

MBA
USP
ESALQ

Otros Modelos de Machine Learning III

João F. Serrajordia R. de Mello

Va a necesitar de...

Preparativos

- Abrir R
- Importar las bibliotecas
- Algo para hacer sus anotaciones



Agenda

Revisión

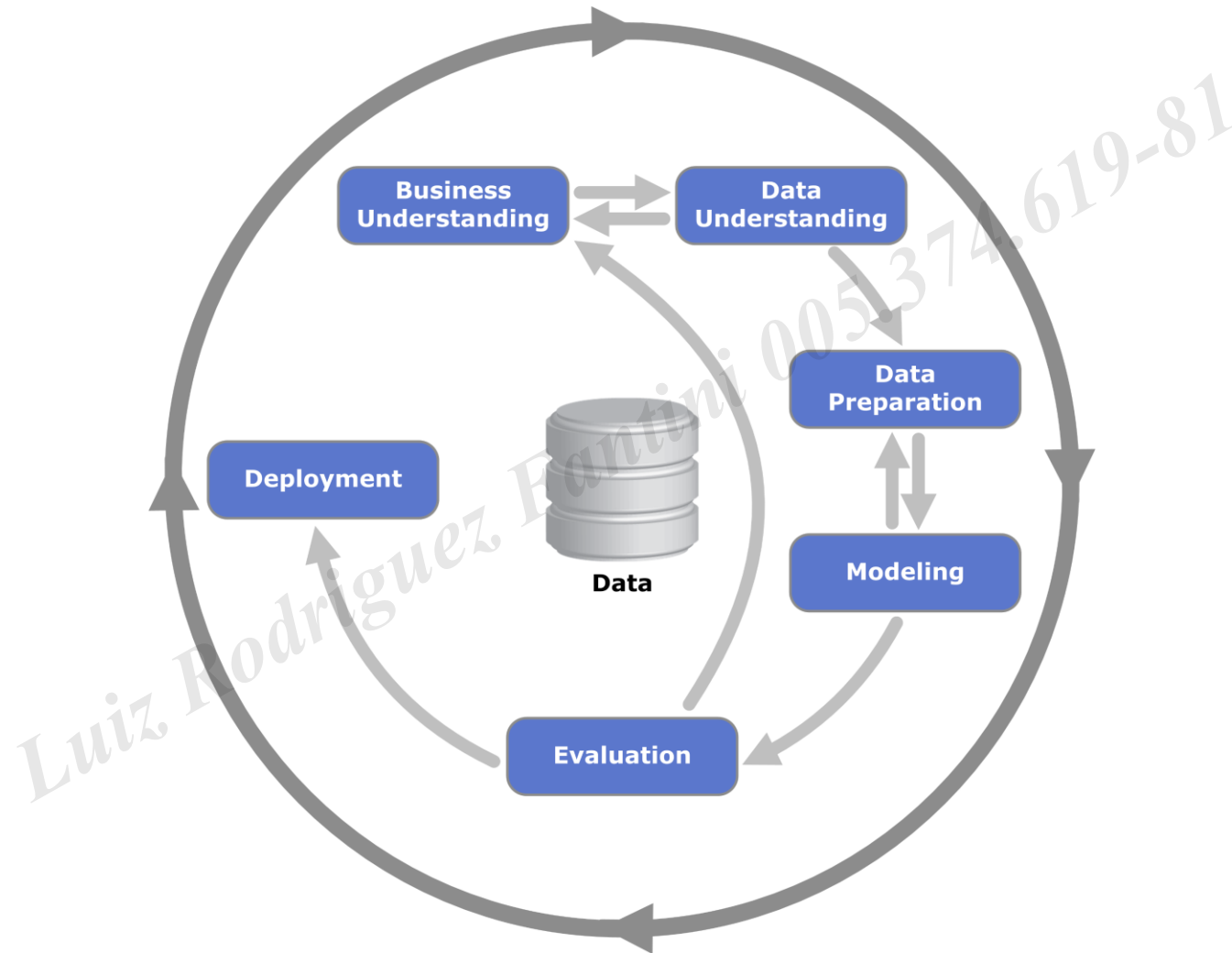
Histórico

Ideas básicas

Usos

Luiz Rodriguez Fantini 005.374.619-81

CRISP-DM



Fuente: <https://www.the-modeling-agency.com/crisp-dm.pdf>



Ensemble

Un ensemble es cualquier mezcla de modelos ya existentes.
Los principales tipos son:

