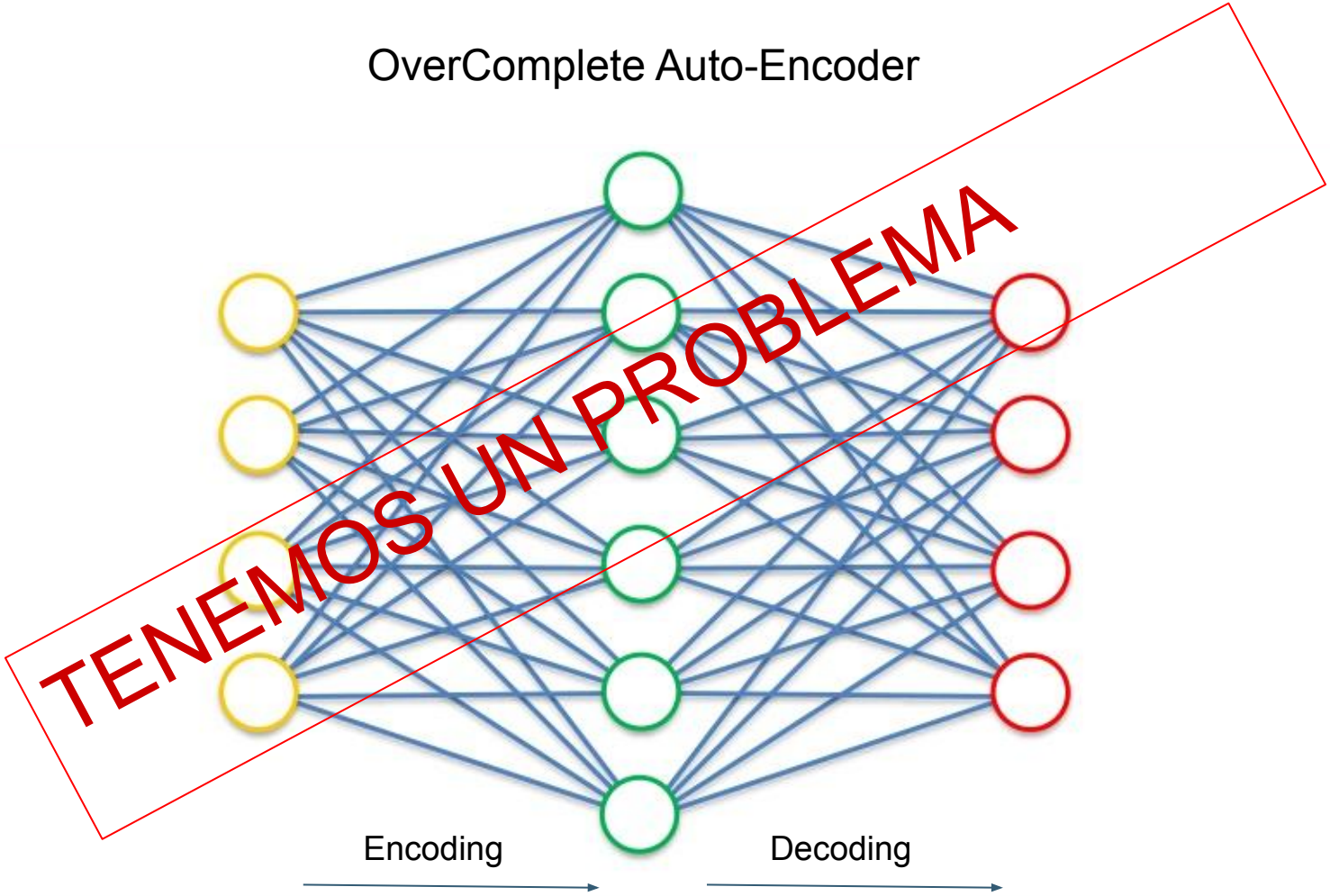
¿Como Funcionan?



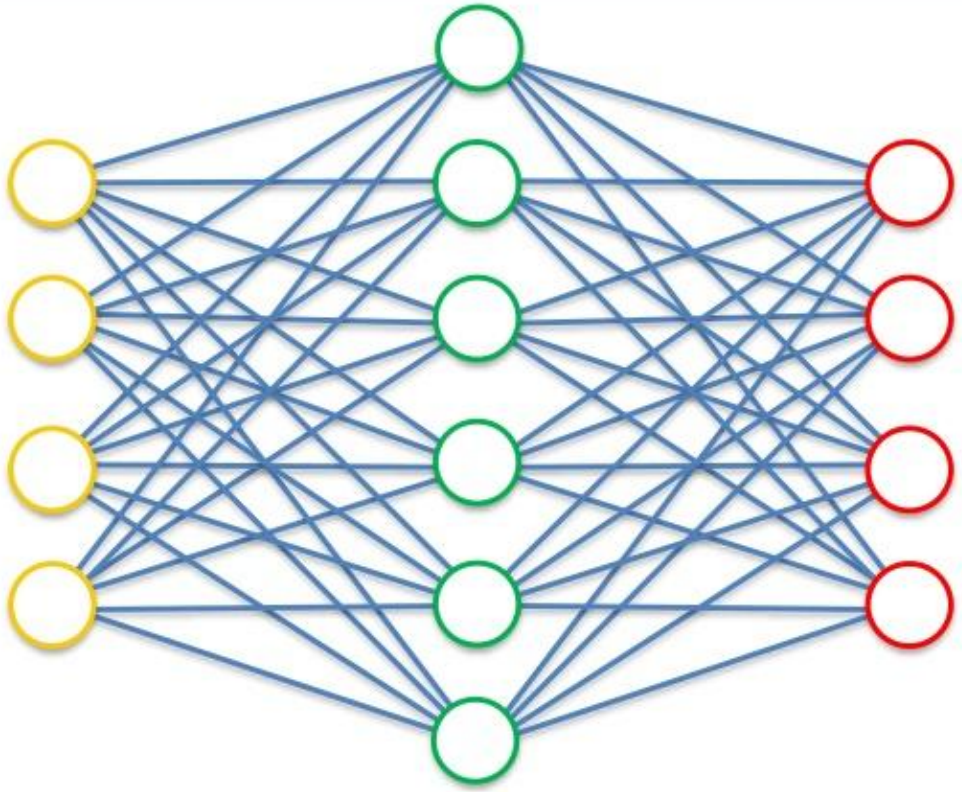
Deep Auto-Encoders



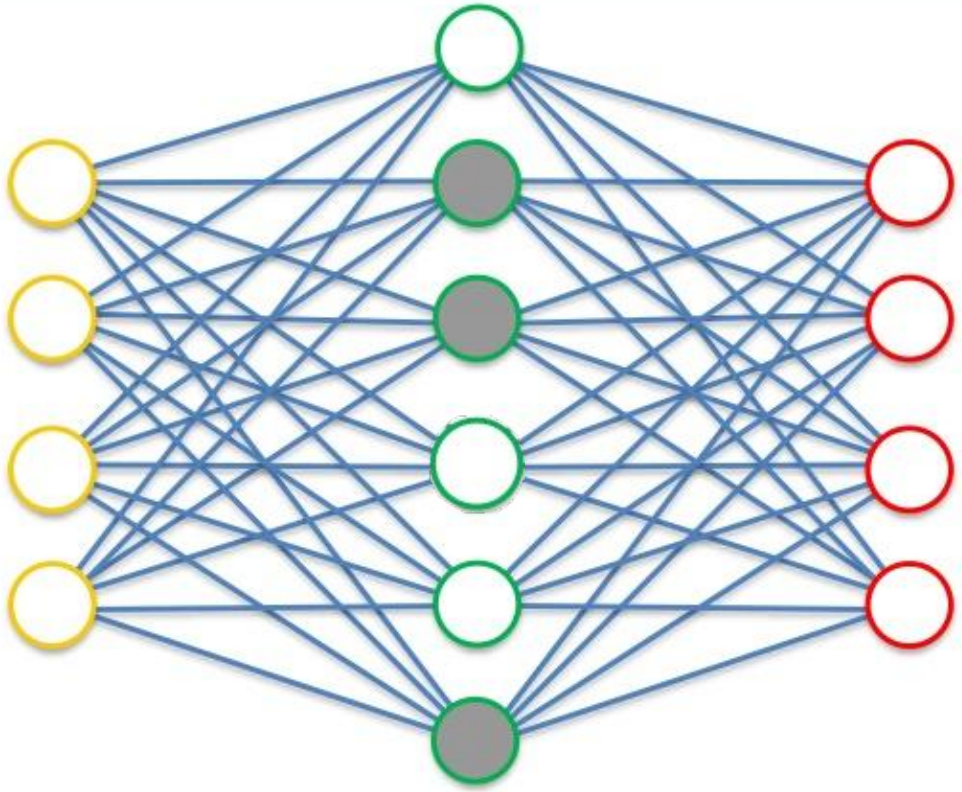
¿y si quiero codificar a más variables?



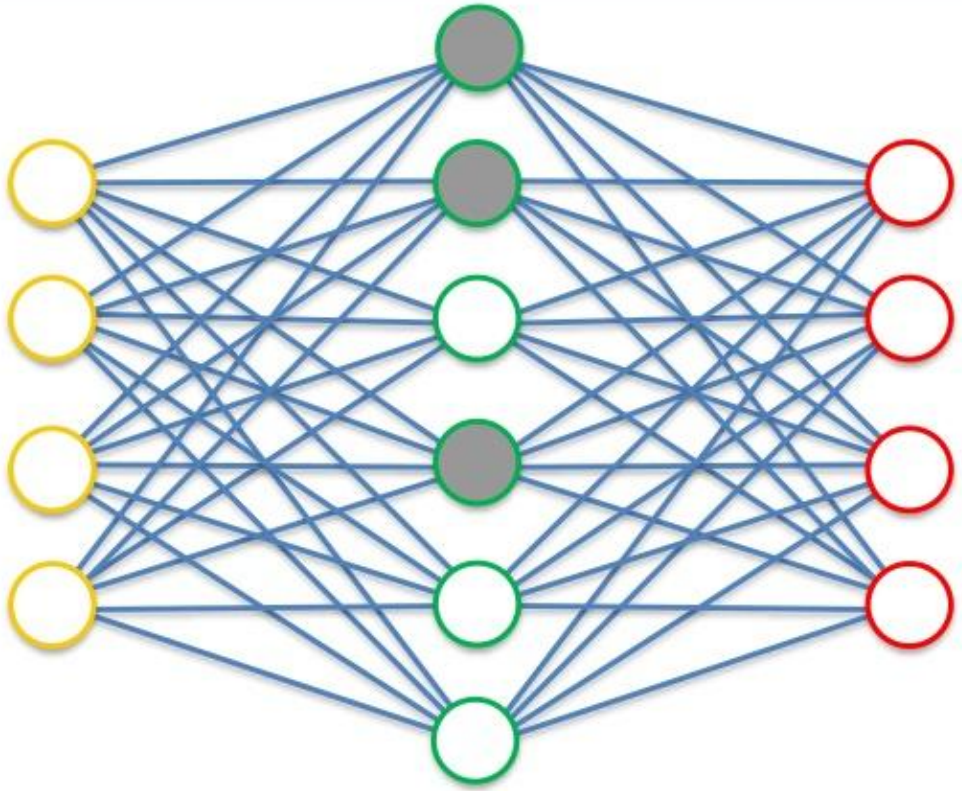
Soluciones: Sparse Autoencoder



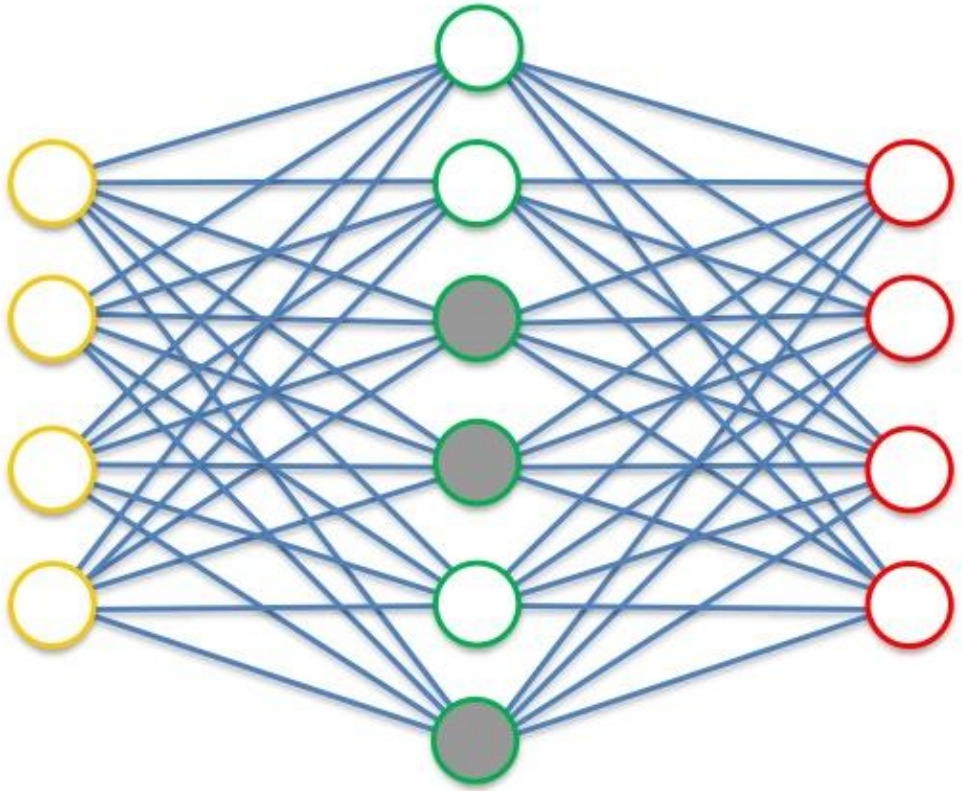
¿Como Funcionan?



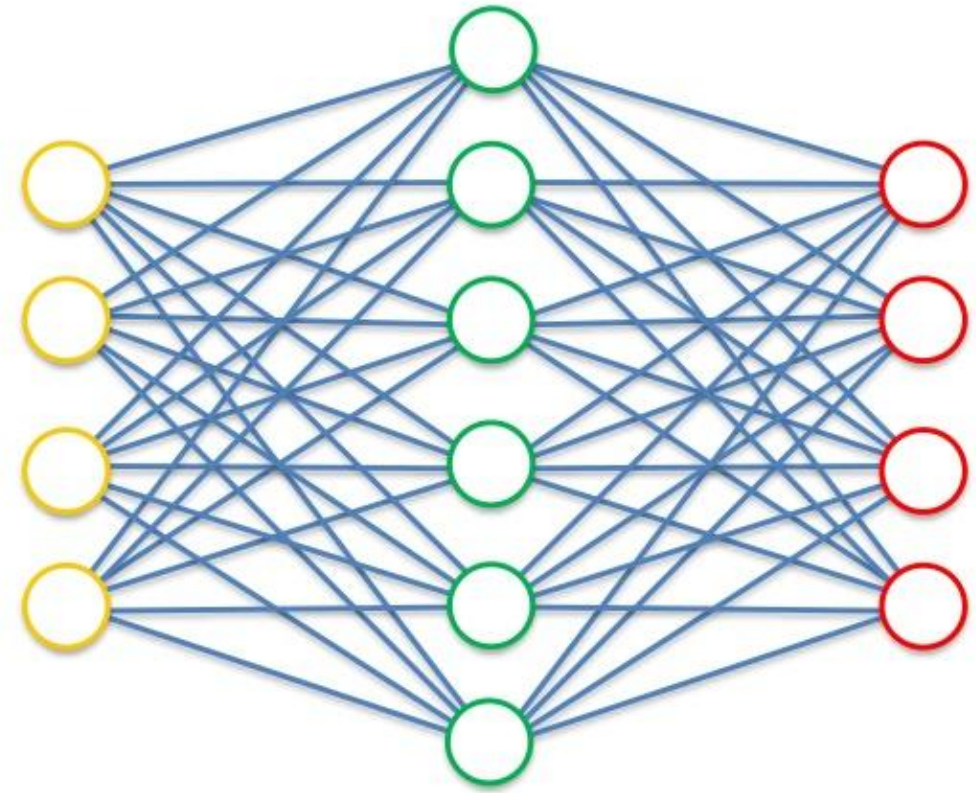
¿Como Funcionan?



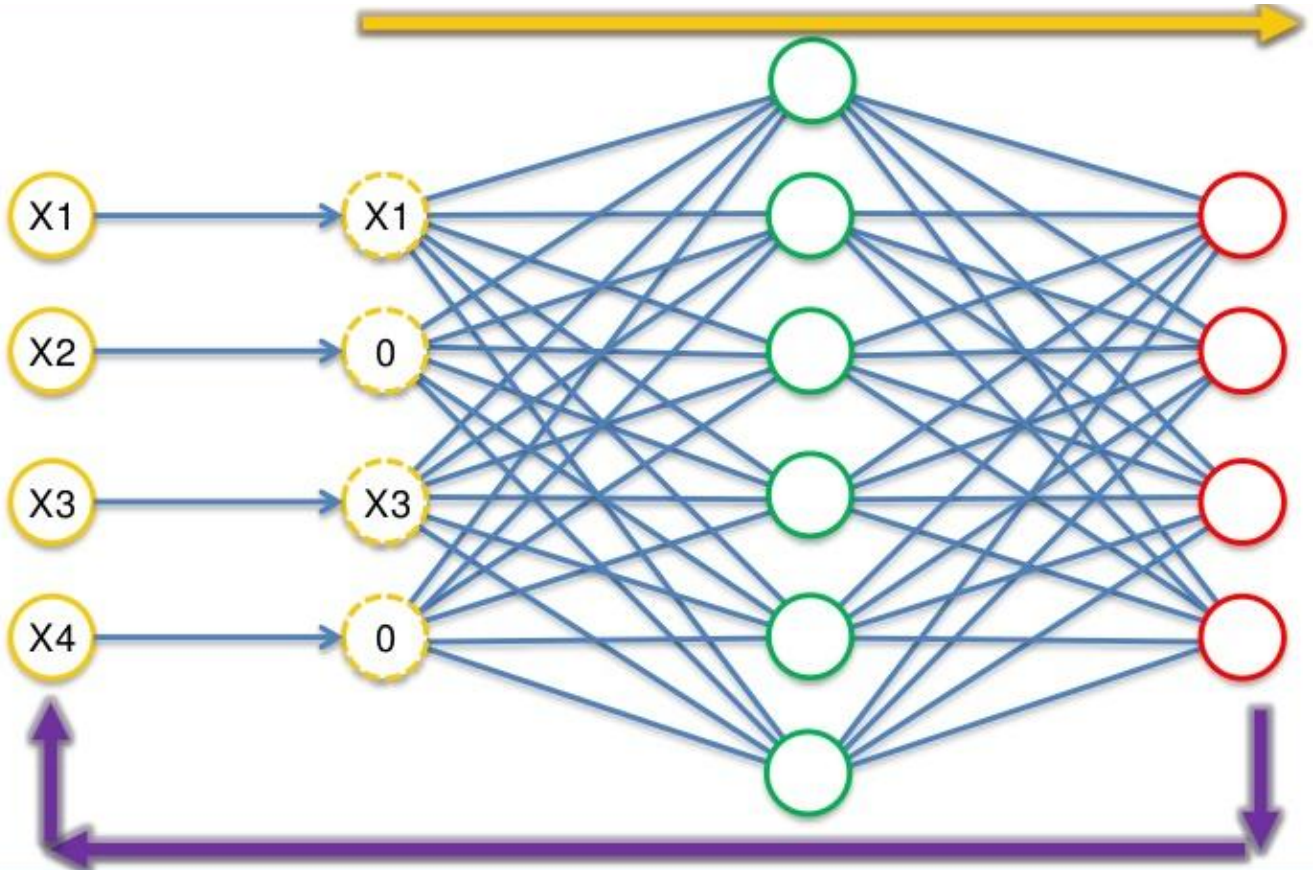
¿Como Funcionan?



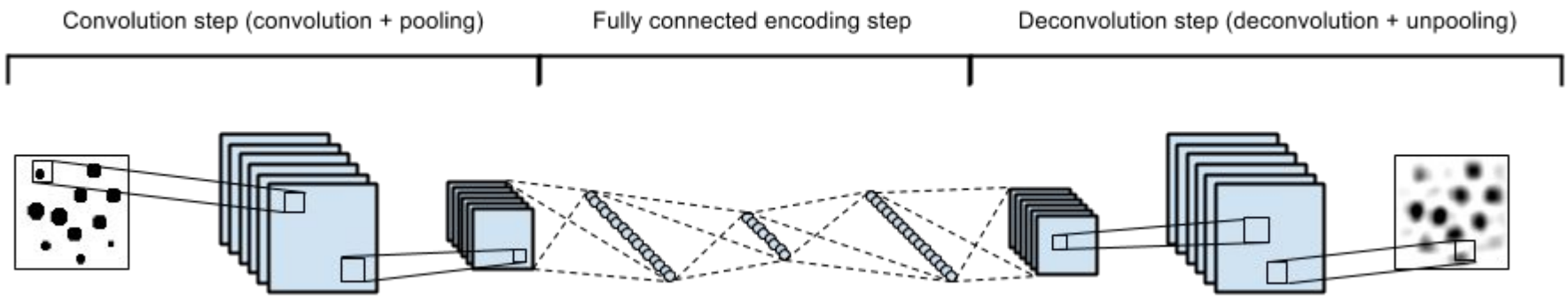
overcomplete Denoising auto-encoder



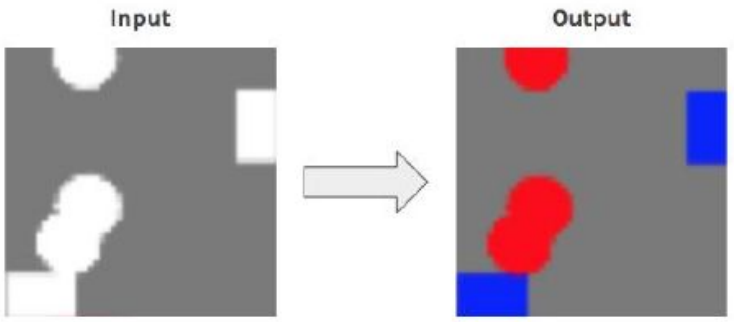
¿Como Funcionan?



Convolutional auto-encoder

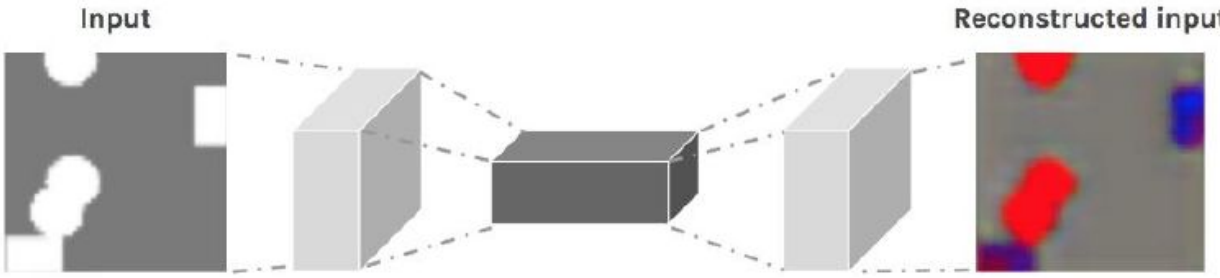


Convolutional auto-encoder



The CAE is trained to colorize the image

IDEA!
proyect!

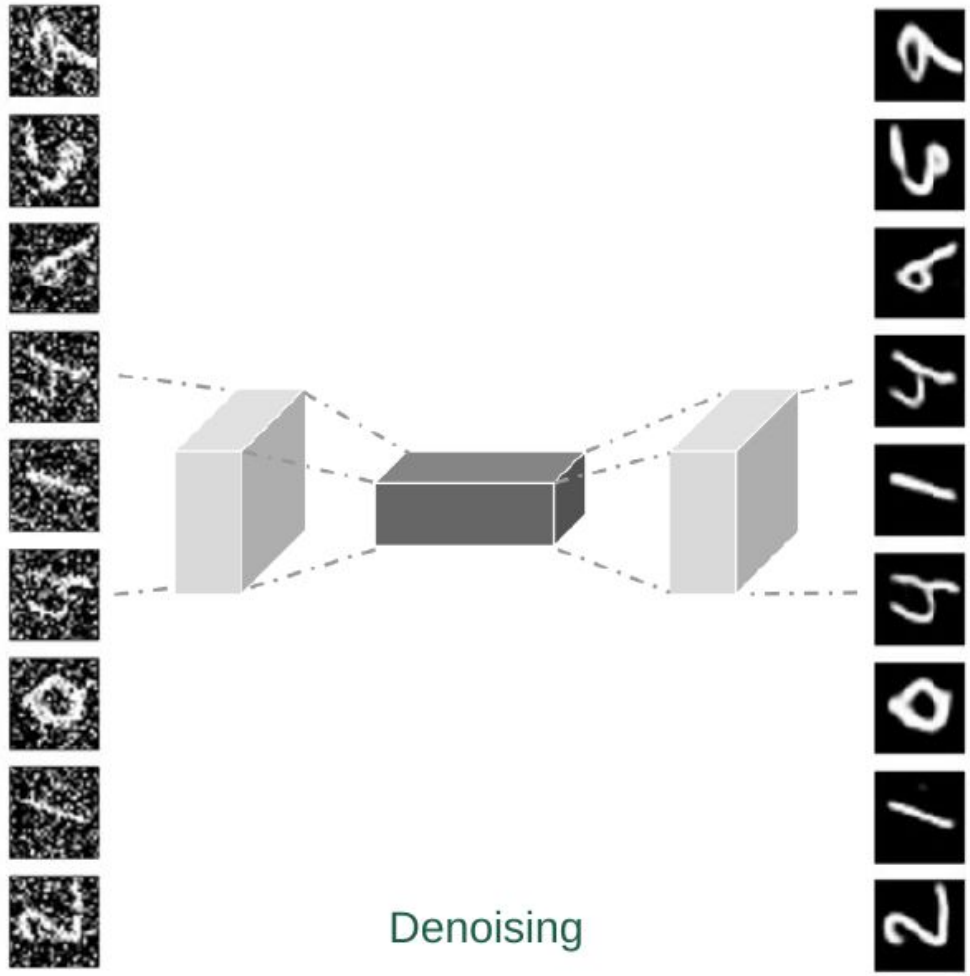


Even though the reconstruction is blurry, the color are mostly right

Image colorization

https://blog.floydhub.com/colorizing-b-w-photos-with-neural-networks/

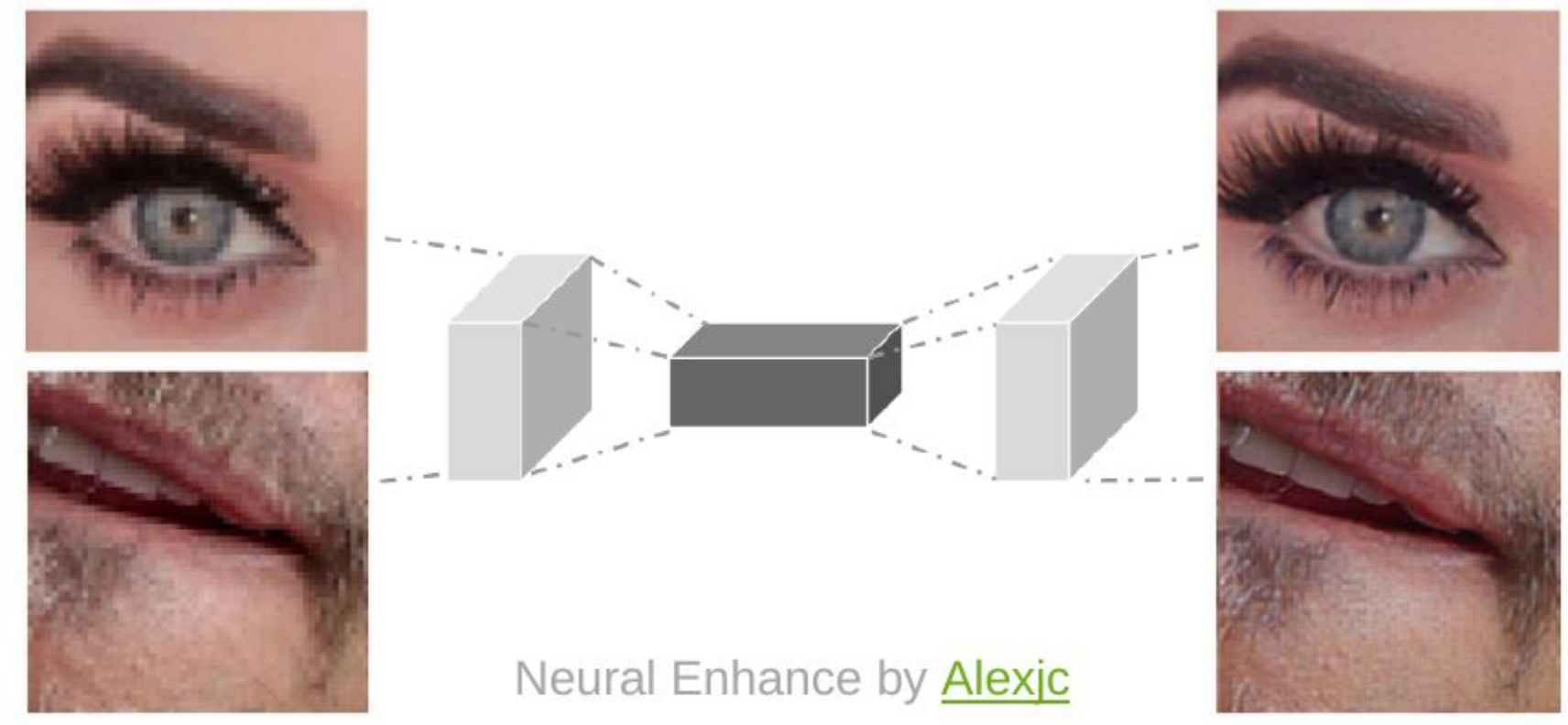
Convolutional auto-encoder



IDEA!
proyect!
(not mnist!)

https://towardsdatascience.com/convolutional-autoencoders-for-image-noise-reduction-32fce9fc1763

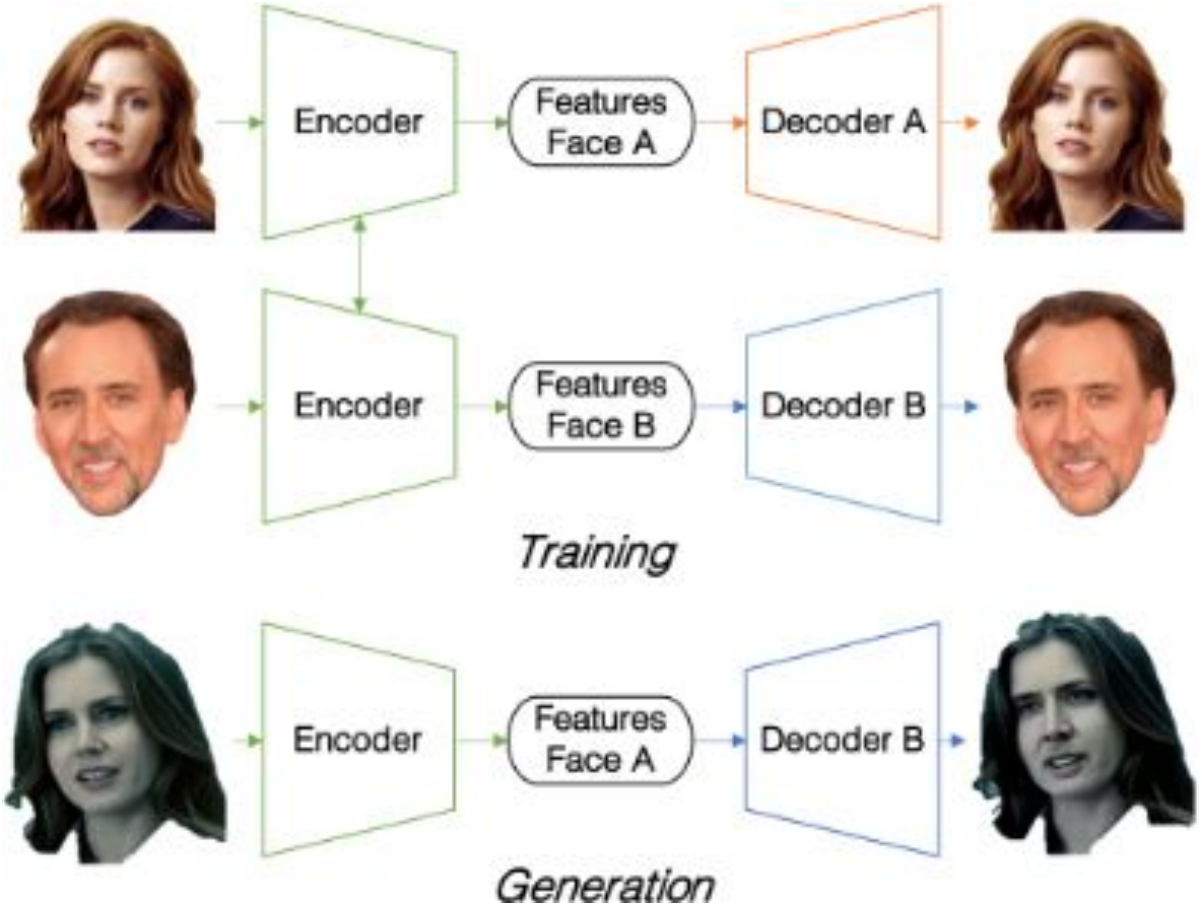
Convolutional auto-encoder



https://github.com/alexjc/neural-enhance

Image resolution enhancement

¿Cómo se hacen los deepFakes?



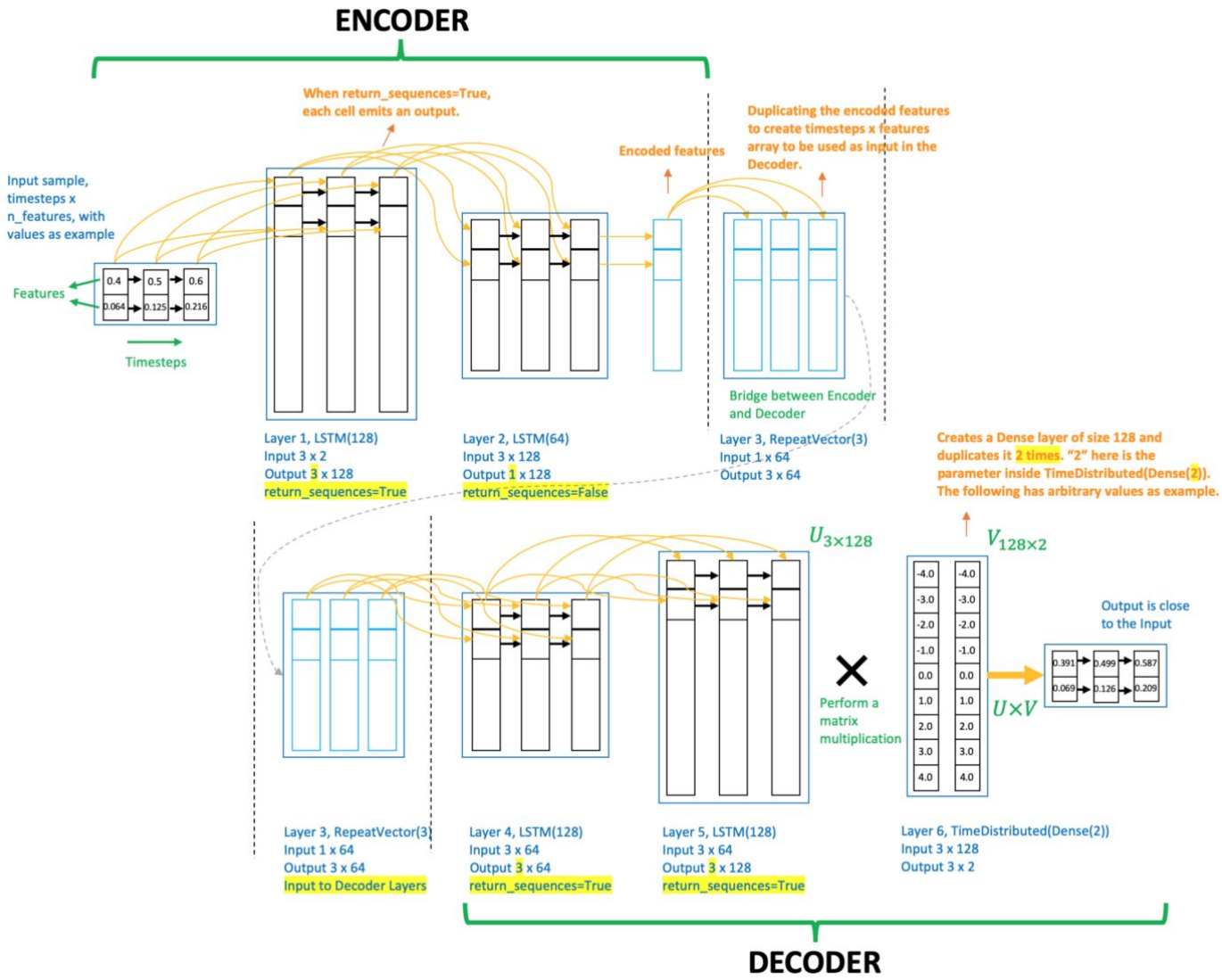
IDEA! proyect! (only with images, something basic CNN Autoencoder)

https://www.alanzucconi.com/2018/03/14/introduction-to-deepfakes/

https://engineering.purdue.edu/~dgueraco/content/deepfake.pdf

https://www.researchgate.net/post/Does_anyone_know_of_a_downloadable_large_faces_dataset

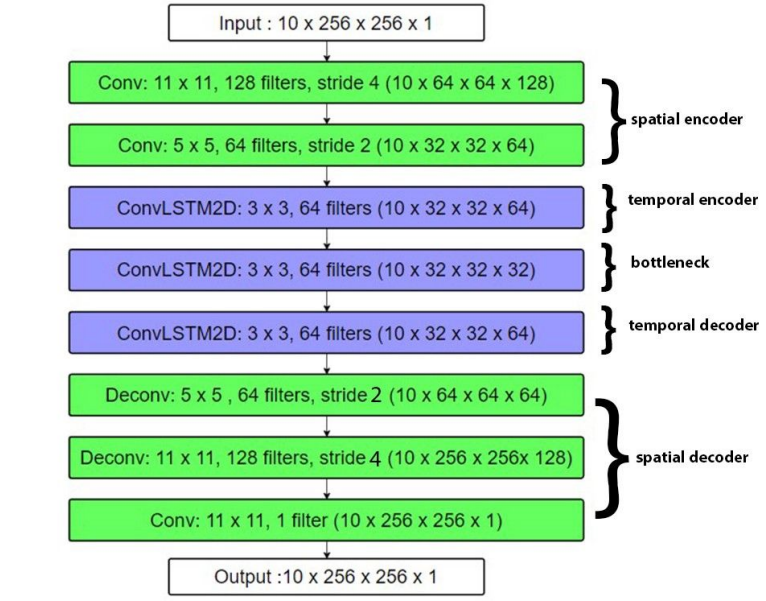
LSTM auto-encoder (ejercicio de clase)



Conv LSTM auto-encoder

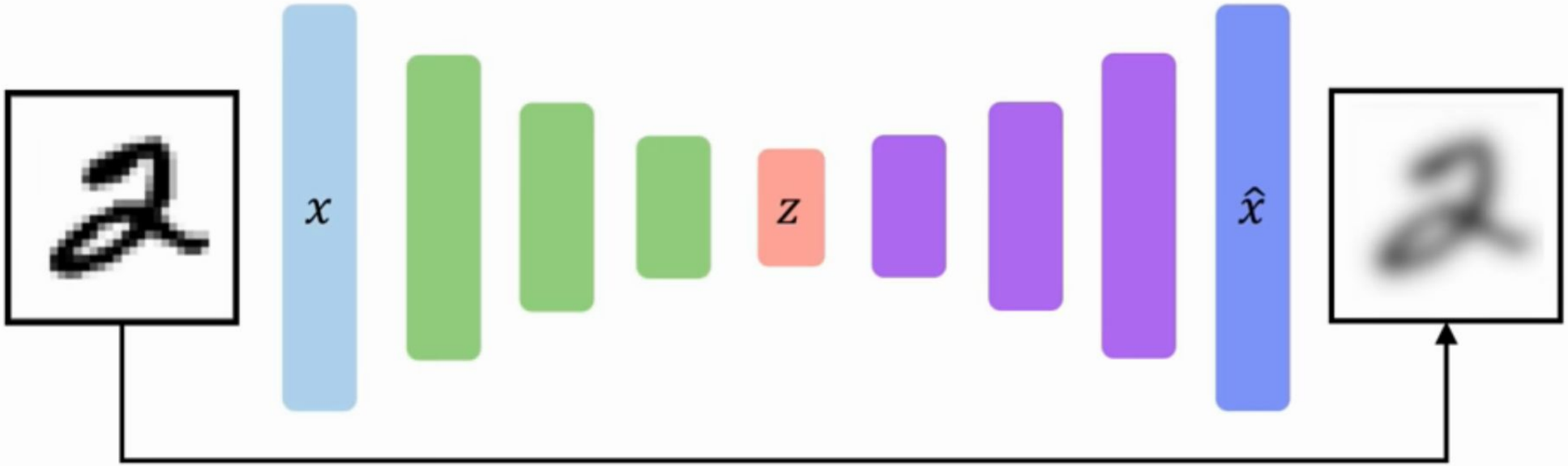
Mondragon Unibertsitatea
Faculty of Engineering

- Muy usado en video, cuando hay que tener en cuenta información entre frames:
  a. entra un frame, y pasa por la parte convolucional, para reducir imagen de entrada, y sacar características
  b. utiliza los layers ya preparados de keras de convolutional LSTMs



Variational Auto-encoders

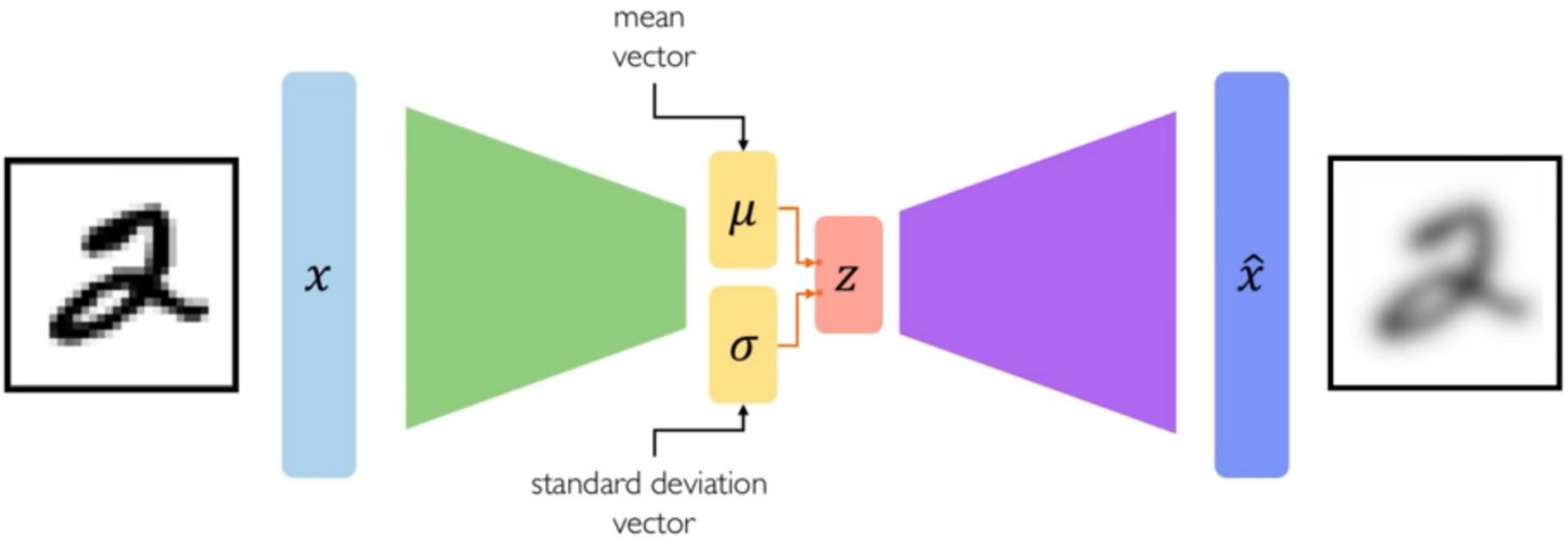
How can we learn this latent space?
Train the model to use these features to **reconstruct the original data**



$$\mathcal{L}(x, \hat{x}) = \|x - \hat{x}\|^2$$

Loss function doesn't use any labels!!