Bagging

Boosting

Stacking

Bootstrap – aggregation (bagging)



El *bagging* con árboles es el famoso *Random Forest*

ID	...	Y
1	...	1
2	...	0
...
N	...	0



Y	P	ERRO
1	75%	25%
0	20%	20%
...
0	40%	40%



ERRO	Δ	P	ERRO
25%	10%	85%	15%
-20%	-10%	10%	-10%
...
-40%	-15%	25%	-25%



ERRO	Δ	P	ERRO
15%	2%	87%	5%
-10%	-1%	9%	5%
...
-25%	-5%	20%	10%

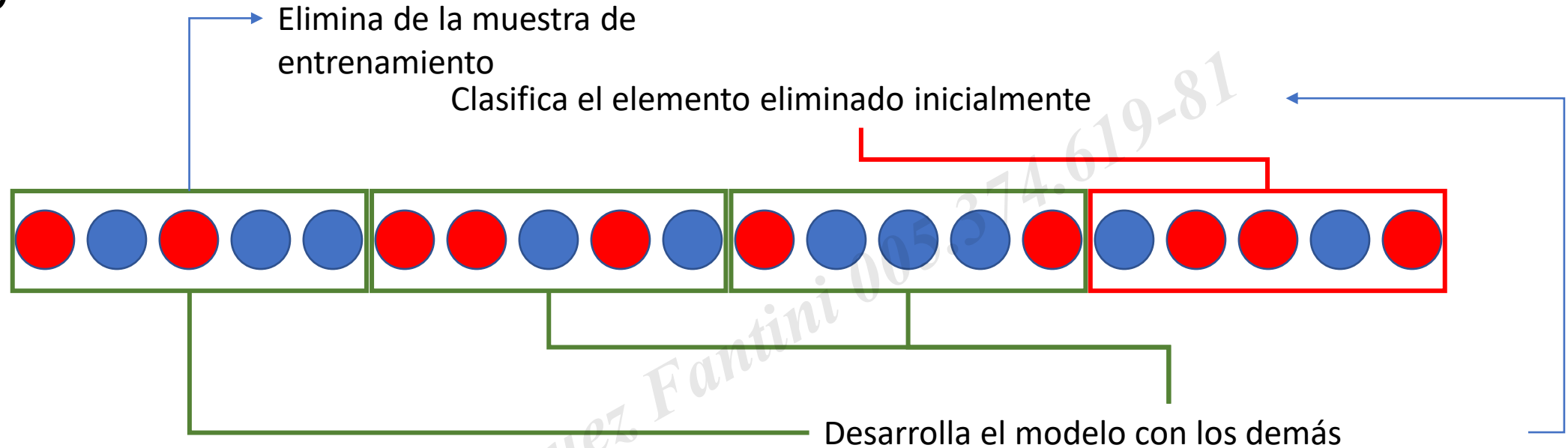
La variable respuesta de una iteración es el 'error' de la anterior.

La variable respuesta de una iteración es el 'error' de la anterior.

Boosting

- Los métodos de *boosting* son modelos secuenciales que intentan mejorar el error del modelo anterior

K-fold



- Dividimos la base en k submuestras
- Para cada submuestra:
 - Eliminamos la submuestra como validación
 - Entrenamos el modelo con las observaciones restantes
 - Utilizamos este modelo para clasificar la submuestra eliminada
 - Evaluamos la métrica de desempeño del modelo
- Calculamos la media de las métricas de desempeño del modelo