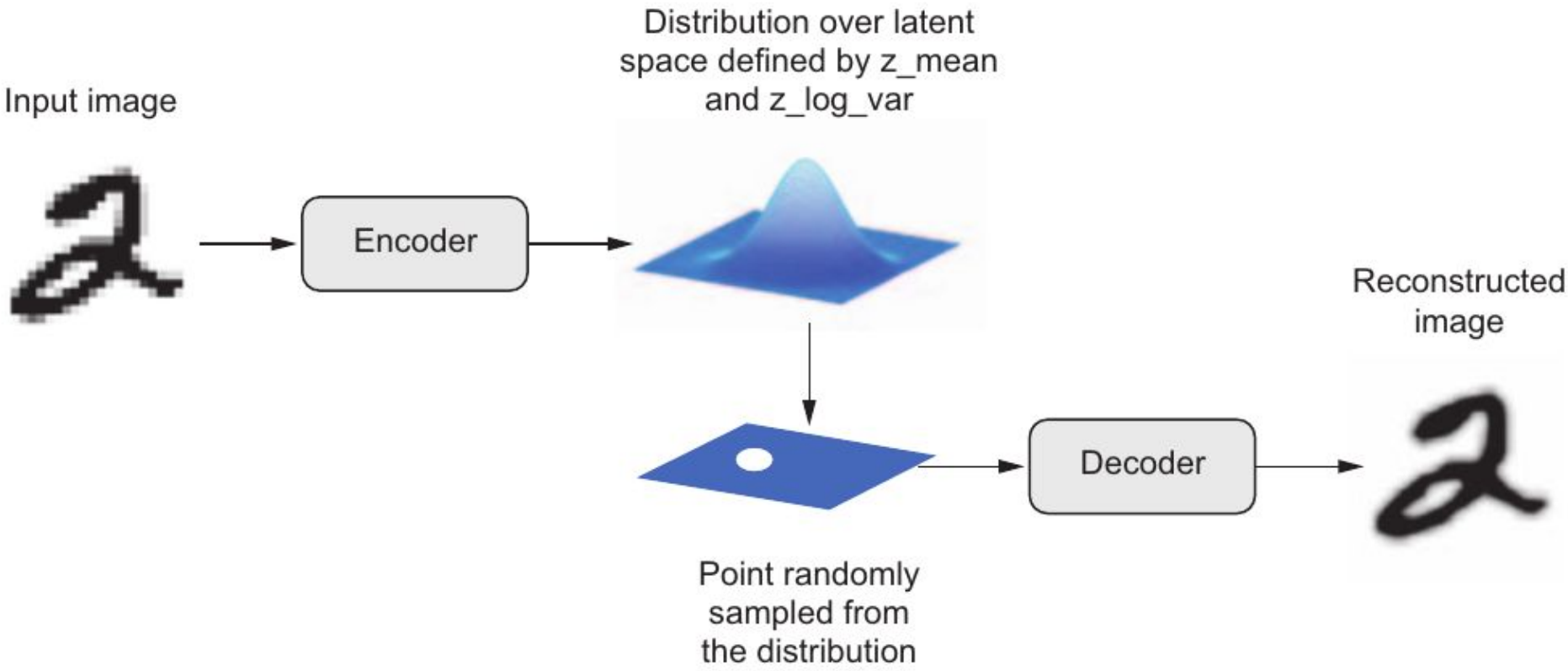
VAE



Variational autoencoders are a probabilistic twist on autoencoders!

Sample from the mean and standard dev. to compute latent sample

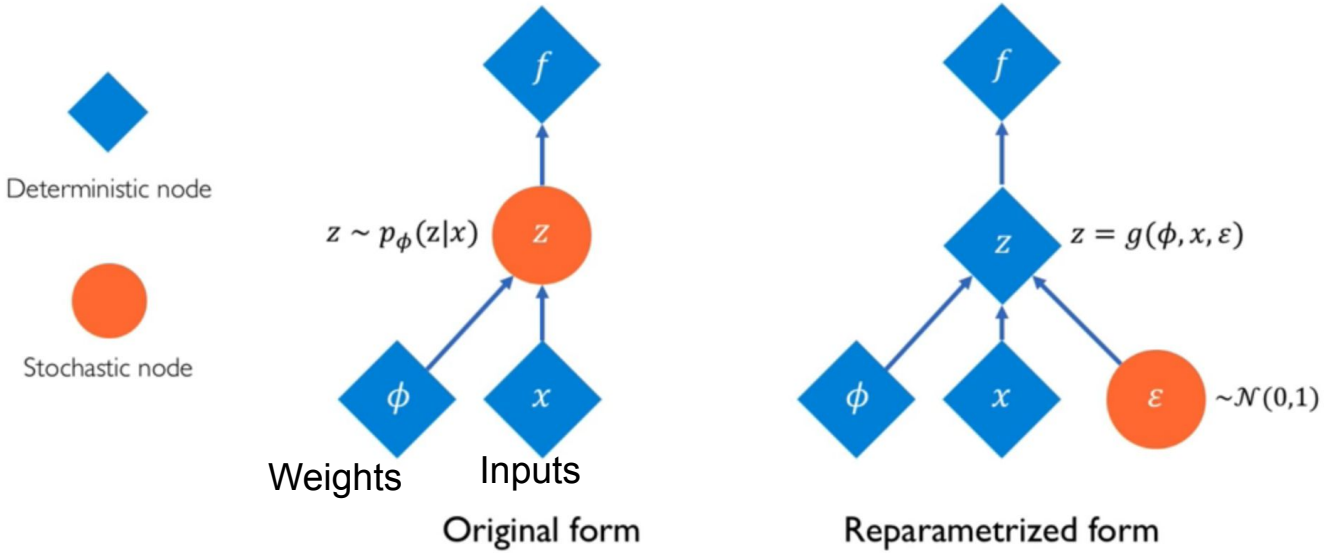
VAE


Input image
Encoder
Distribution over latent space defined by z_mean and z_log_var
Point randomly sampled from the distribution
Decoder
Reconstructed image

PROBLEM!!!!

- We can not BACPROPAGATE error through a stochastic LAYER!

Reparametrizing the sampling layer



Key Idea:

~~$z \sim \mathcal{N}(\mu, \sigma^2)$~~

Consider the sampled latent vector z as a sum of

- a fixed μ vector,
- and fixed σ vector, scaled by random constants drawn from the prior distribution

$$\Rightarrow z = \mu + \sigma \odot \varepsilon$$

where $\varepsilon \sim \mathcal{N}(0,1)$

VAE

- each of the latent variables mean something, sometimes it is human interpretable, other times not.
- We can sample changing only one latent variable:

Slowly increase or decrease a **single latent variable**

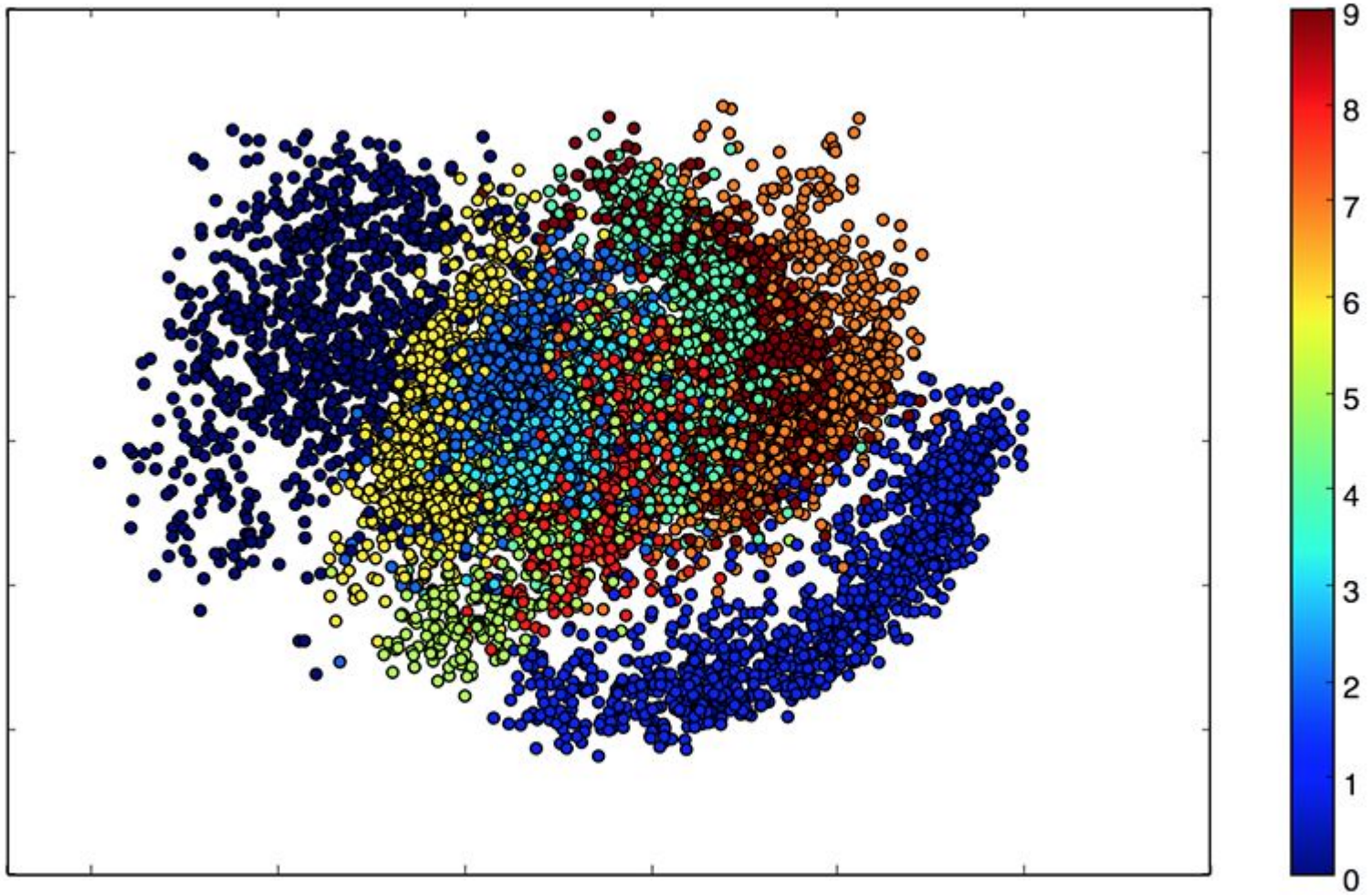
Keep all other variables fixed

Head pose

Different dimensions of z encodes **different interpretable latent features**

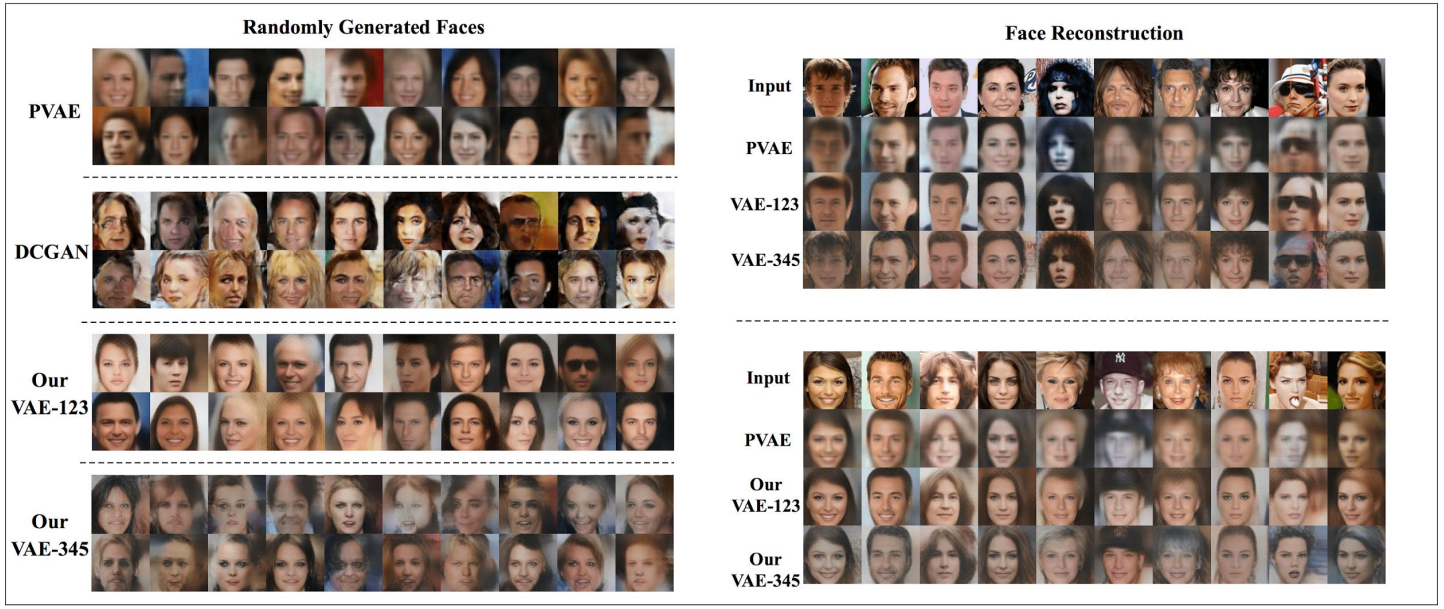
VAE

- As normal AutoEncoders, we can also plot the results of the encoding in the latent space.



VAE

- Once we have our Latent Space distribution modelled, we can start creating new data, sampling random samples of the latent space, and passing them through the DECODER.
- It is being used for data augmentation.



https://houxianxu.github.io/assets/project/dfcvae

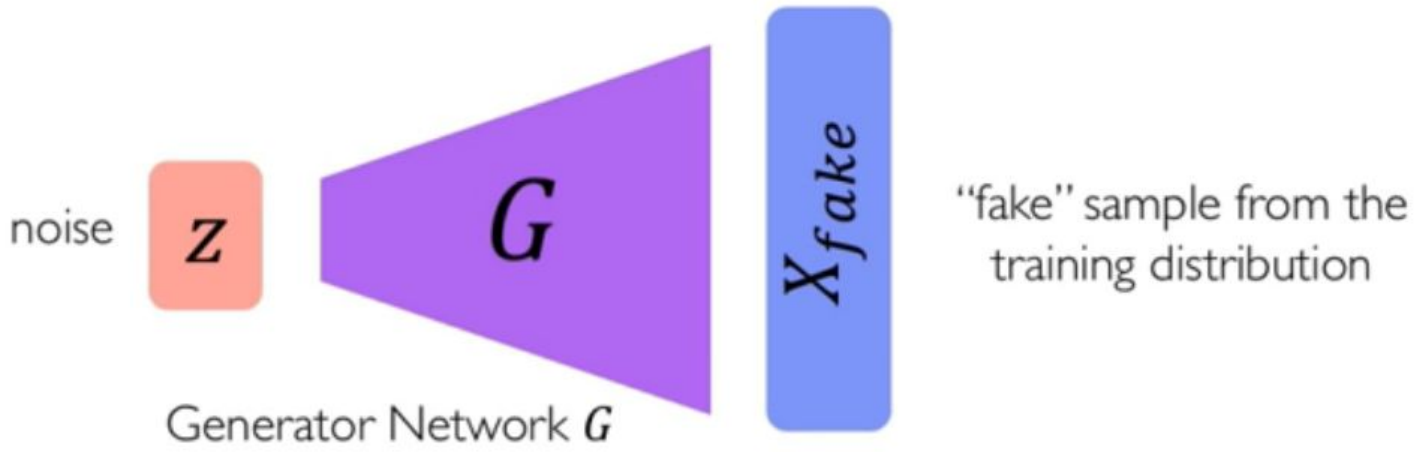
IDEA! proyect! Generate something (music, images, …) with VAE (can be CNNVAE, LSTMVAE….)

Generative Adversarial Networks

Idea: don't explicitly model density, and instead just sample to generate new instances.

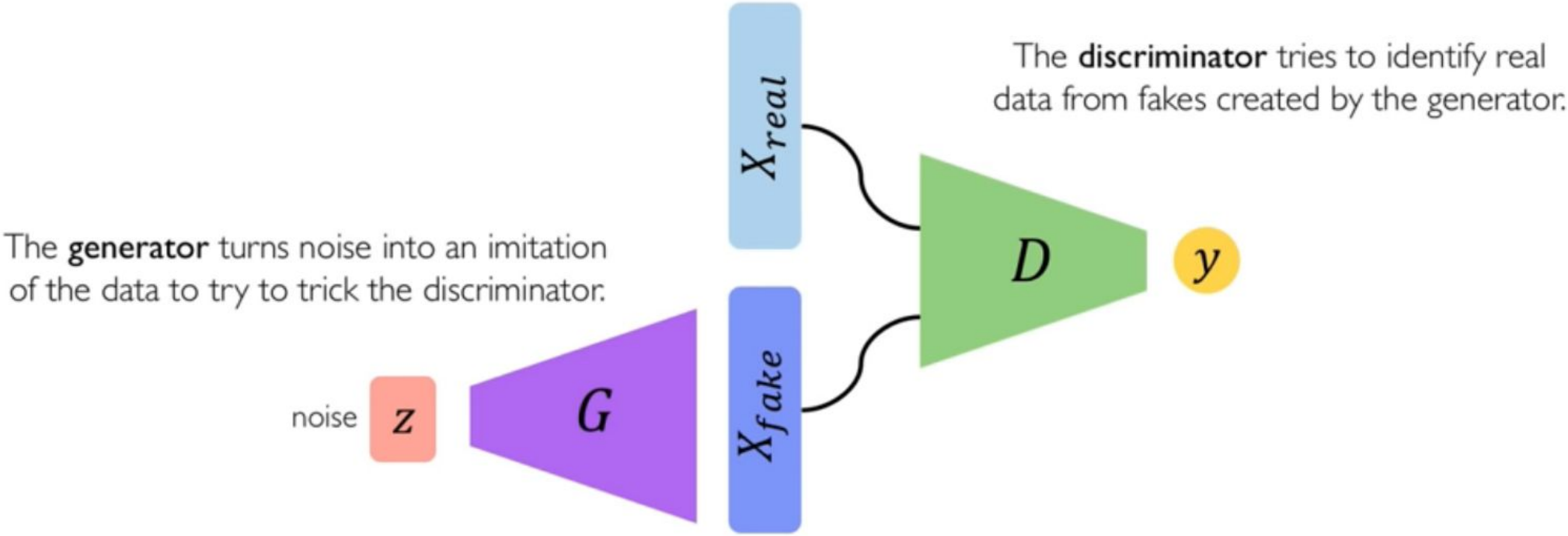
Problem: want to sample from complex distribution – can't do this directly!

Solution: sample from something simple (noise), learn a transformation to the training distribution.

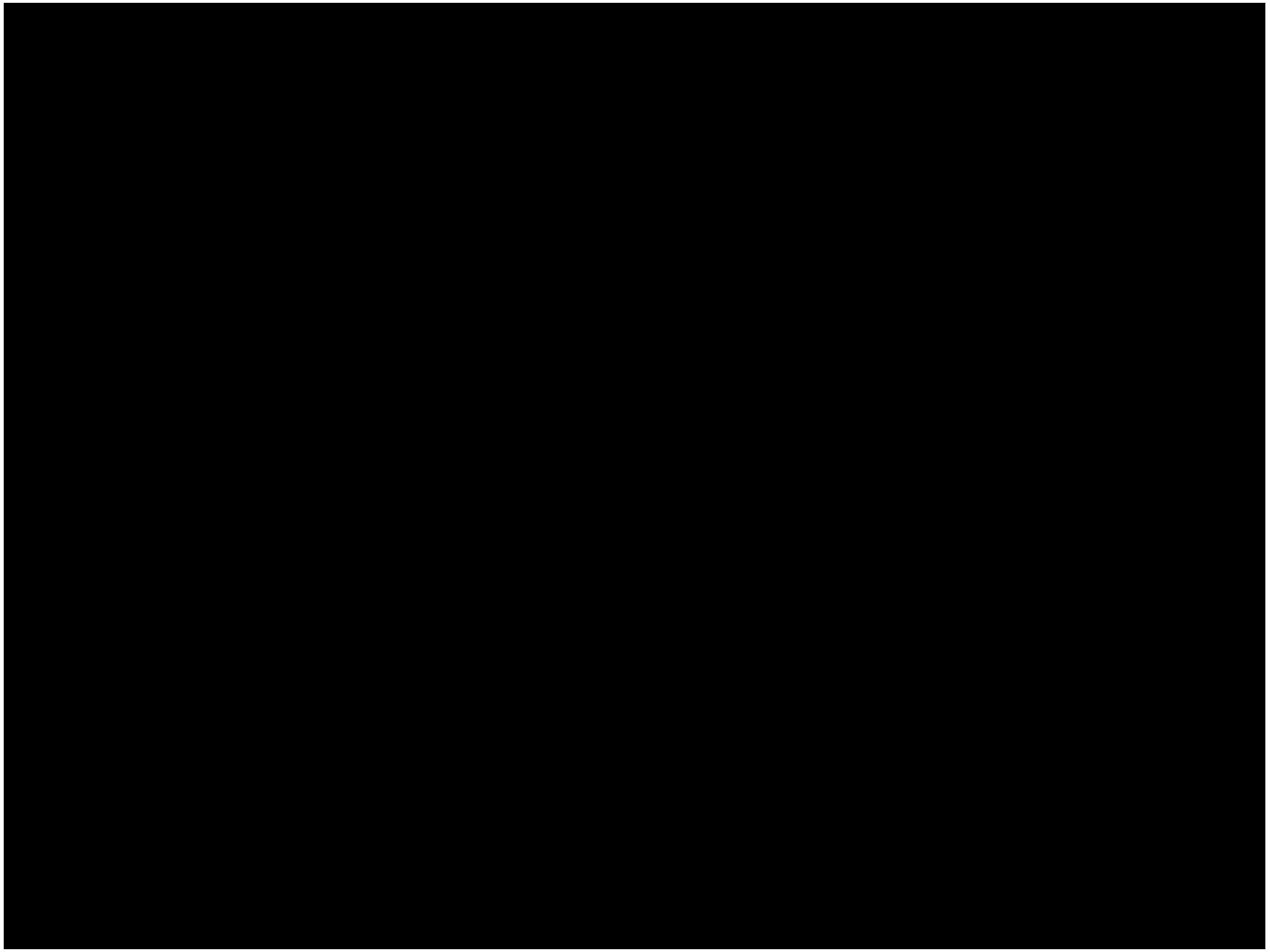


GANs

Generative Adversarial Networks (GANs) are a way to make a generative model by having two neural networks compete with each other.

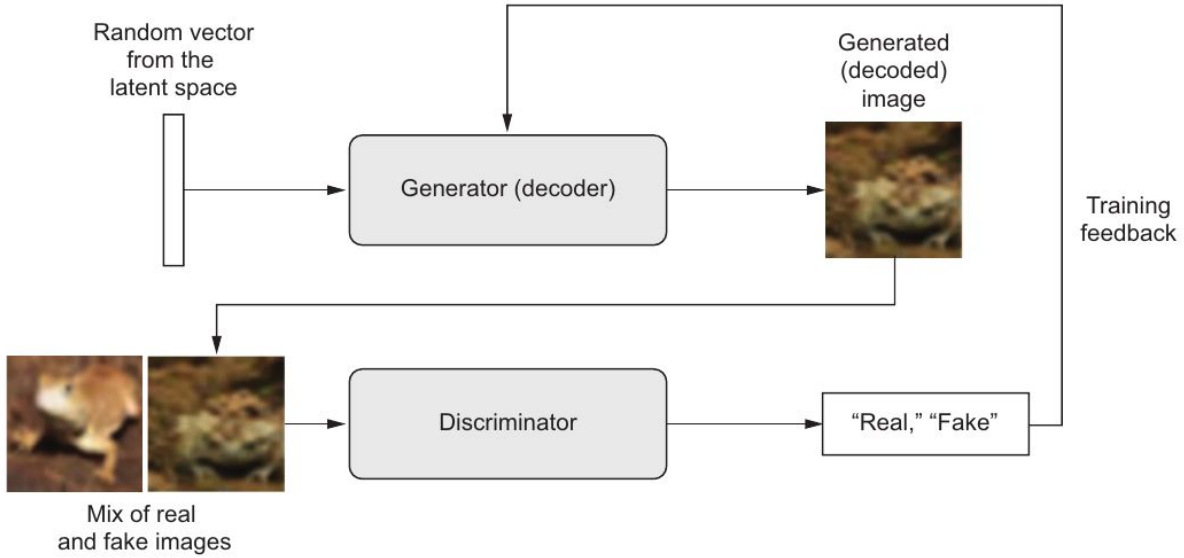


GANs



GANs Learning phase

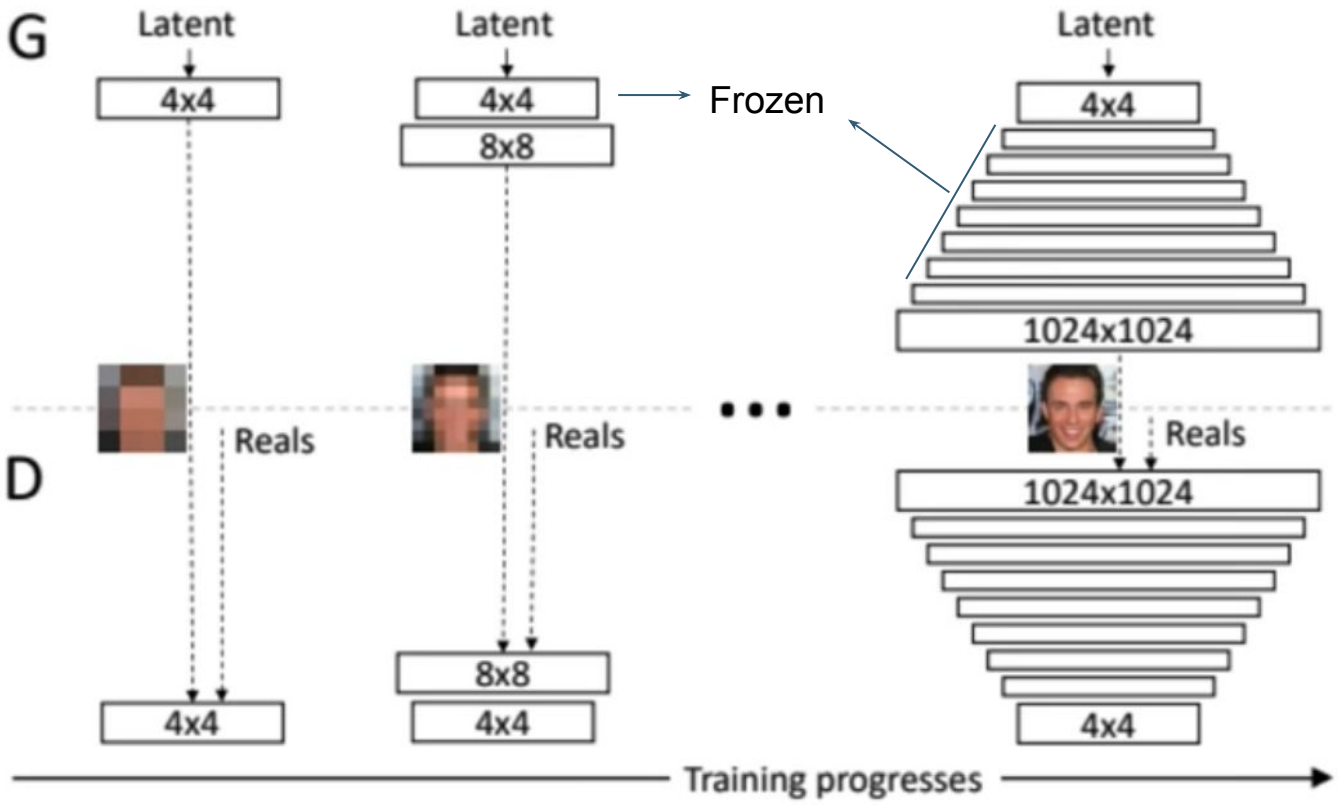
- there are 2 NNs:
  - Generator -> creates images from noise
  - Discriminator -> compares real and fake images, and tries to distinguish between them
- Generator starts with random noise, and generates samples, that are presented to the discriminator.
- Every step, the generator is trained to be able to fool the discriminator, and evolves generating more realistic samples
- at the end, the discriminator is not able to distinguish between real and fakes.

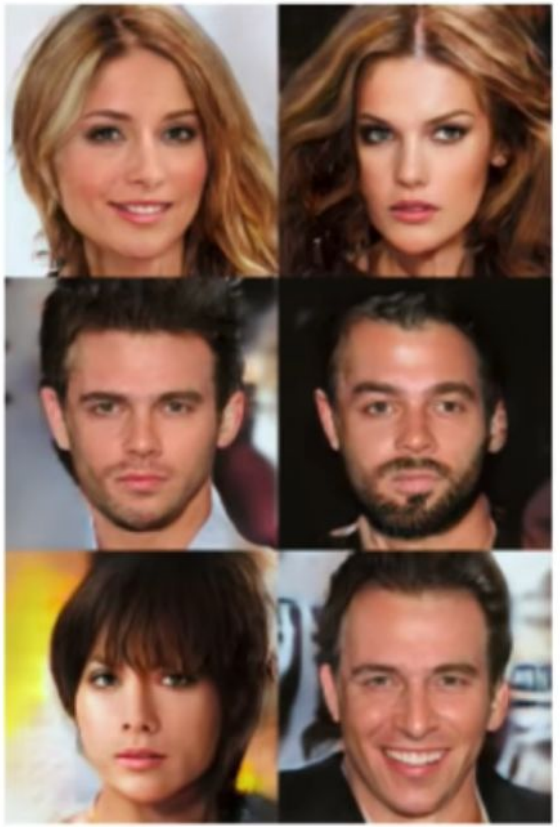


GANs Learning phase

1. Draw random points in the latent space (random noise).
2. Generate images with generator using this random noise.
3. Mix the generated images with real ones.
4. Train discriminator using these mixed images, with corresponding targets: either “real” (for the real images) or “fake” (for the generated images).
5. Draw new random points in the latent space.
6. Train gan using these random vectors, with targets that all say “these are real images.” This updates the weights of the generator (only, because the discriminator is frozen inside gan) to move them toward getting the discriminator to predict “these are real images” for generated images: this trains the generator to fool the discriminator.

Progressive growing of GANs (NVIDIA)





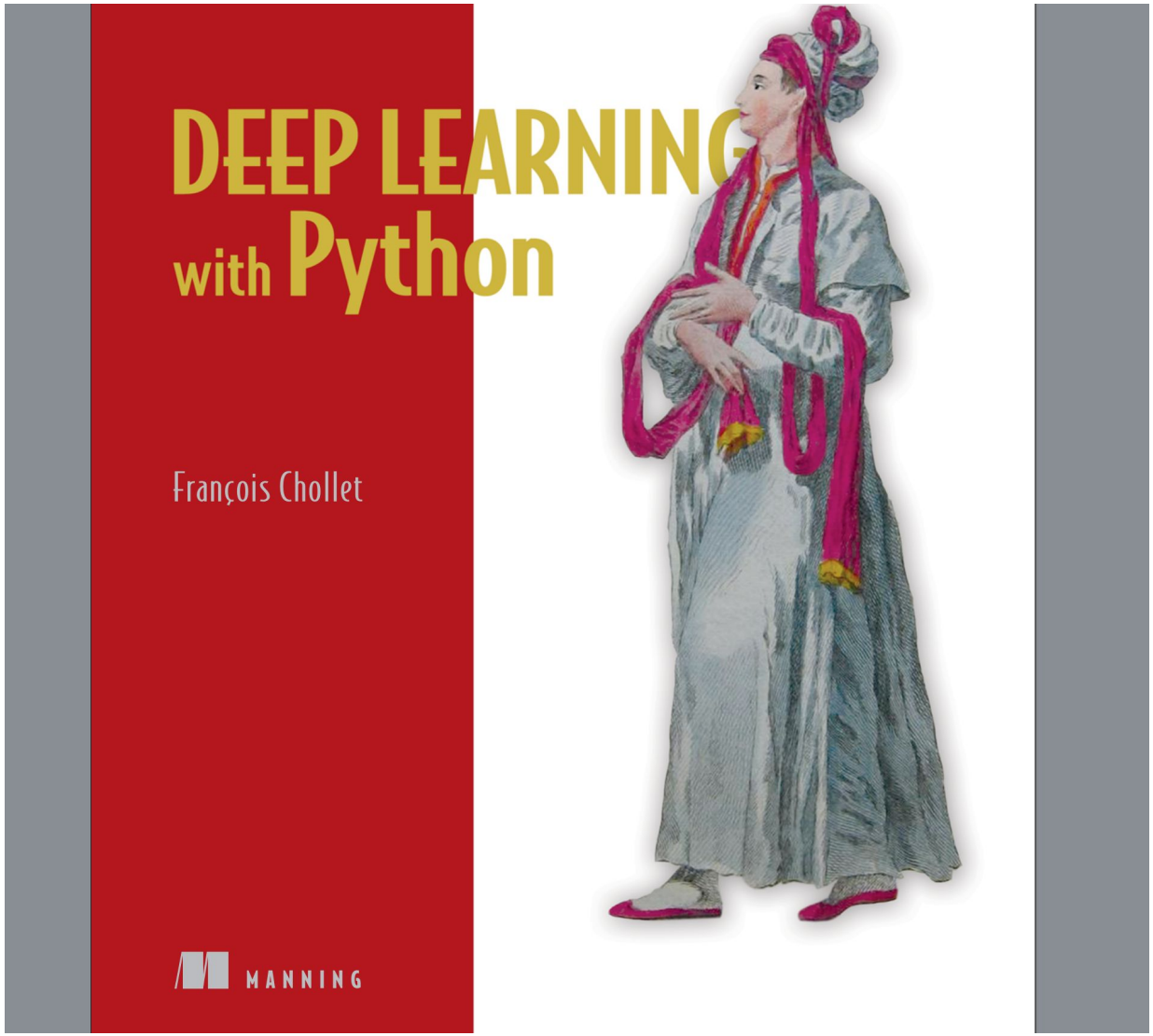
GAN - Ejemplo Friki (para may the 4th)

Mondragon Unibertsitatea
Faculty of Engineering

SharinGAN: generador de Sharingans

https://reddit.com/r/deeplearning/comments/gc6786/ai_generates_a_new_sharingan_using_gan_to/

GAN how to implement

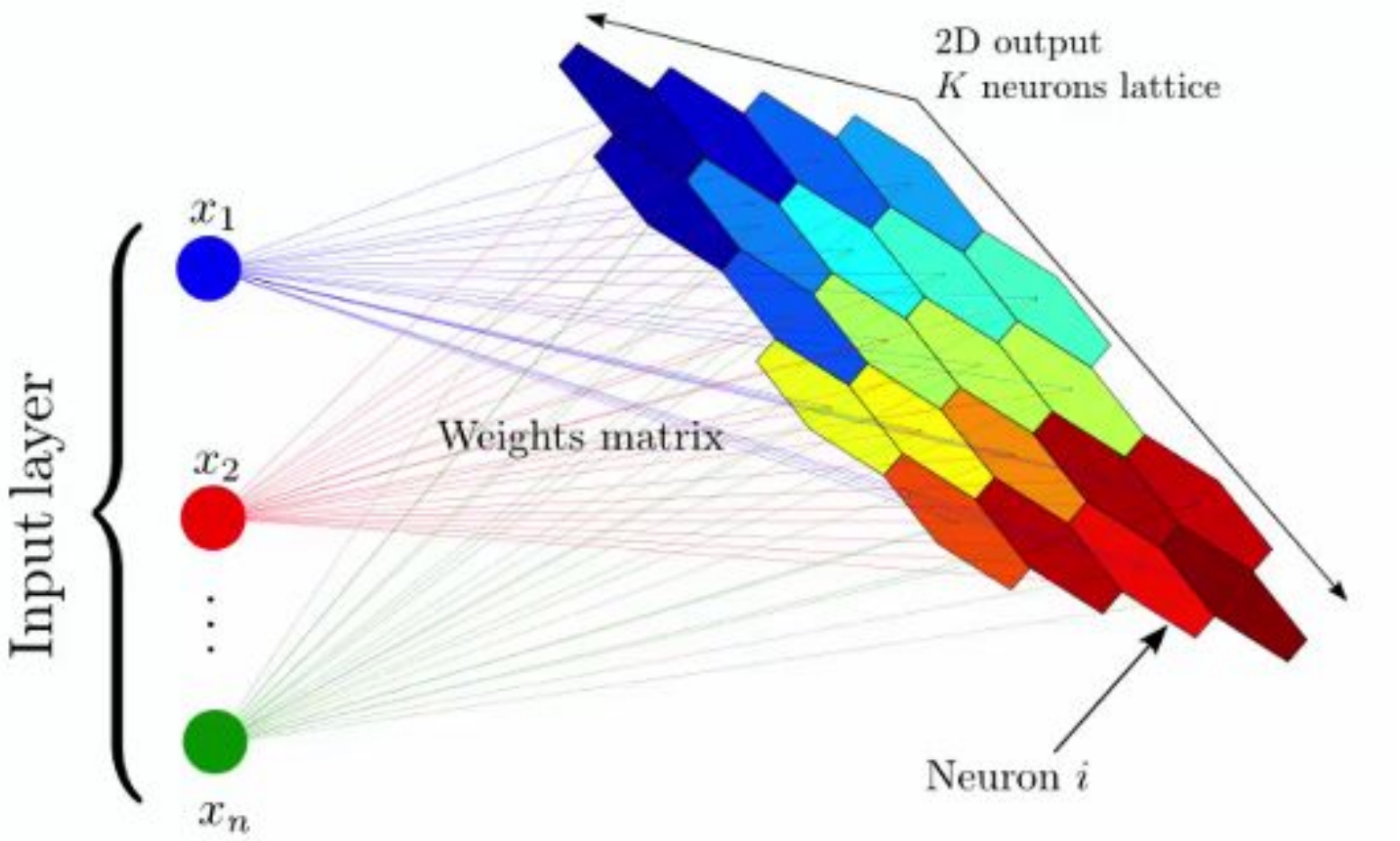


IDEA!
proyect!
Same as VAE

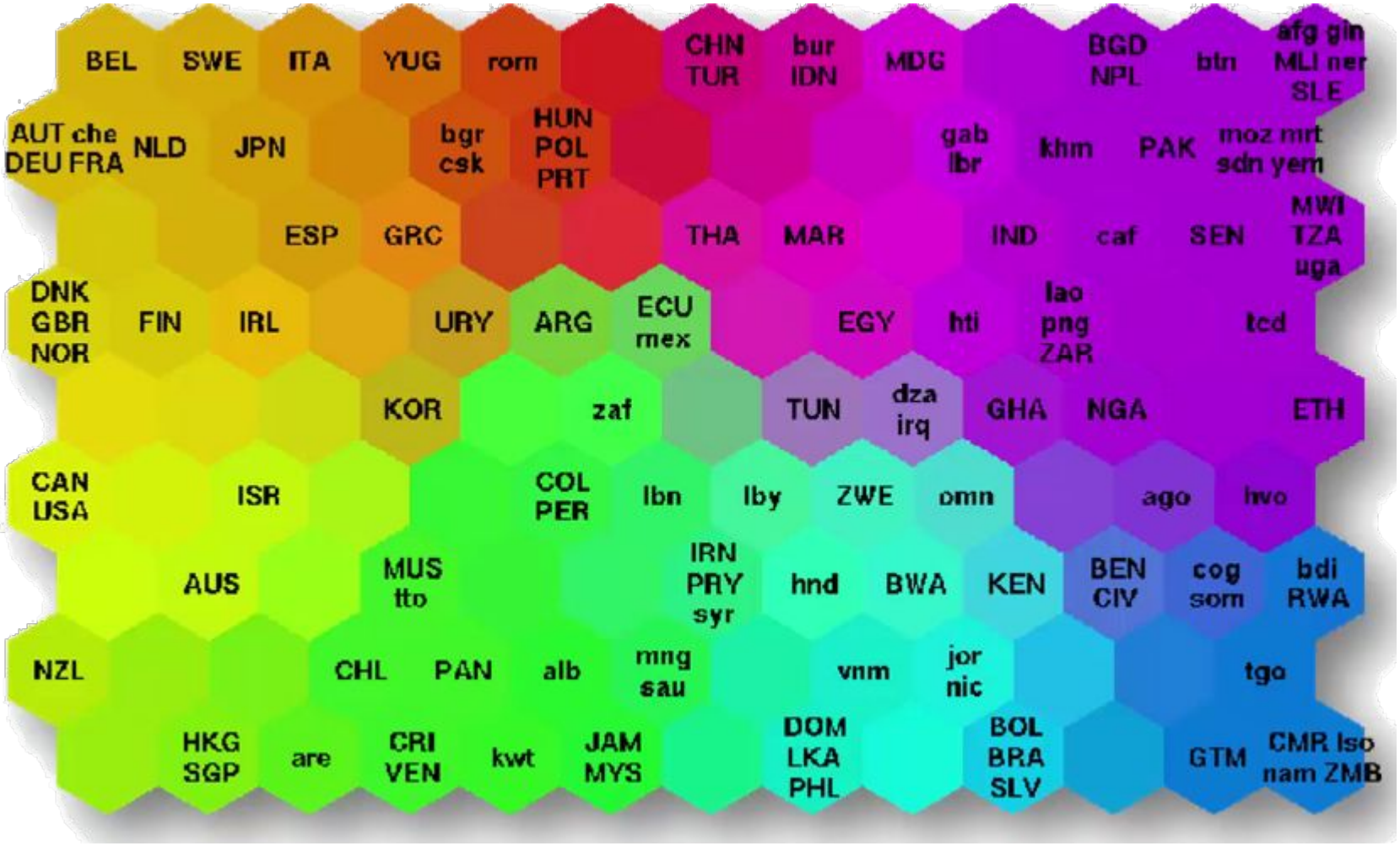
Self Organizing Maps
SOM

Teuvo Kohonen

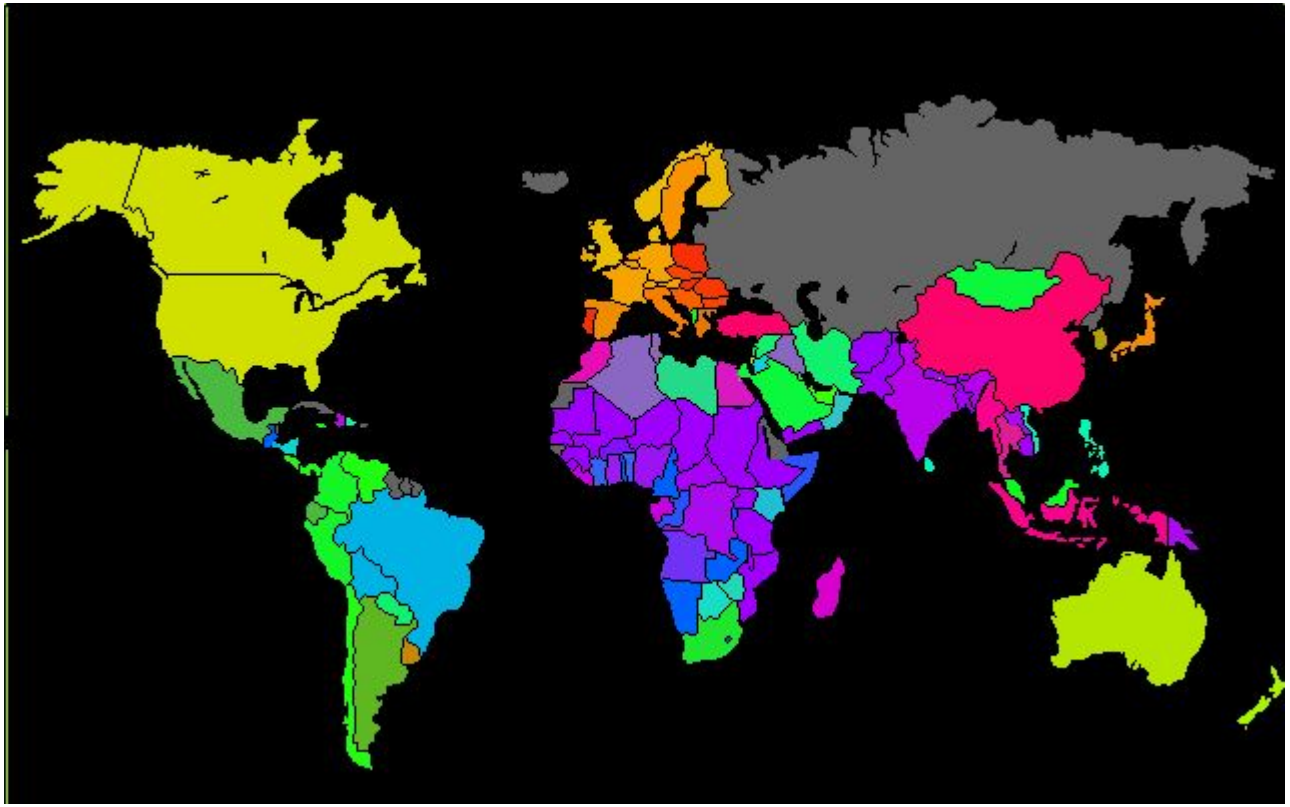
¿Que son?