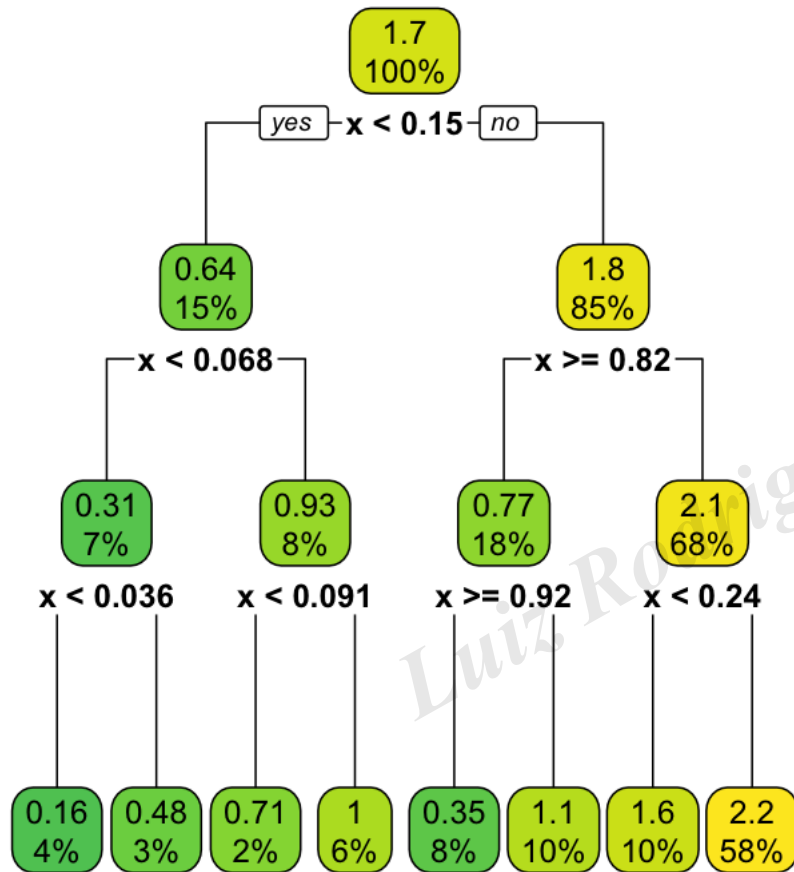
Árboles de regresión

Son muy semejantes a los árboles de clasificación

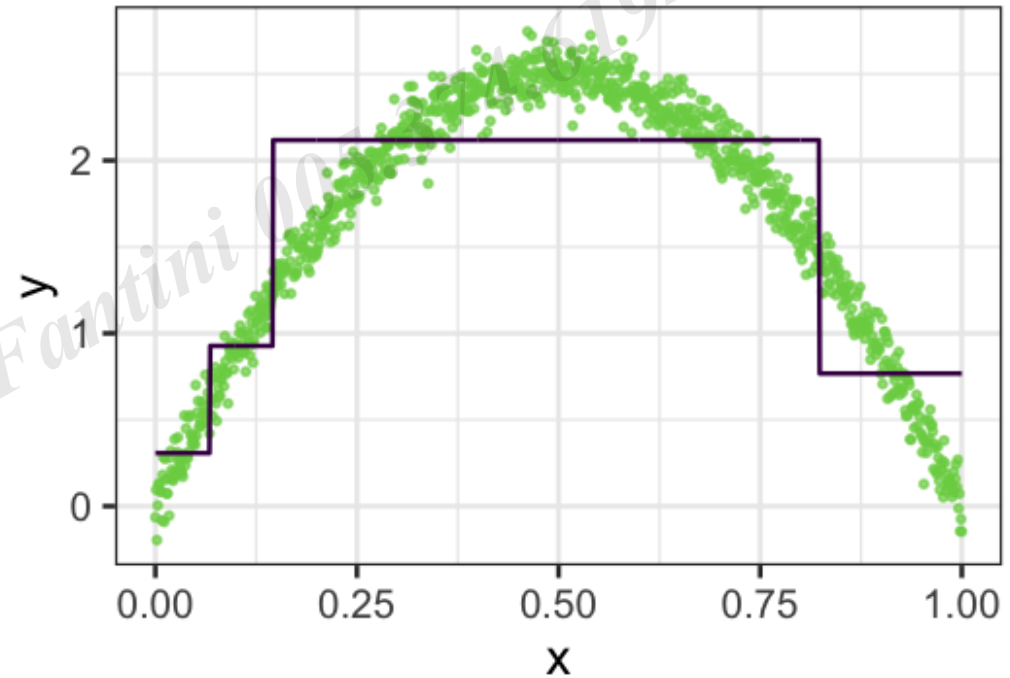
Lo que cambia es el criterio de impureza

$$SQE = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Árboles de regresión



Valores observados vs esperados



Dato: — Esperado — Observado

Problemas de predictivos y de clasificación



¿Cuál es la eficacia de una vacuna?



¿El cliente pagará el préstamo?



¿Cuánto petróleo tiene el pozo?



¿El cliente va a comprar mi producto?



¿Qué está haciendo la persona?



¿Cuán ecológico es ese vehículo?

clasificación

Luiz Rodriguez Fantini 005.374.619-81

Clasificación de los algoritmos

Supervisados

- Regresión
- GLM
- GLMM
- Support vector machines
- Naive Bayes
- K-nearest neighbors
- Redes Neurales
- Decision Trees



No Supervisados

- K-Means
- Métodos jerárquicos
- Mezcla Gaussiana
- DBScan
- Mini-Batch-K-Means



¡Estamos aquí!

Clasificación de los algoritmos



Respuesta continua

- Regresión
- GLM
- GLMM
- Support vector machines
- K-nearest neighbors
- Redes Neurales
- Regression Trees



Respuesta discreta

- Regresión logística
- Clasification trees
- Redes Neurales
- GLM
- GLMM

¡Estamos aquí!

Clasificación de los algoritmos

Métodos Machinelárnicos

- Árboles de decisión
- Bagging
- Boosting
- K-NN
- Redes Neurales
- Support vector machines

Métodos Machinelárnico- estadísticos

- Regresión
- GLM
- GLMM
- ANOVA

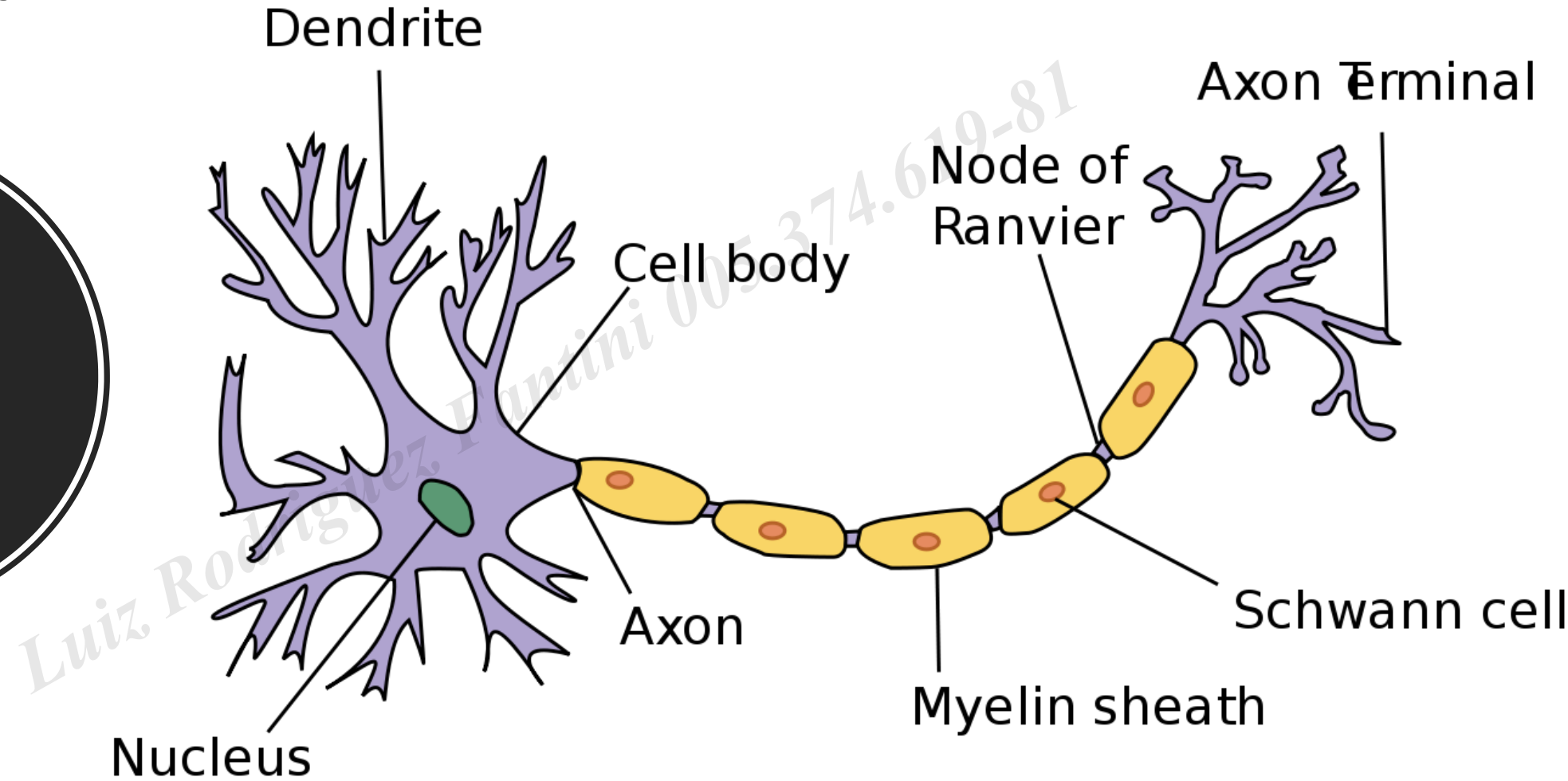
¡Estamos aquí!