2D output
K neurons lattice
x_1
x_2
x_n
Input layer
Weights matrix
Neuron i

Ejemplo: clasificación de países



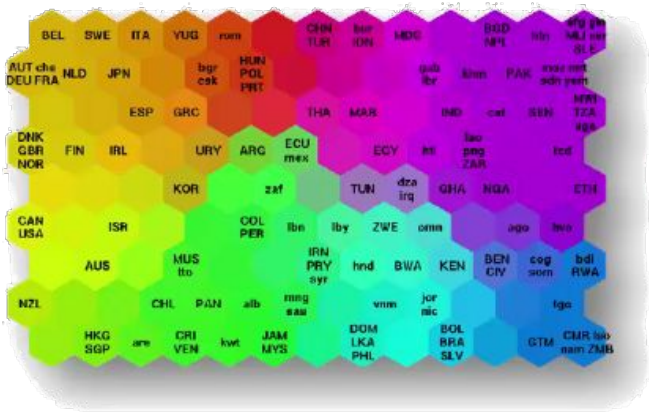
Ejemplo: clasificación de países

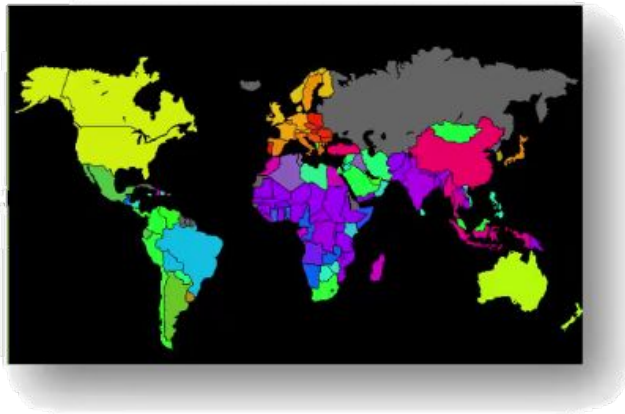


Ejemplo: clasificación de países

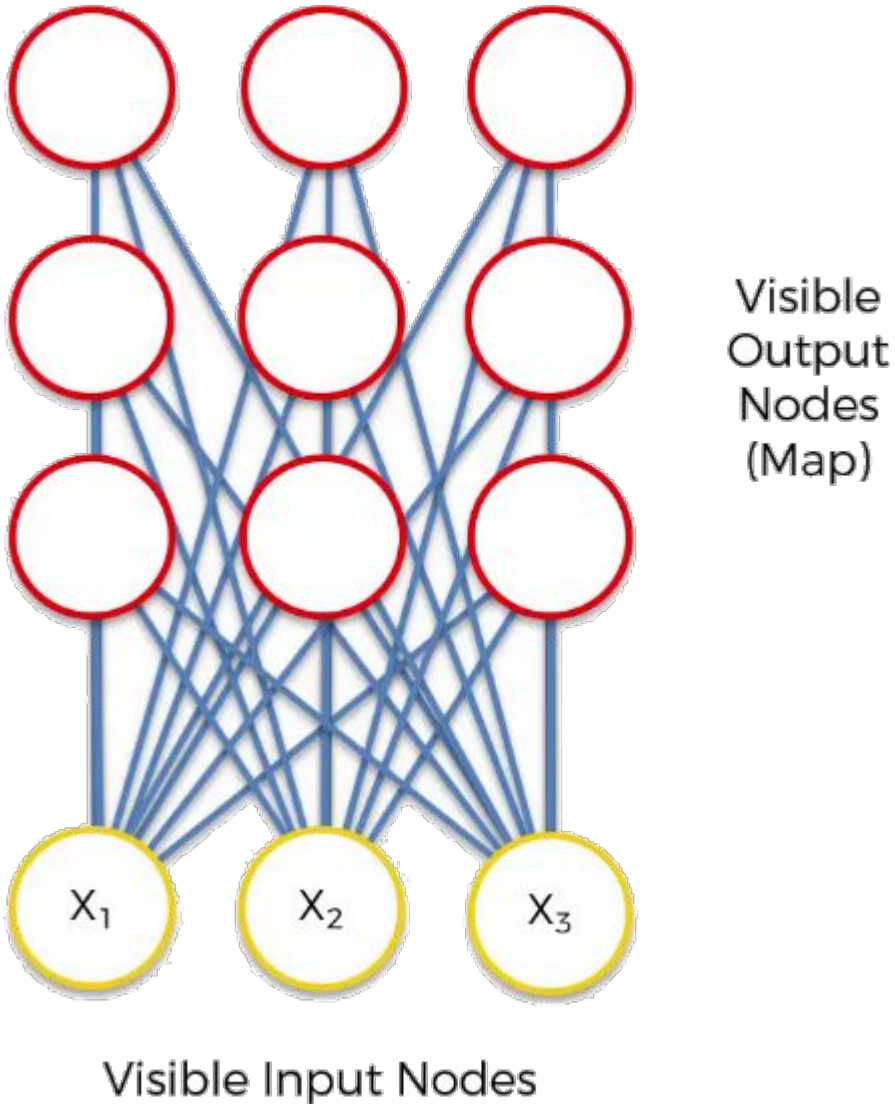
Mondragon Unibertsitatea
Faculty of Engineering

	A	B	C	D	E
1	Country	Country C	Health Ex	Education E	Inflation
2	Aruba	ABW	9.418971	5.92467022	-2.13637
3	Afghanist	AFG	4.371774		-8.28308
4	Angola	AGO	5.791339		13.73145
5	Albania	ALB	6.75969		2.280502
6	Andorra	AND	4.57058	3.1638701	
7	Arab Wor	ARB	4.049924		3.524814
8	United Ar	ARE	7.634758		
9	Argentina	ARG	4.545323	4.88997984	6.282774
10	Armenia	ARM		3.84079003	3.406767
11	American	ASM	4.862062		
12	Antigua a	ATG	9.046056	2.55447006	-0.55016
13	Australia	AUS	11.19444	5.09262991	1.820112
14	Austria	AUT	5.85024	5.7674098	0.506313
15	Azerbaija	AZE	6.964187	3.22430992	1.401056
16	Burundi	BDI	10.39434	6.3197999	10.98147
17	Belgium	BEL	4.46431	6.41535997	-0.05315
18	Benin	BEN	7.405431	4.22204018	2.15683

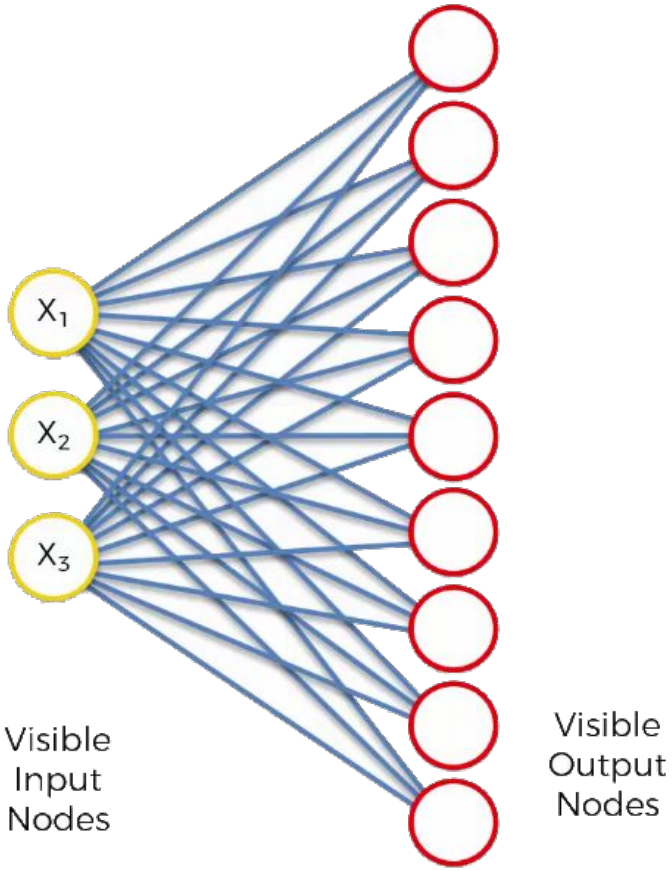




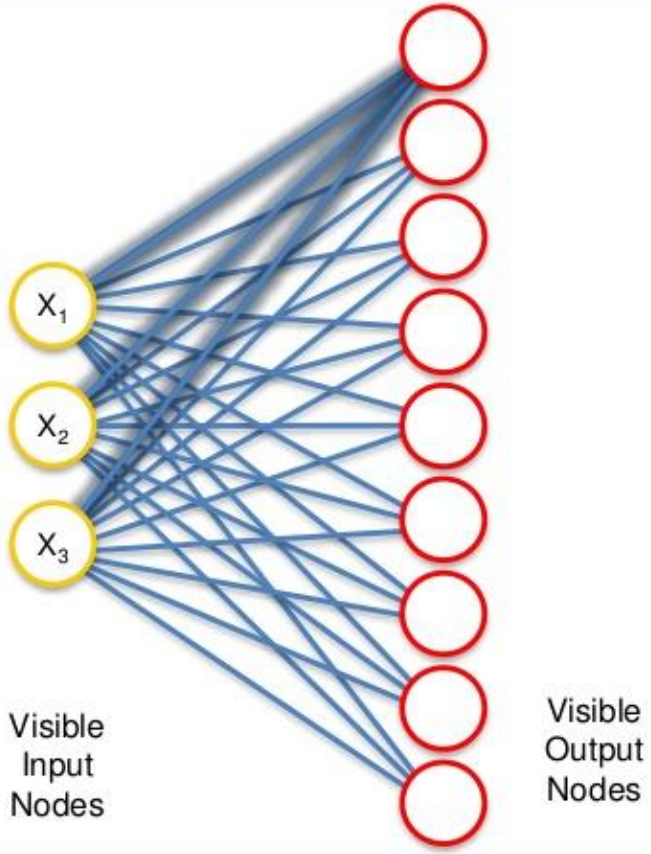
¿Cómo aprenden los SOM?



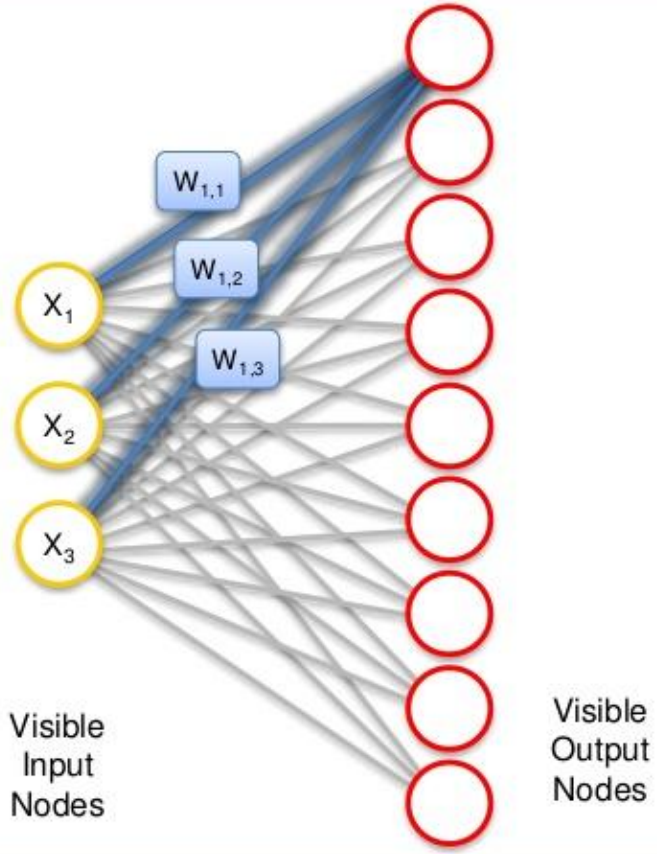
¿Cómo aprenden los SOM?



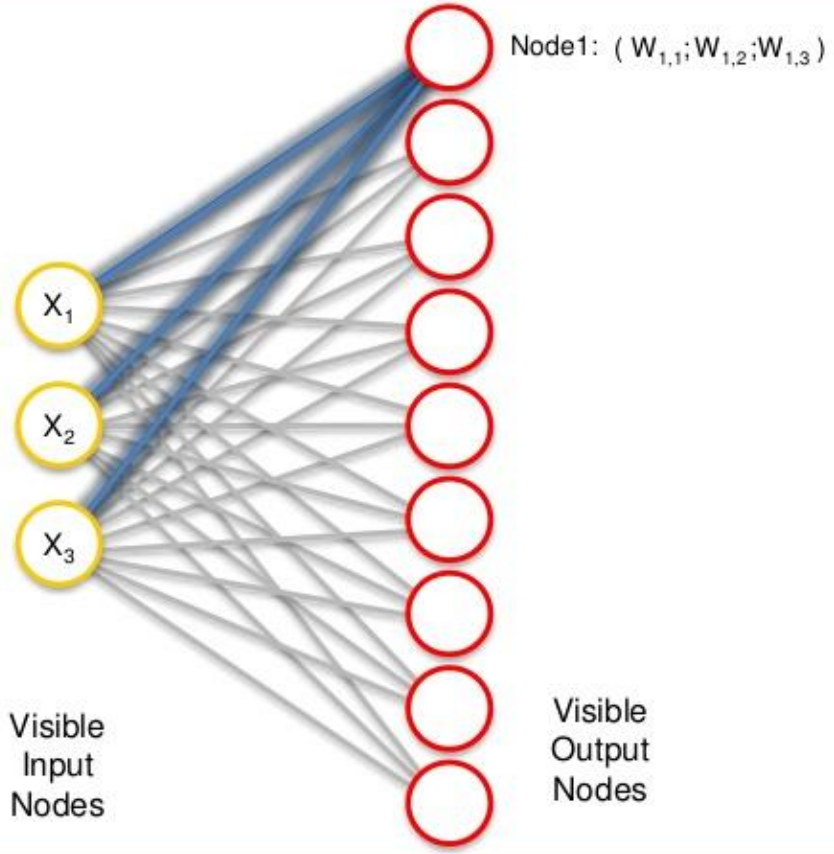
¿Cómo aprenden los SOM?



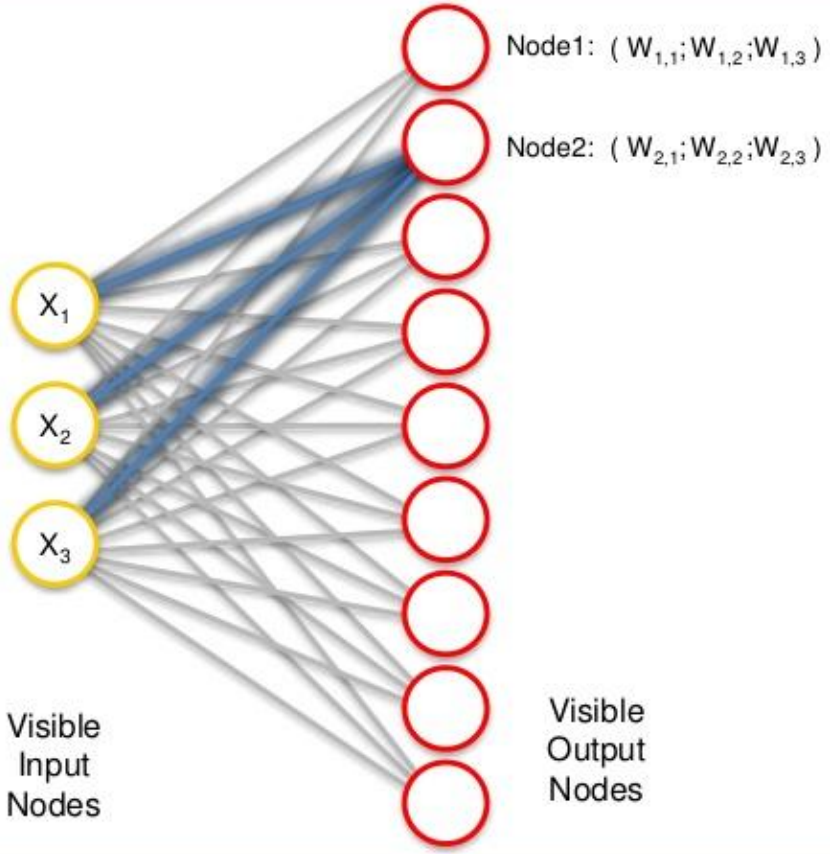
¿Cómo aprenden los SOM?



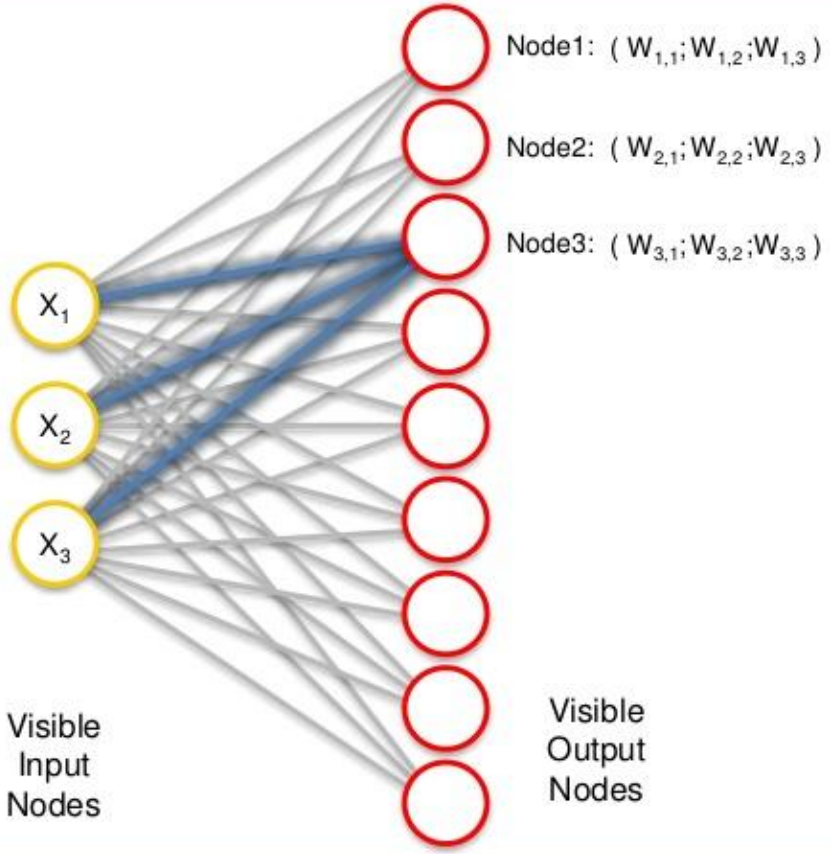
¿Cómo aprenden los SOM?



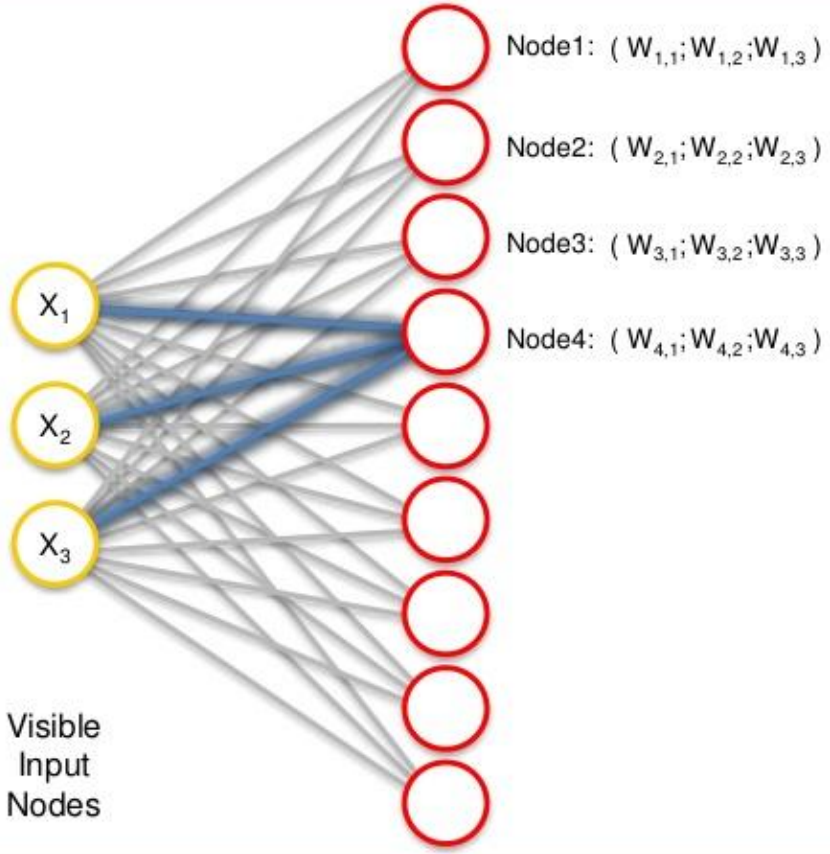
¿Cómo aprenden los SOM?



¿Cómo aprenden los SOM?



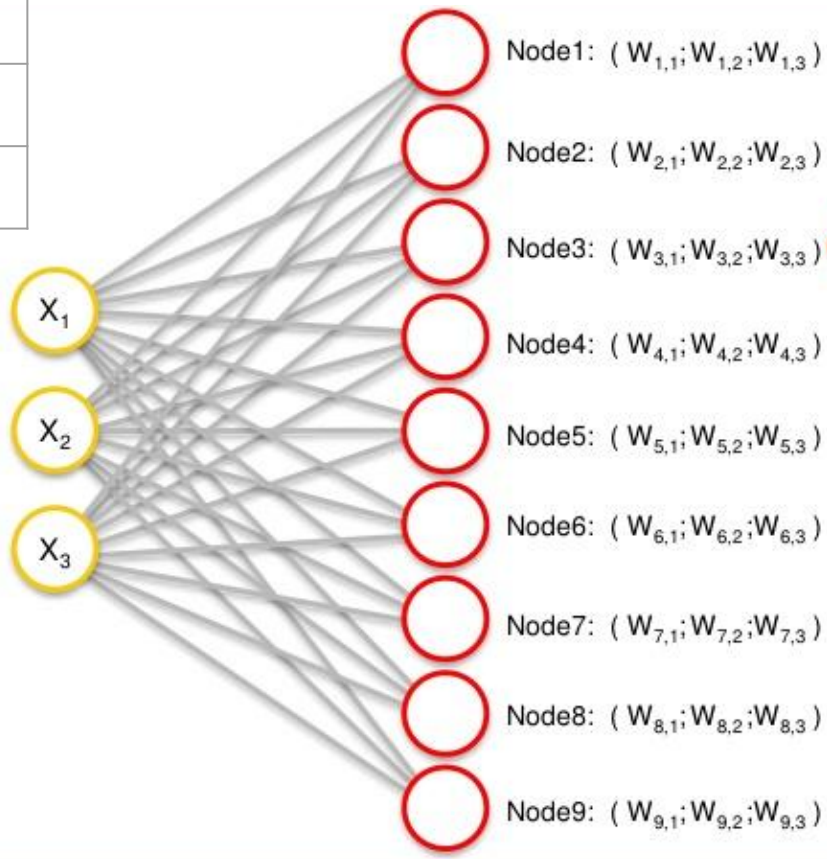
¿Cómo aprenden los SOM?