Redes Neurales Artificiales

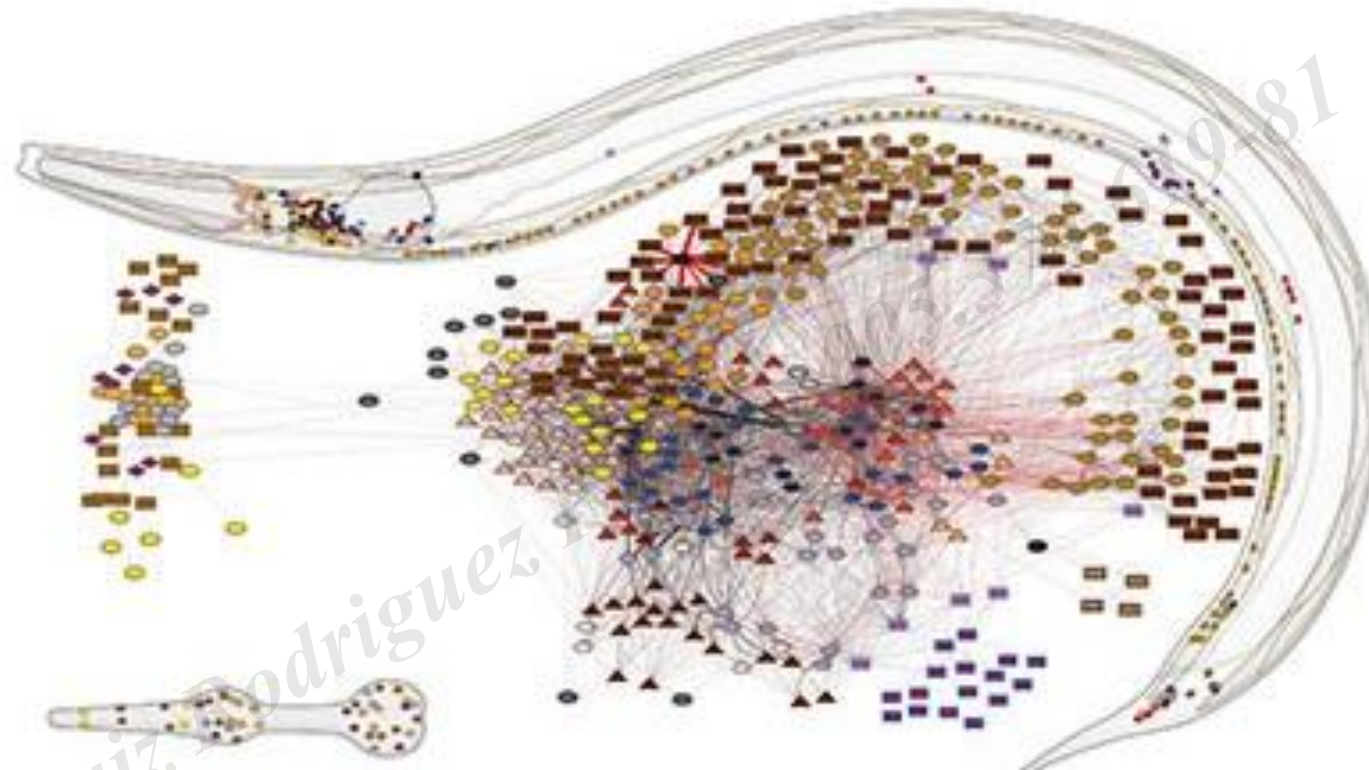
Juiz Rodriguez, Portini 005.374.079-81

Metáfora