¿Cómo aprenden los SOM?

Mondragon Unibertsitatea
Faculty of Engineering

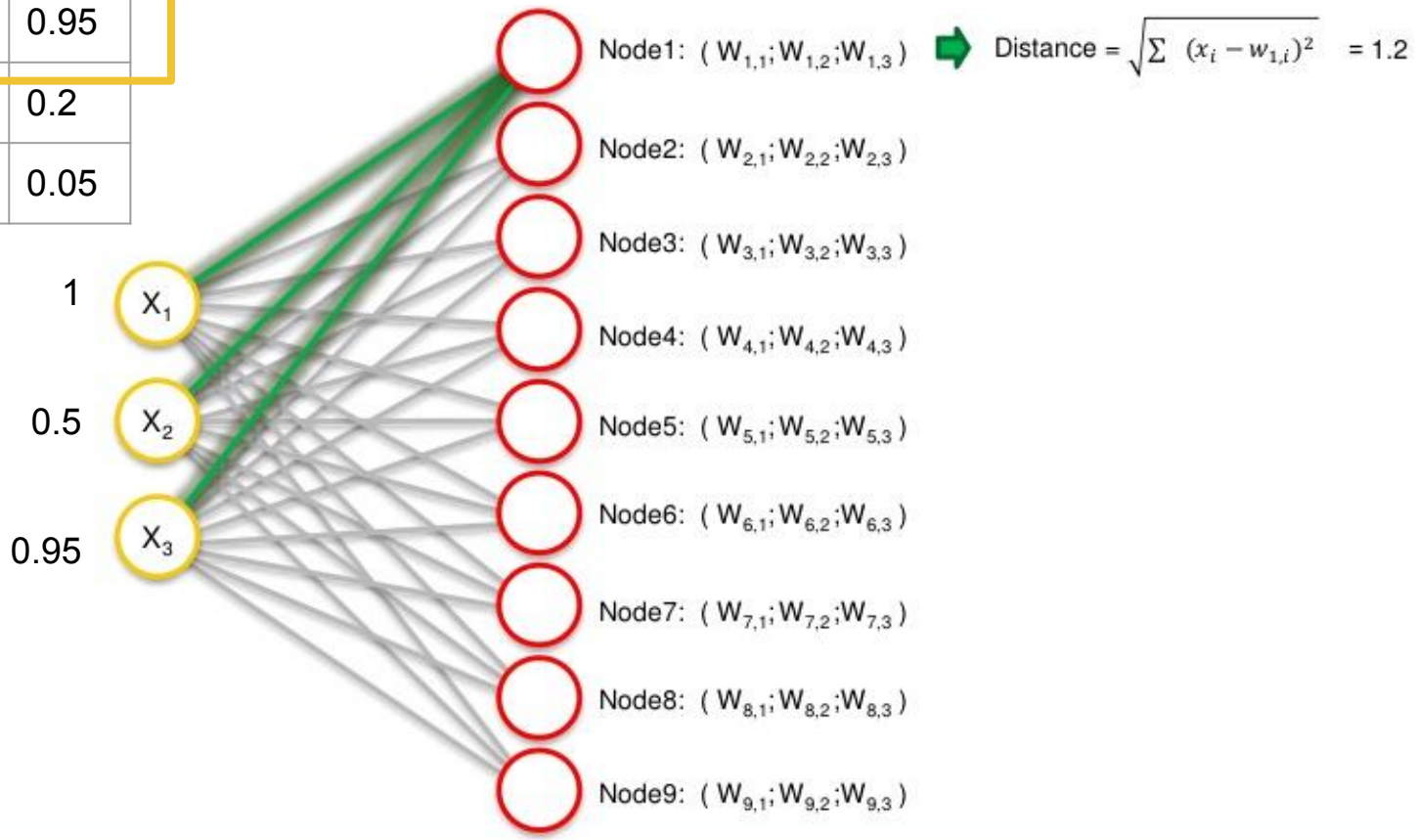
filaN	var1	var2	var3
1	1	0.5	0.95
2	0.5	0	0.2
3	1	0.75	0.05



¿Cómo aprenden los SOM?

Mondragon Unibertsitatea
Faculty of Engineering

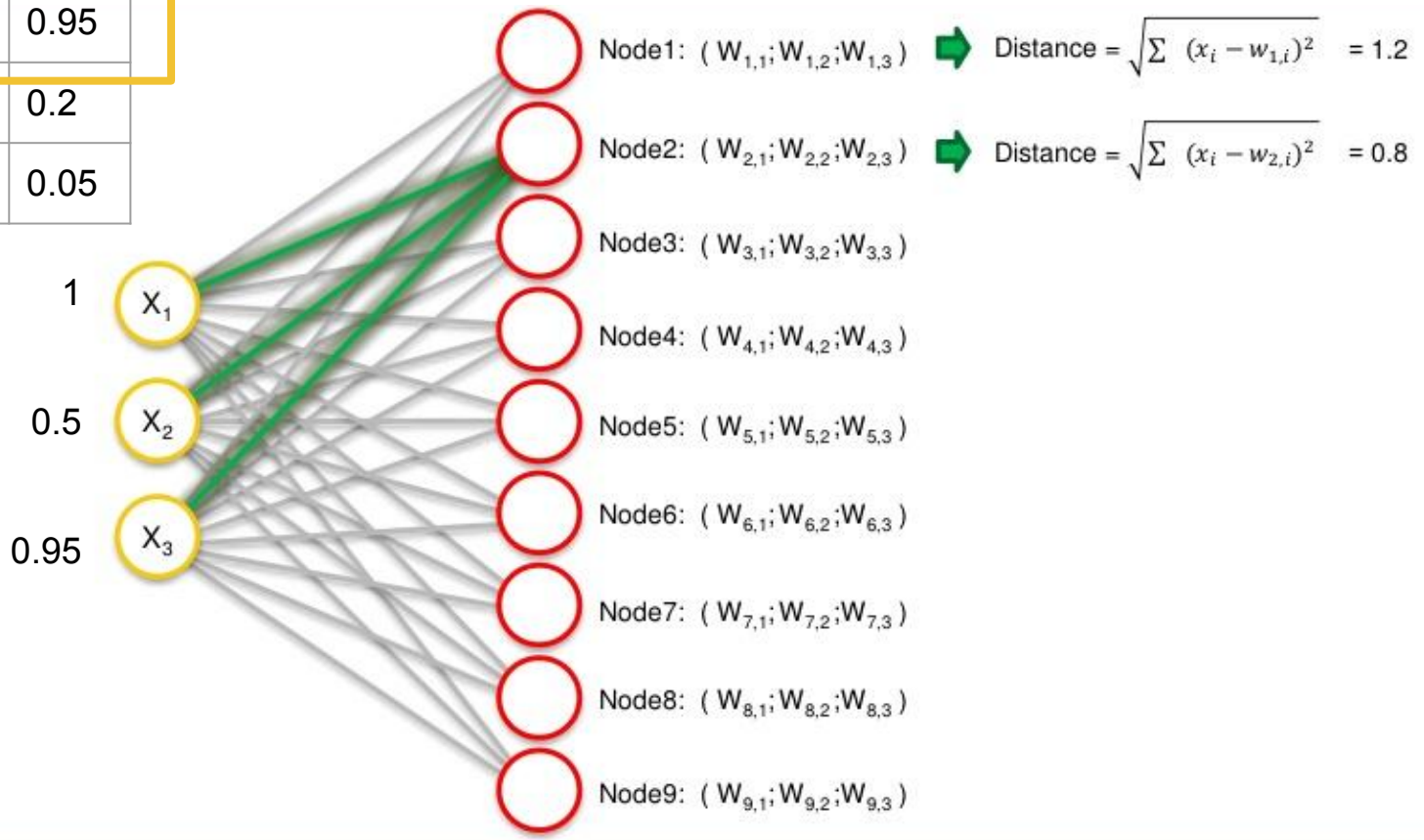
filaN	var1	var2	var3
1	1	0.5	0.95
2	0.5	0	0.2
3	1	0.75	0.05



¿Cómo aprenden los SOM?

Mondragon Unibertsitatea
Faculty of Engineering

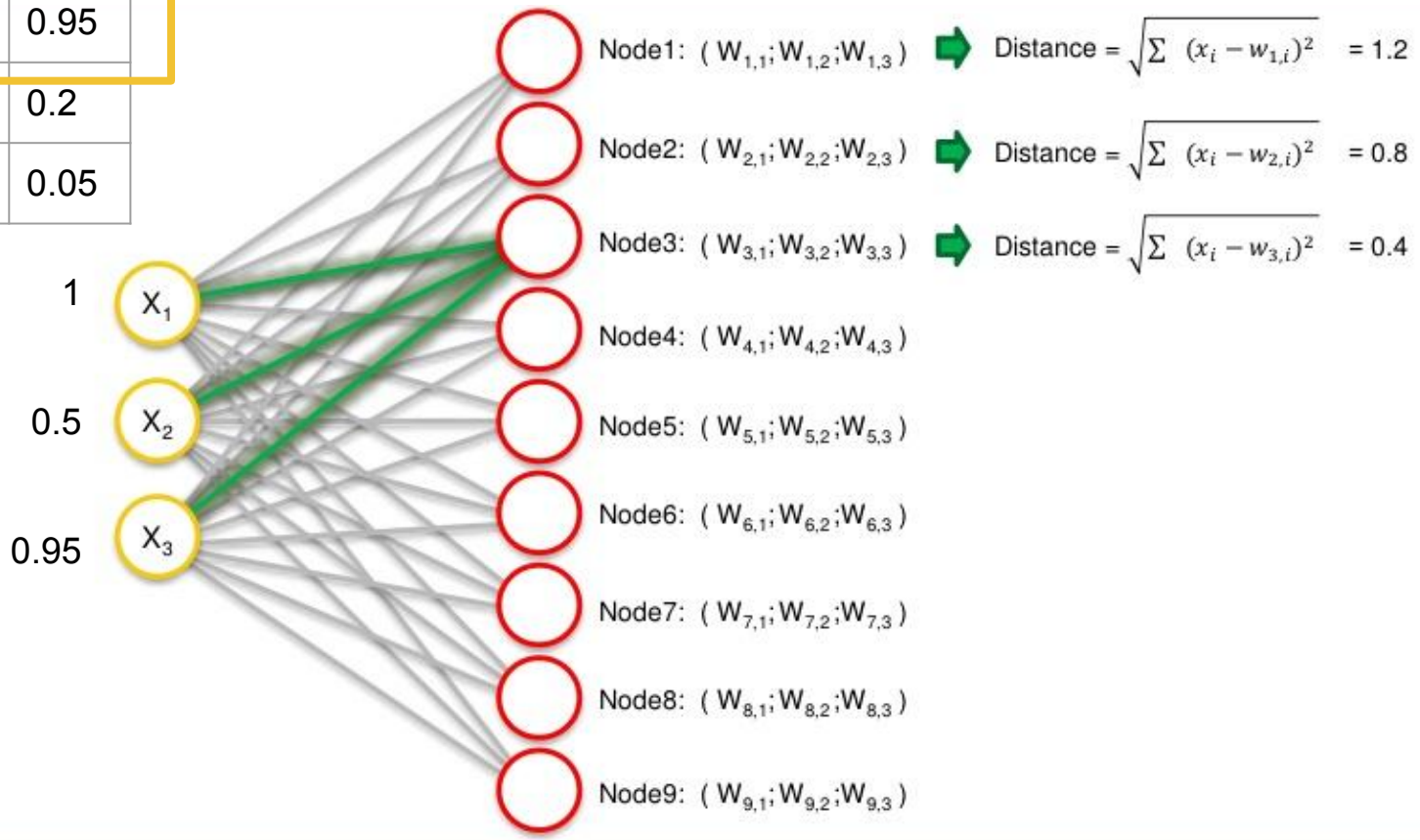
filaN	var1	var2	var3
1	1	0.5	0.95
2	0.5	0	0.2
3	1	0.75	0.05



¿Cómo aprenden los SOM?

Mondragon Unibertsitatea
Faculty of Engineering

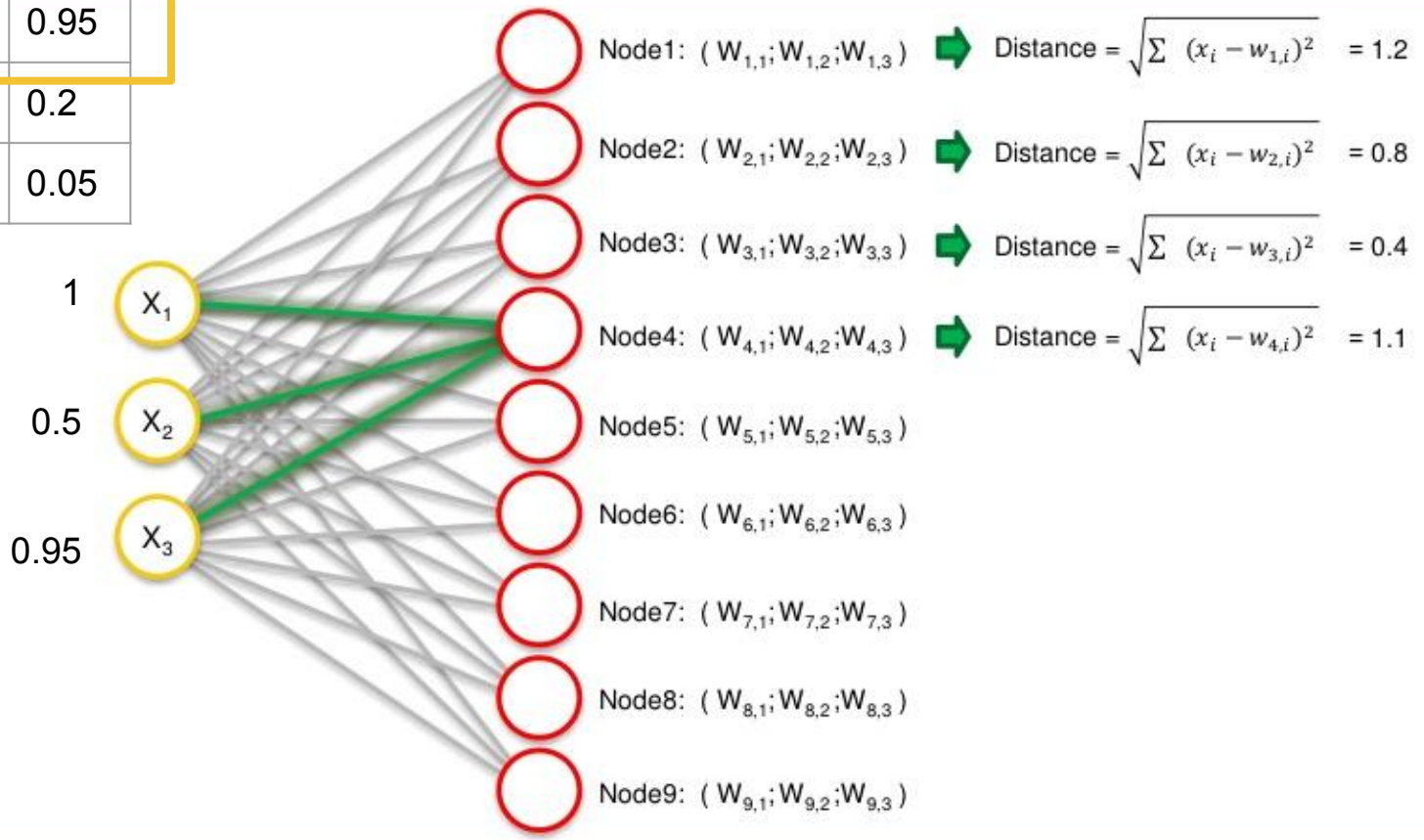
filaN	var1	var2	var3
1	1	0.5	0.95
2	0.5	0	0.2
3	1	0.75	0.05



¿Cómo aprenden los SOM?

Mondragon Unibertsitatea
Faculty of Engineering

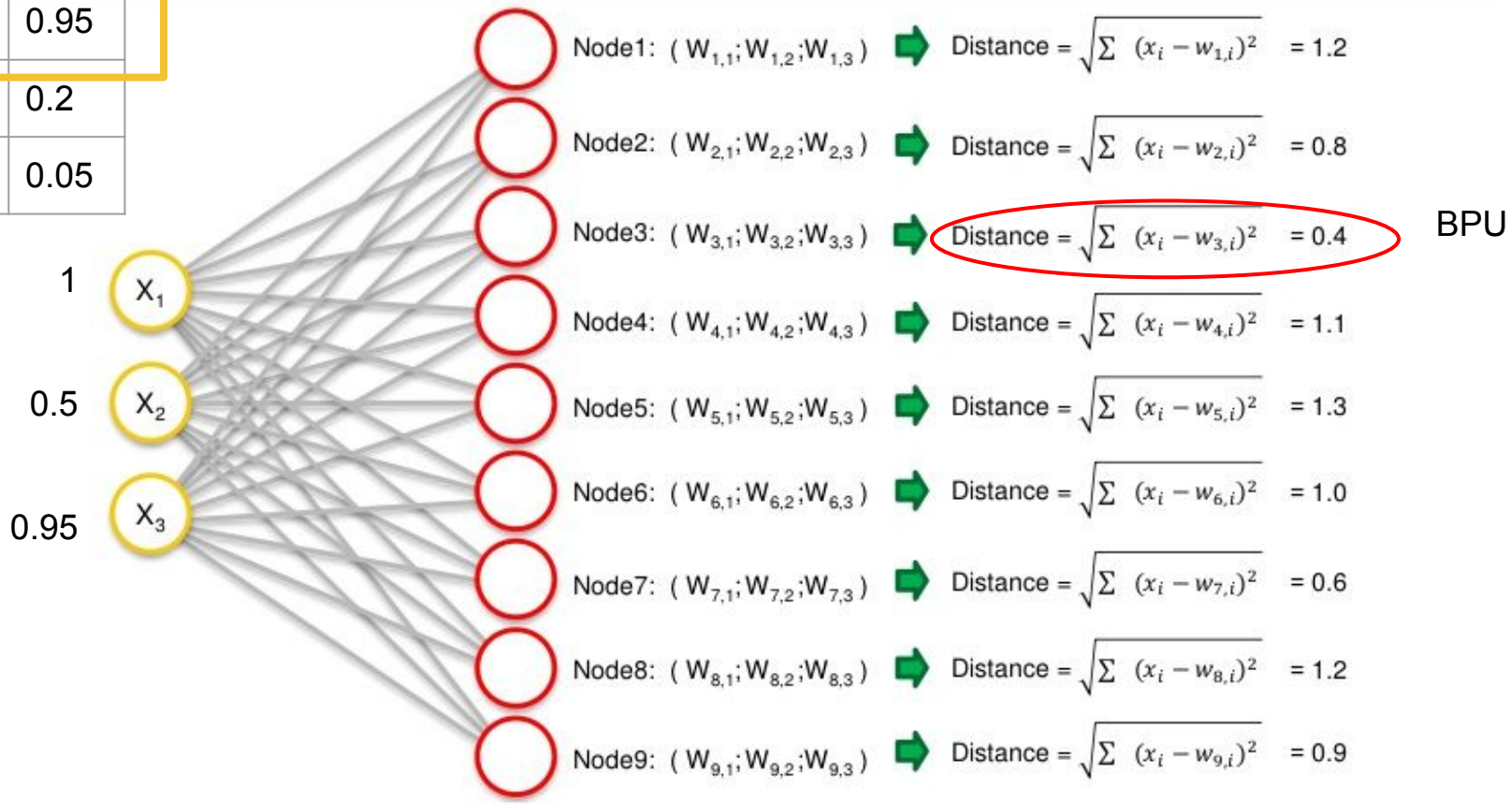
filaN	var1	var2	var3
1	1	0.5	0.95
2	0.5	0	0.2
3	1	0.75	0.05



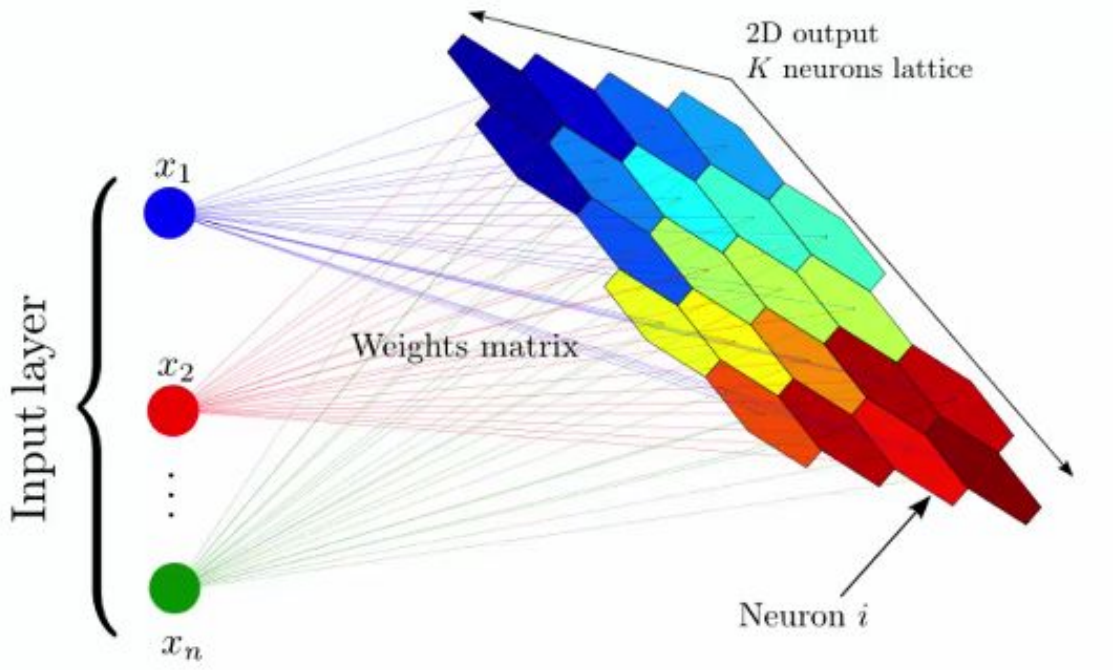
¿Cómo aprenden los SOM?

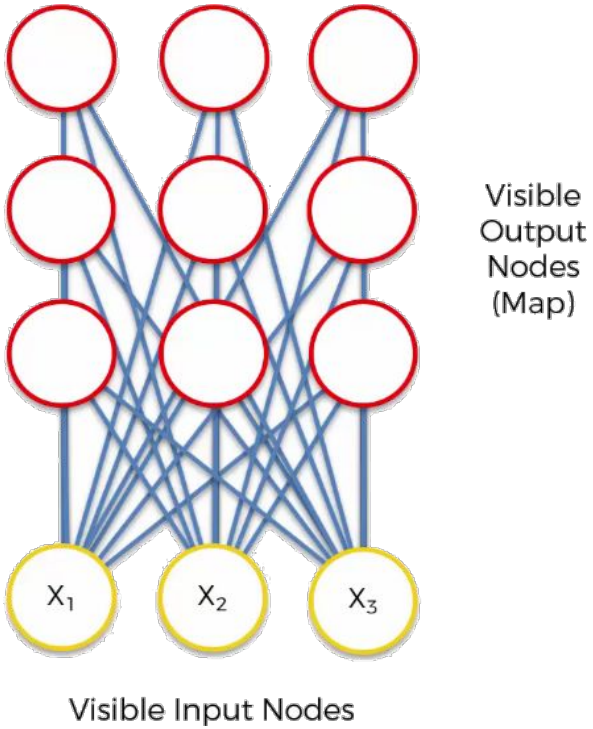
Mondragon Unibertsitatea
Faculty of Engineering

filaN	var1	var2	var3
1	1	0.5	0.95
2	0.5	0	0.2
3	1	0.75	0.05



¿Cómo aprenden los SOM?

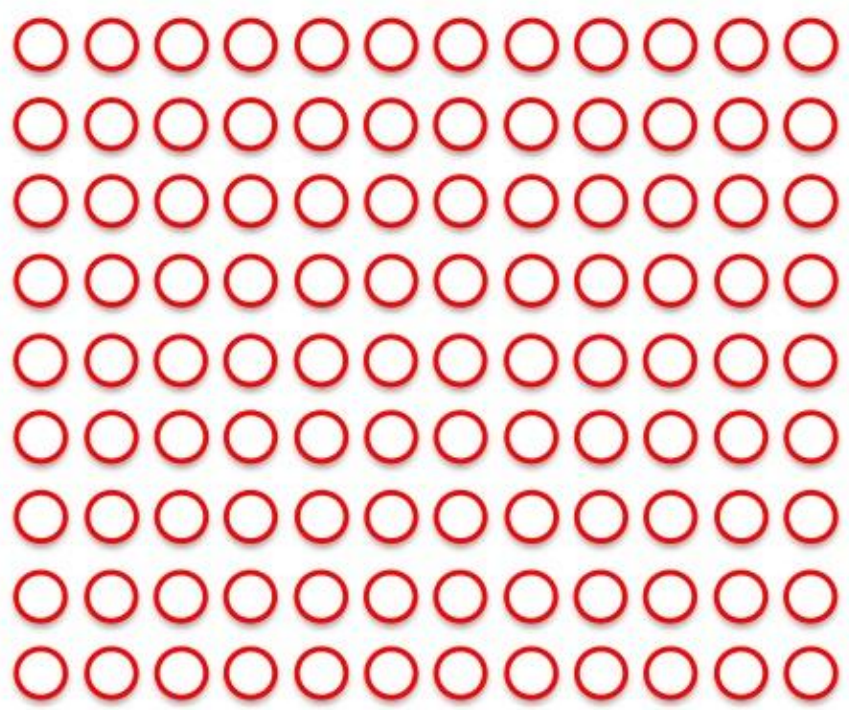




¿Cómo aprenden los SOM?

Mondragon Unibertsitatea
Faculty of Engineering

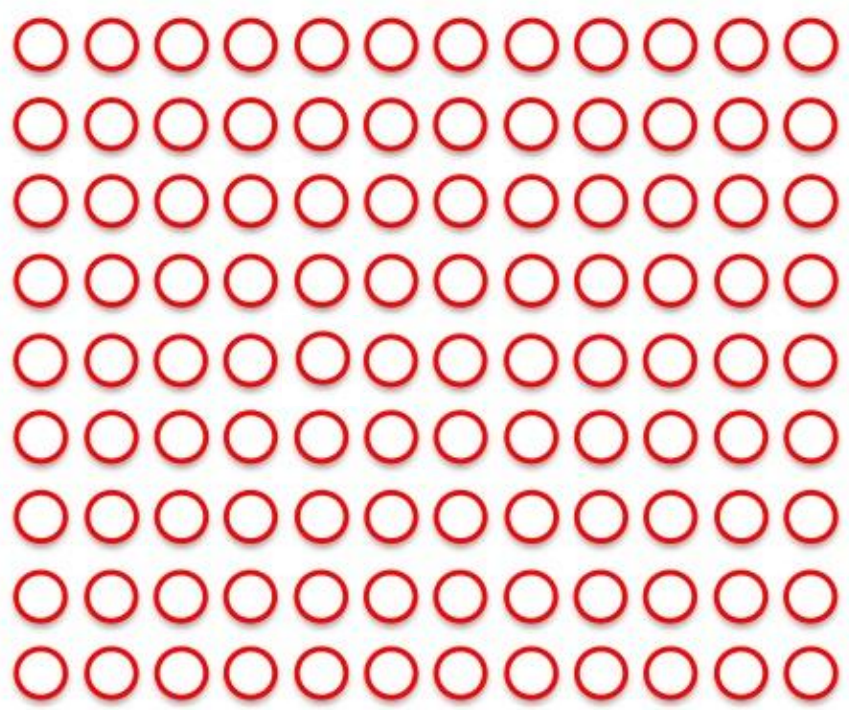
filaN	var1	var2	var3
1	1	0.5	0.95
2	0.5	0	0.2
3	1	0.75	0.05



¿Cómo aprenden los SOM?

Mondragon Unibertsitatea
Faculty of Engineering

filaN	var1	var2	var3
1	1	0.5	0.95
2	0.5	0	0.2
3	1	0.75	0.05



¿Cómo aprenden los SOM?

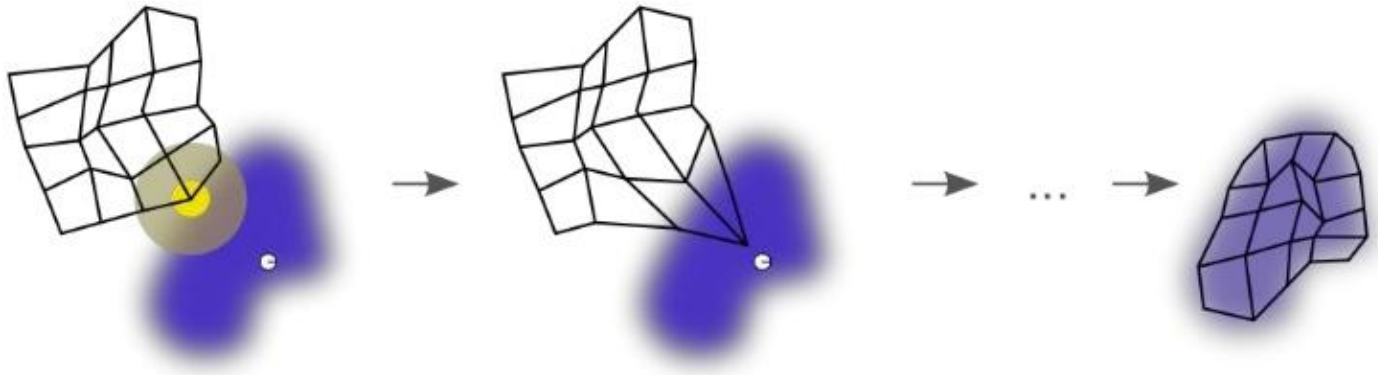
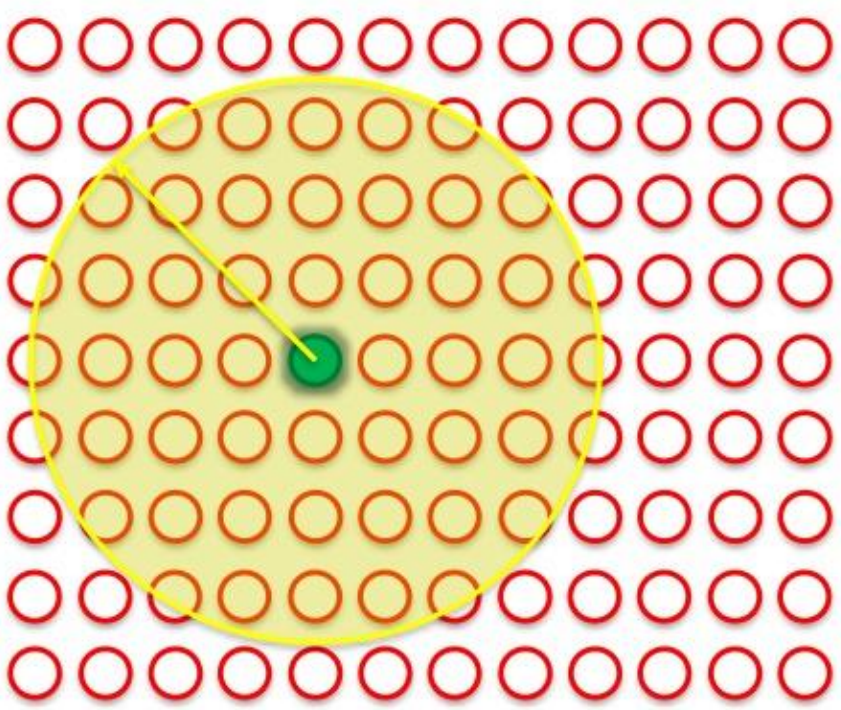


Image Source: Wikipedia

¿Cómo aprenden los SOM?

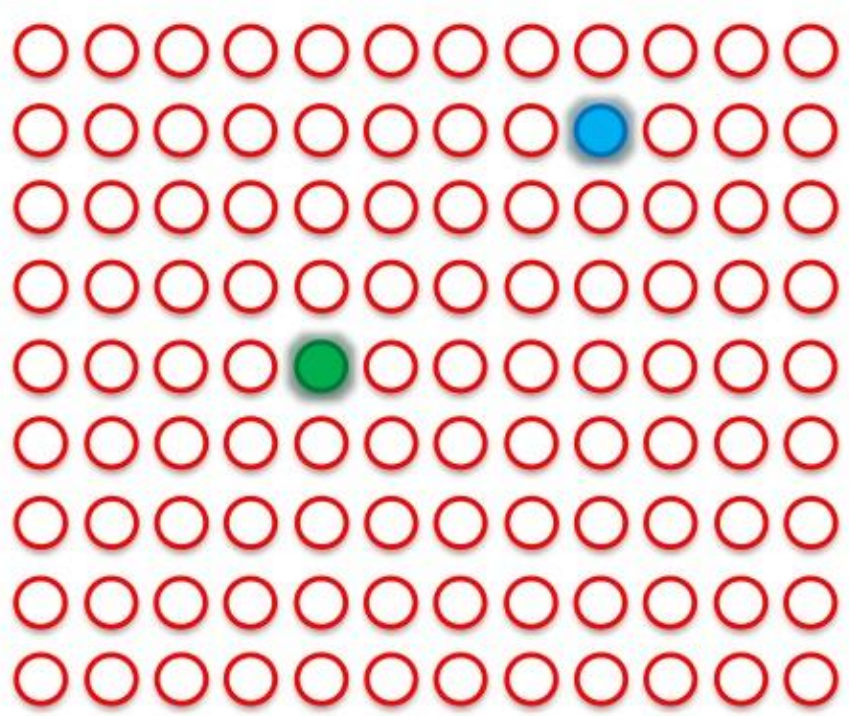
filaN	var1	var2	var3
1	1	0.5	0.95
2	0.5	0	0.2
3	1	0.75	0.05



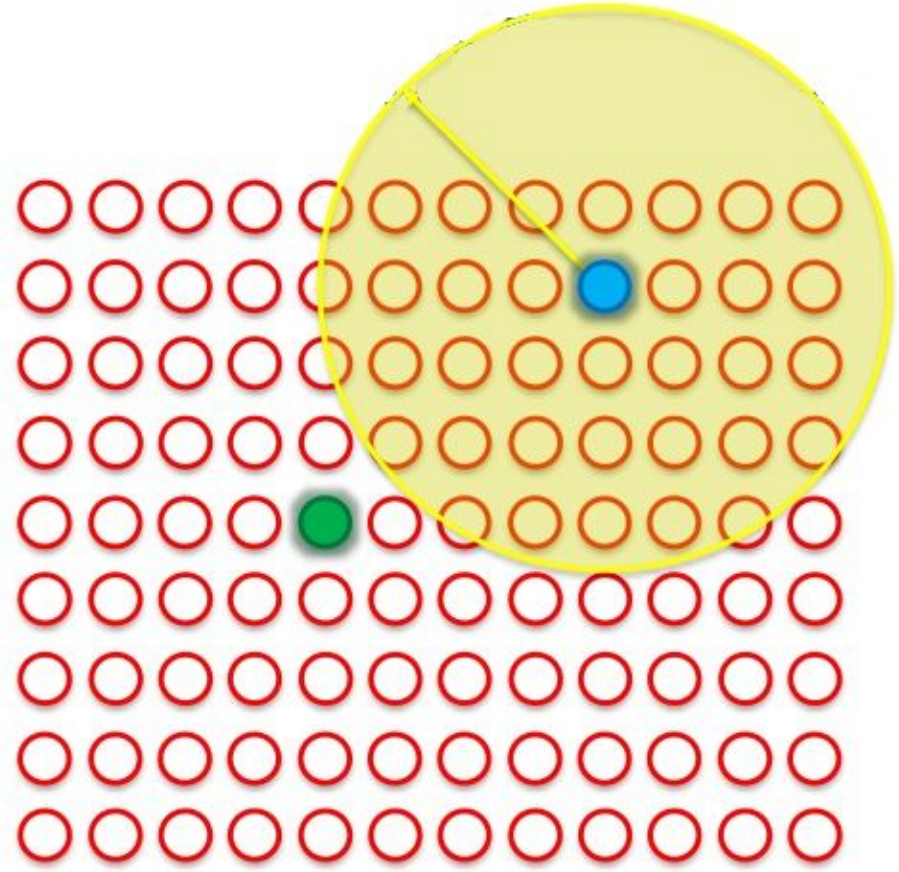
¿Cómo aprenden los SOM?

Mondragon Unibertsitatea
Faculty of Engineering

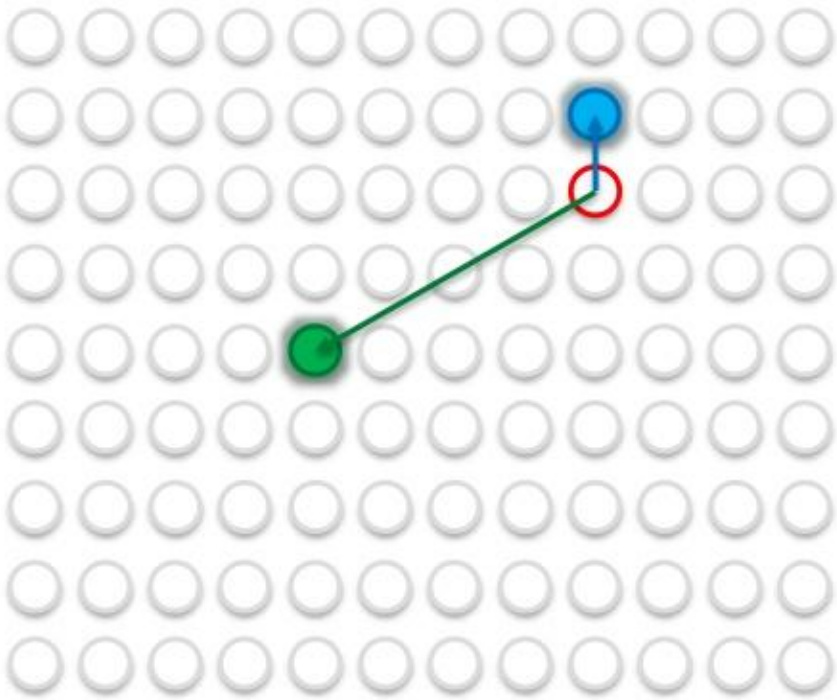
filaN	var1	var2	var3
1	1	0.5	0.95
2	0.5	0	0.2
3	1	0.75	0.05



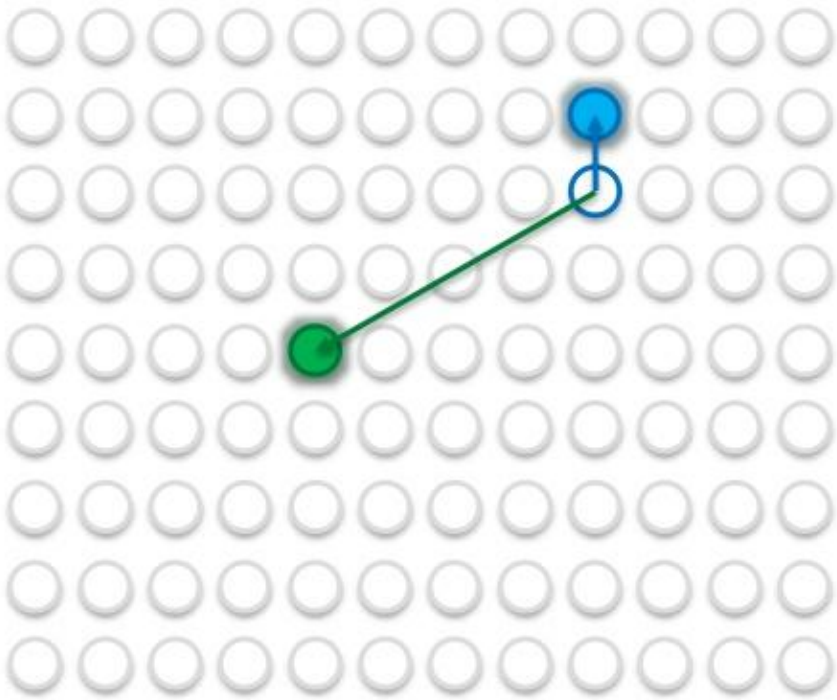
¿Cómo aprenden los SOM?



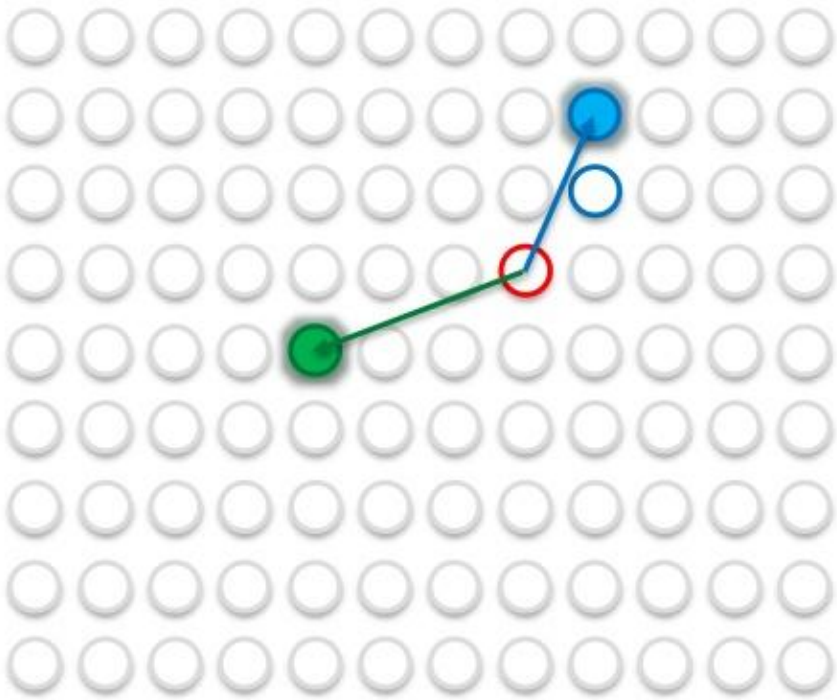
¿Cómo aprenden los SOM?



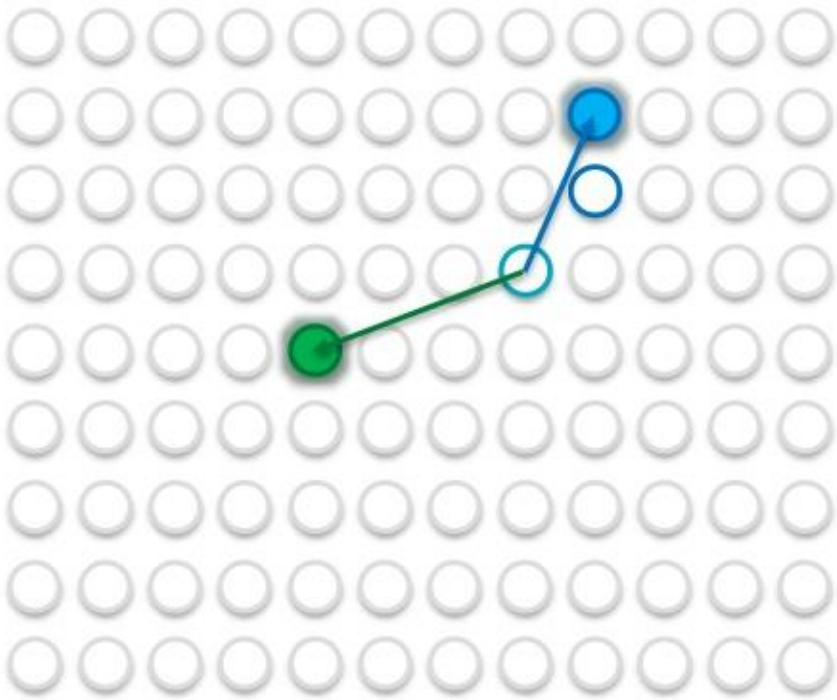
¿Cómo aprenden los SOM?