<https://en.wikipedia.org/wiki/Myelin>

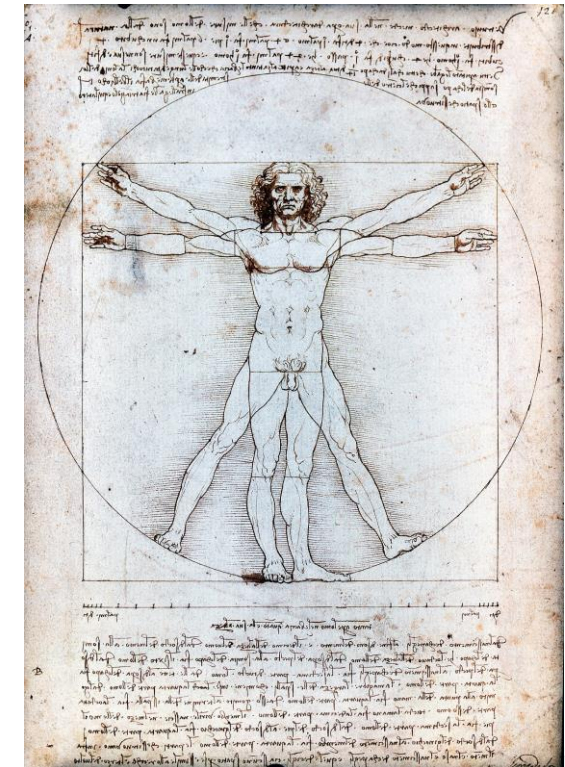
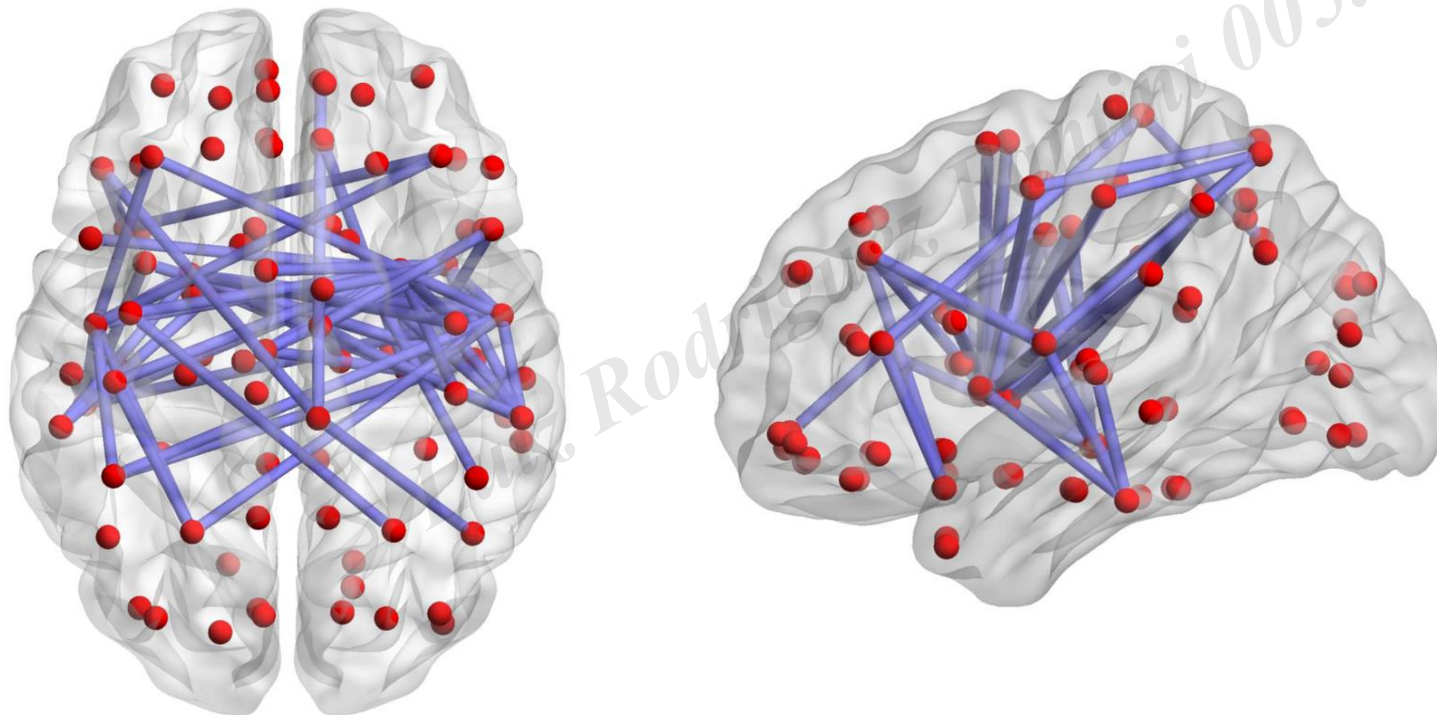
Ejemplo biológico