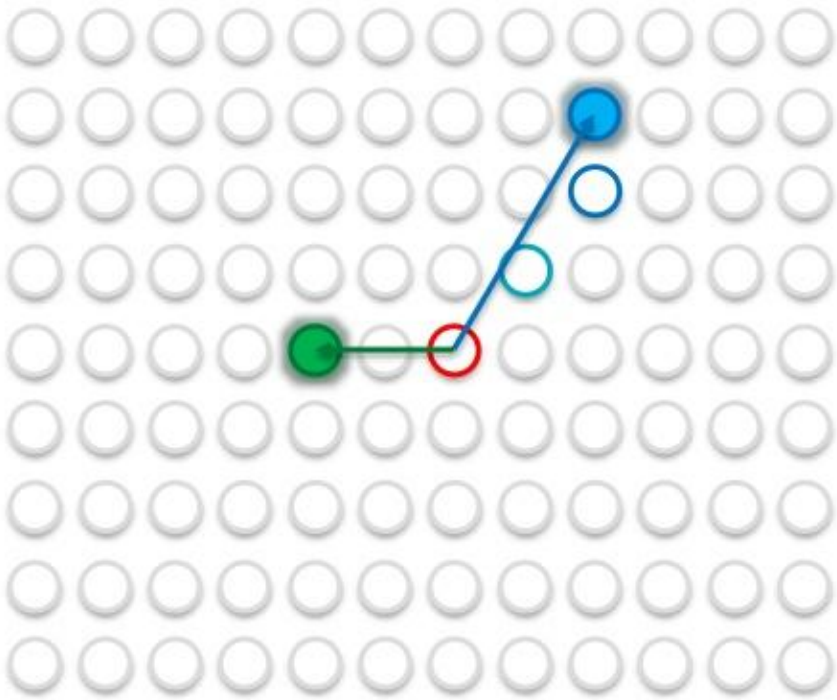
¿Cómo aprenden los SOM?



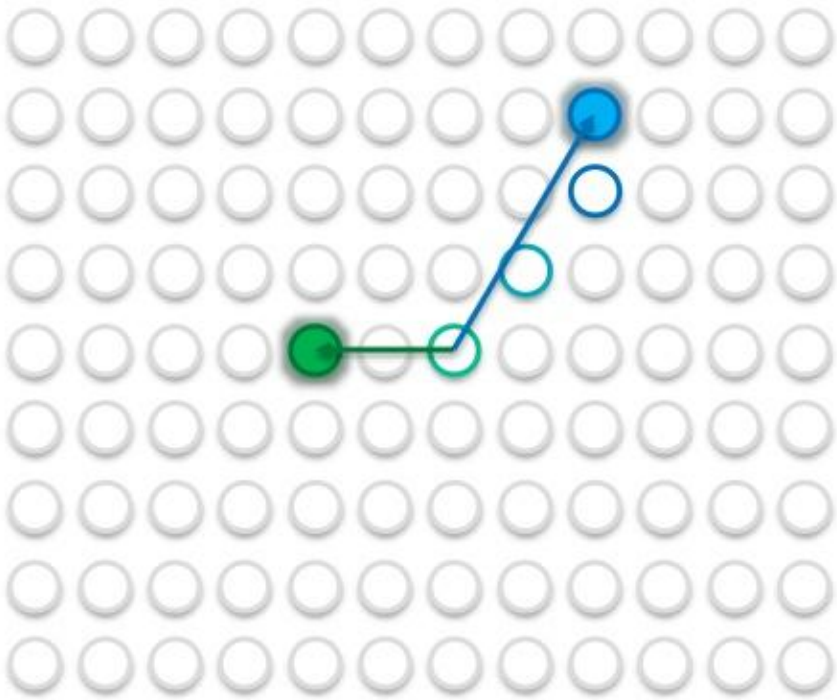
¿Cómo aprenden los SOM?



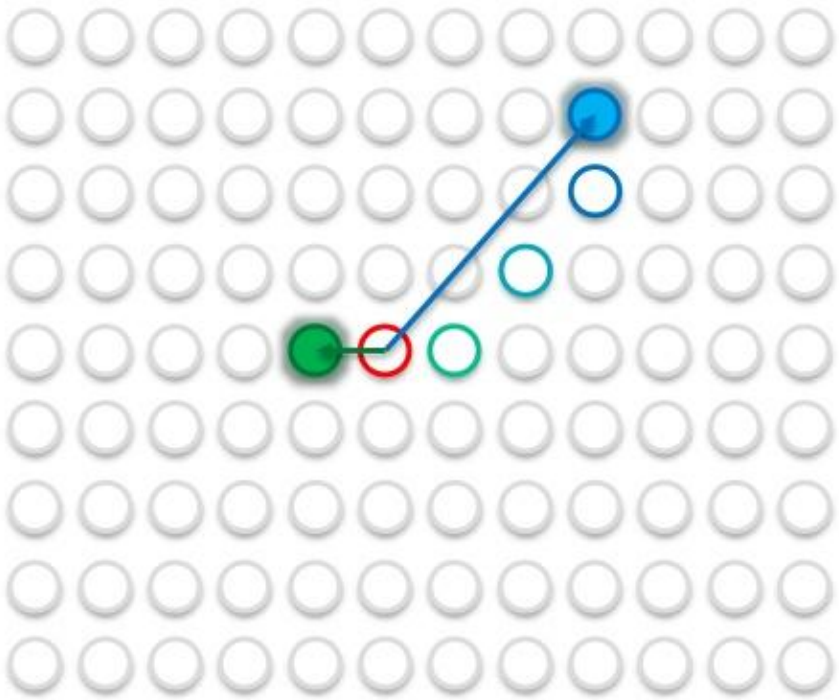
¿Cómo aprenden los SOM?



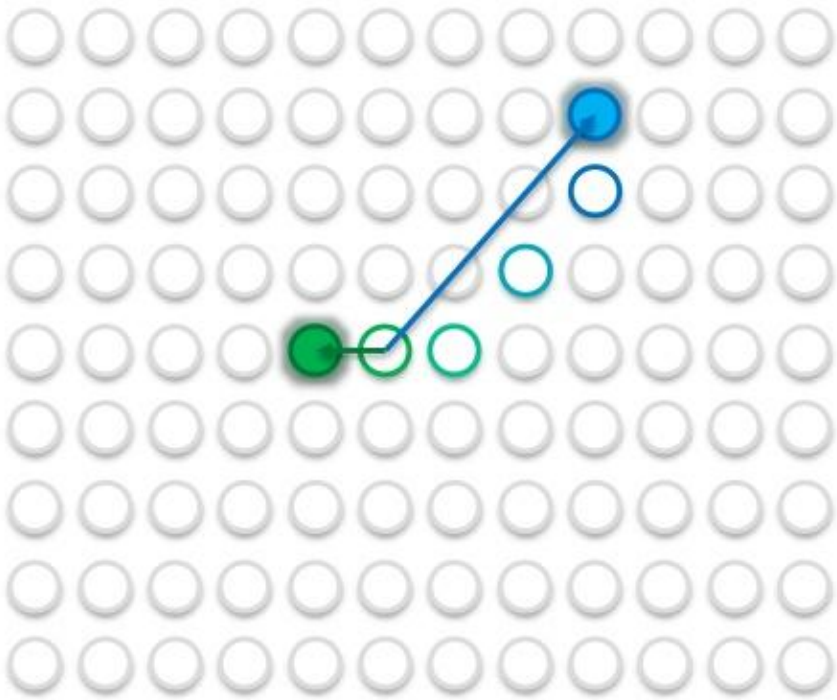
¿Cómo aprenden los SOM?



¿Cómo aprenden los SOM?



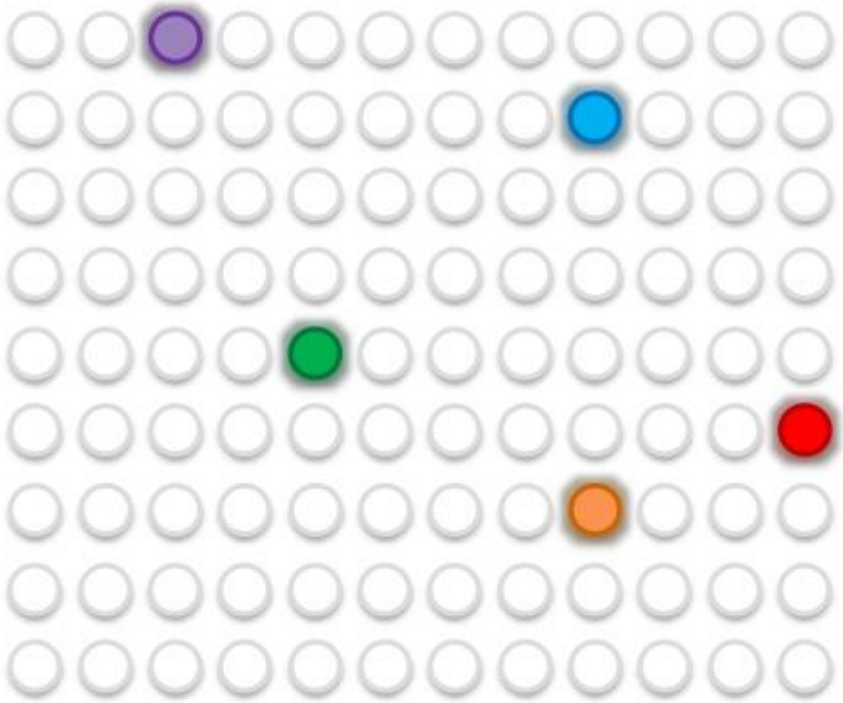
¿Cómo aprenden los SOM?



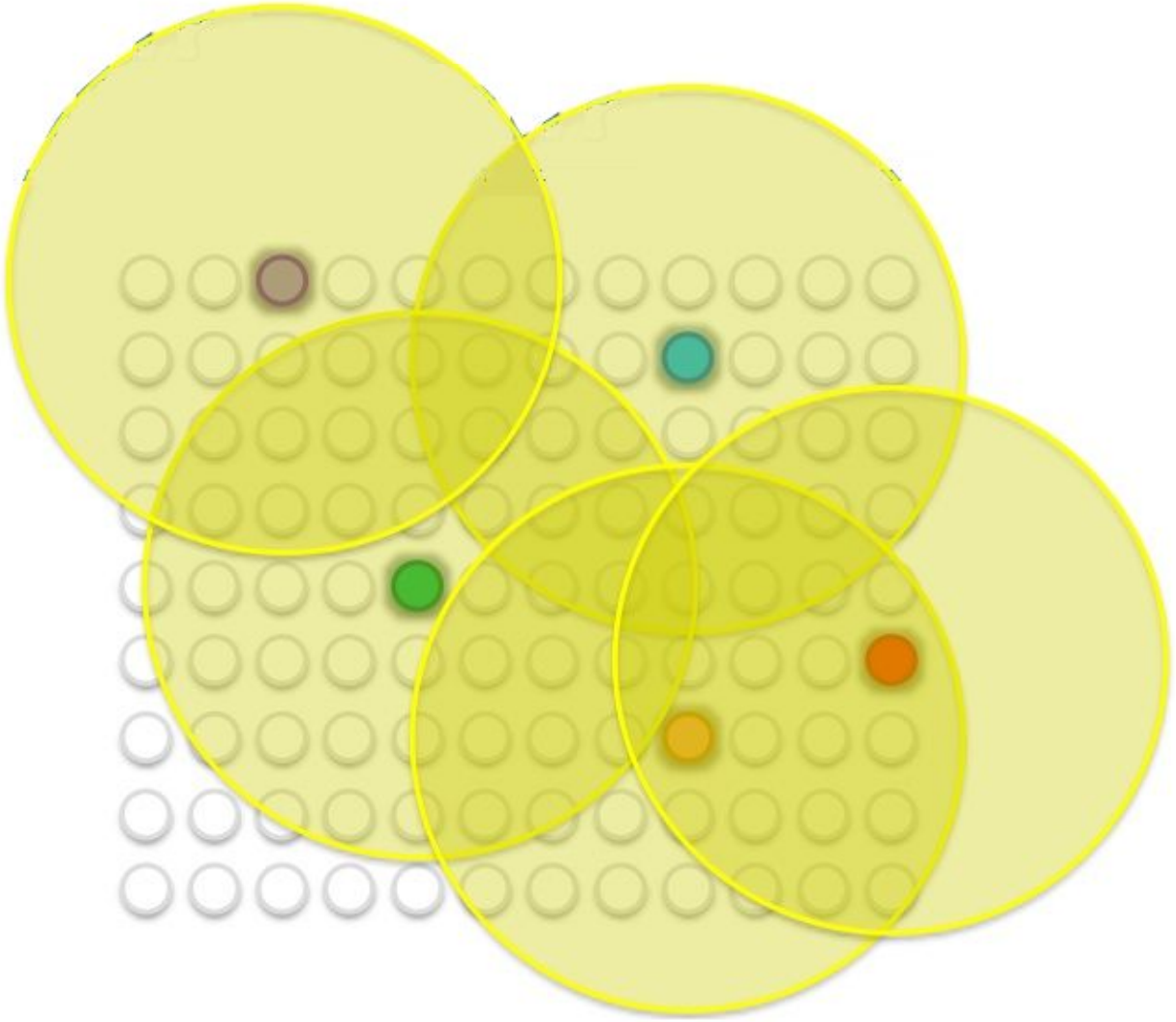
¿Cómo aprenden los SOM?

Mondragon Unibertsitatea
Faculty of Engineering

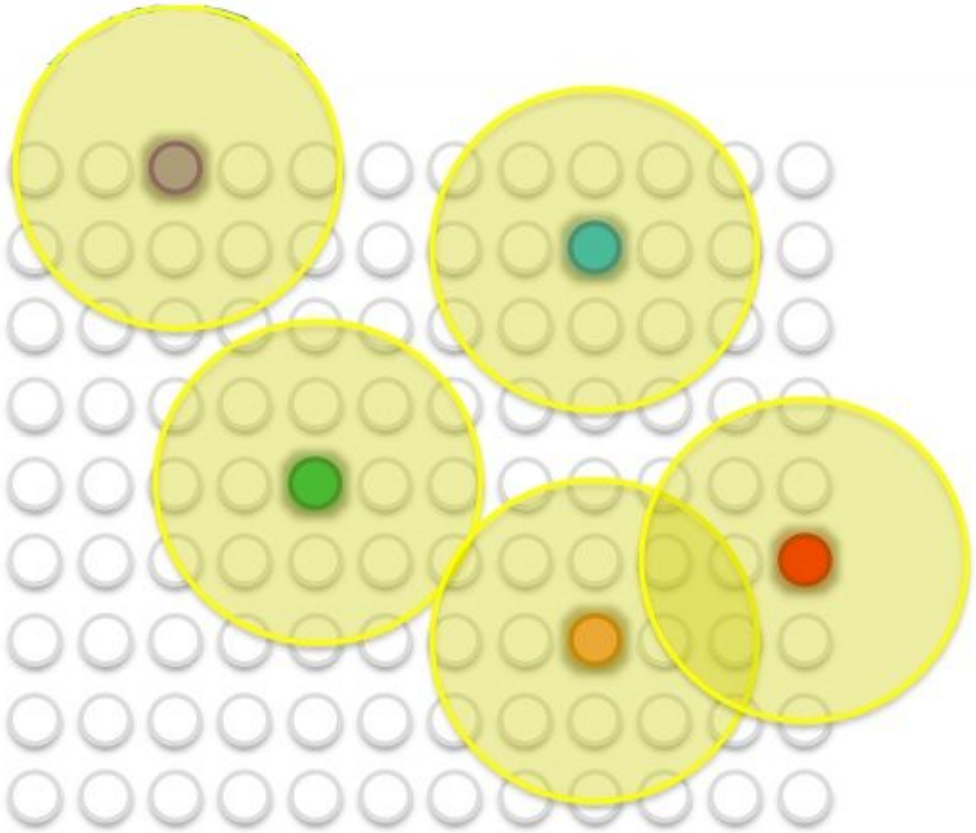
filaN	var1	var2	var3
1	1	0.5	0.95
2	0.5	0	0.2
3	1	0.75	0.05
4	0.75	1	0.1
5	0.1	0.6	0.4



¿Cómo aprenden los SOM?



¿Cómo aprenden los SOM?



¿Cómo aprenden los SOM?

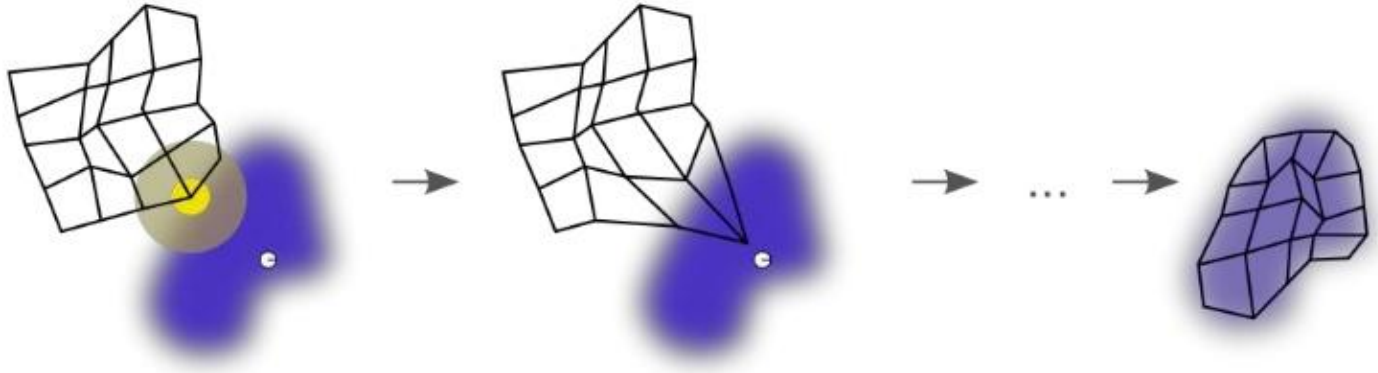
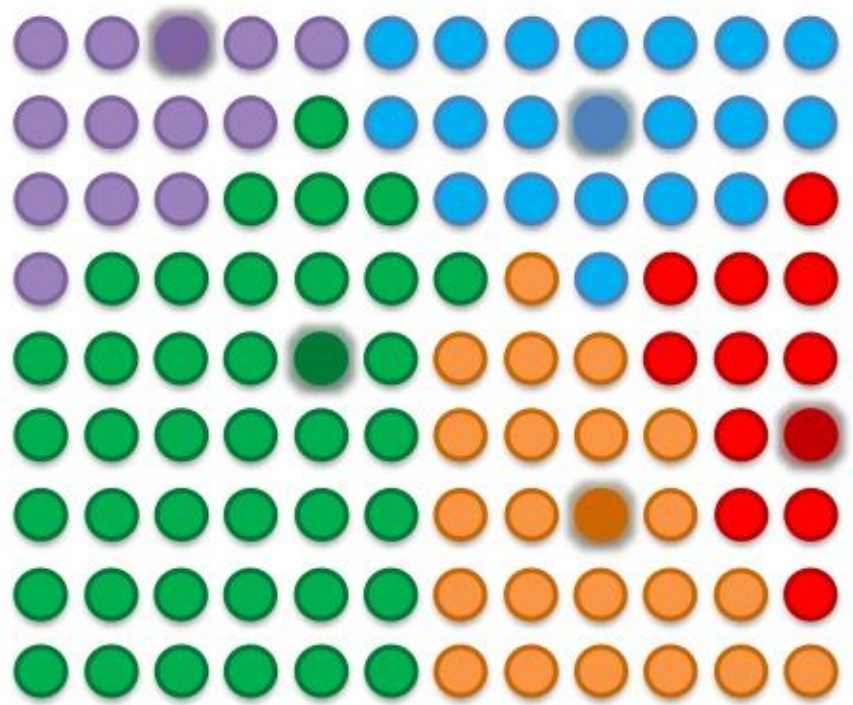


Image Source: Wikipedia

¿Cómo aprenden los SOM?



Cosas importantes para recordar

- Los SOMs mantienen la topología del dataset de entrada
- Los SOMs revelan correlaciones que no se ven fácilmente
- Clasifican datos sin supervisión
- NO HAY “BACPROPAGATION” -> no hay targets!
- no hay conexiones entre nodos de salida

Ejemplo de SOM

Mondragon Unibertsitatea
Faculty of Engineering

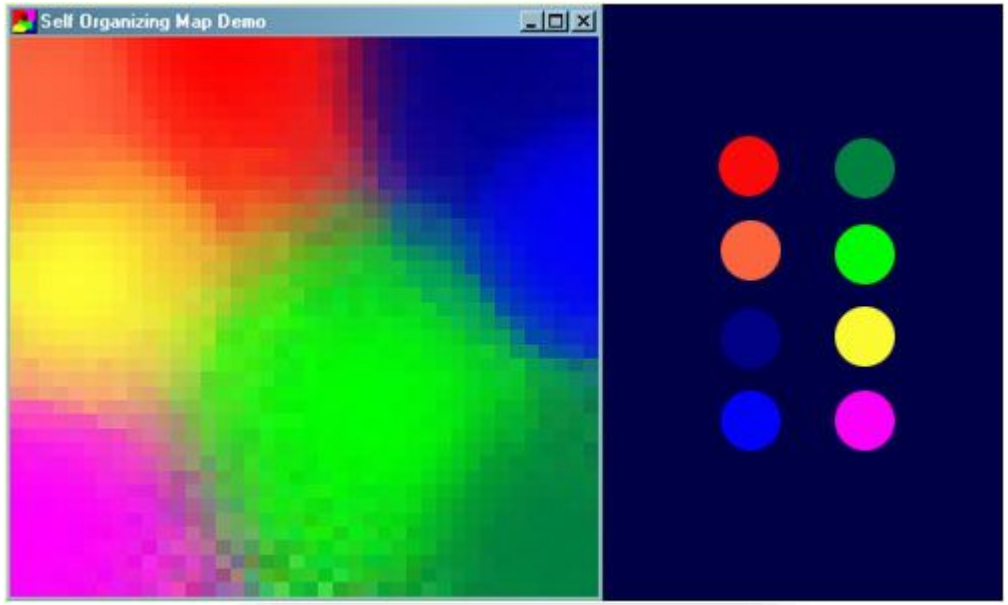
Additional Reading:

Kohonen's Self Organizing Feature Maps

By Mat Buckland

Link:

http://www.ai-junkie.com/ann/som/som1.html



Ejercicios!!!!!!!

Mondragon Unibertsitatea
Faculty of Engineering

1. Clusterización de Animales

2. Detección de anomalías en peticiones de tarjetas de crédito
   2.1. Fase1: unsupervised anomaly detection
   2.2. Fase2: supervised anomaly training

Mondragon Unibertsitatea

Faculty of Engineering

Eskerrik asko

Muchas gracias

Thank you

Ekhi Zugasti
ezugasti@mondragon.edu

Loramendi, 4. Apartado 23
20500 Arrasate – Mondragon
T. 943 71 21 85
info@mondragon.edu