- *Nematelminto*: 302 neuronas

Red Neural Humana

- *Homo sapiens*: 100.000.000 de neuronas

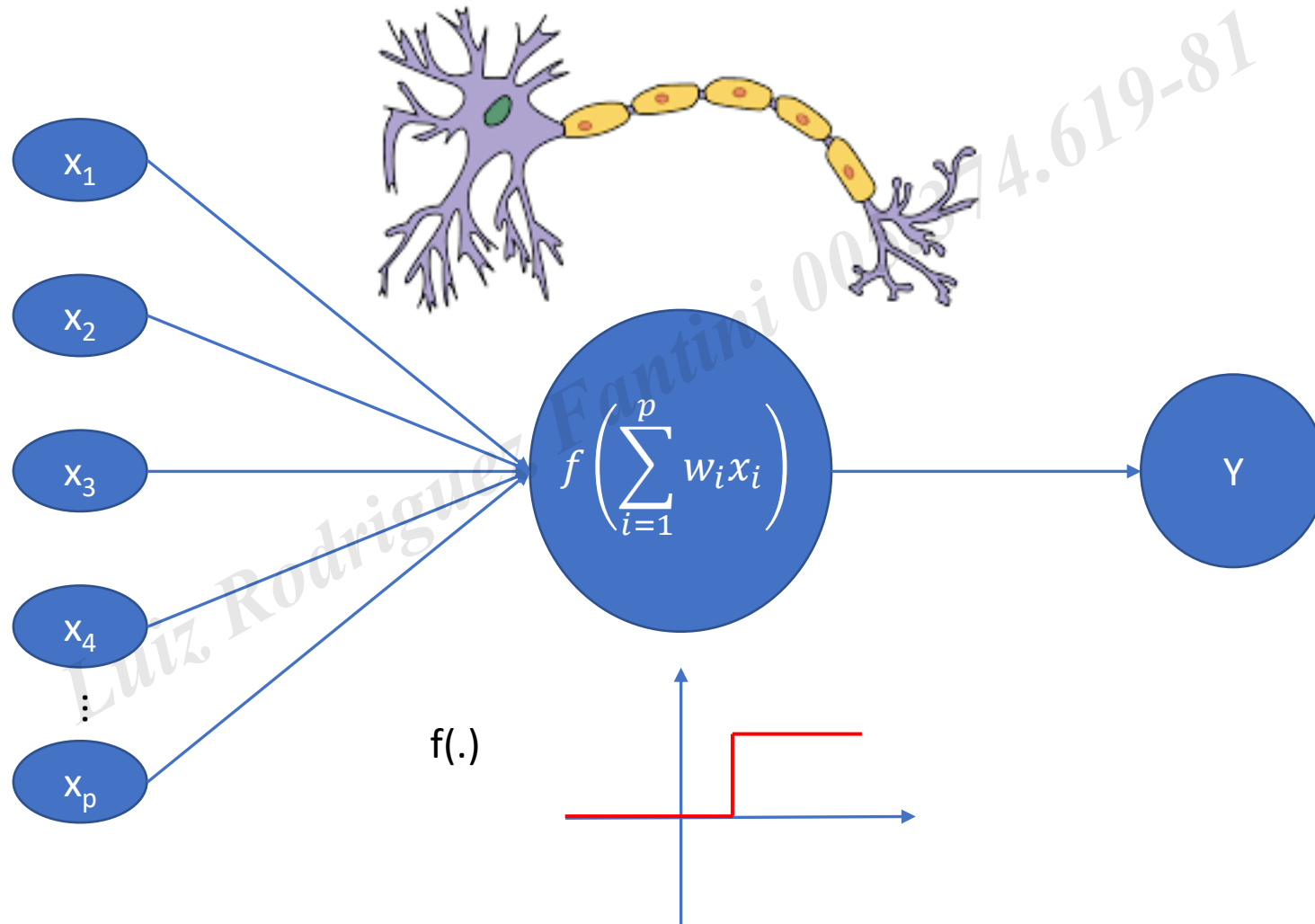


¿Dónde viven?

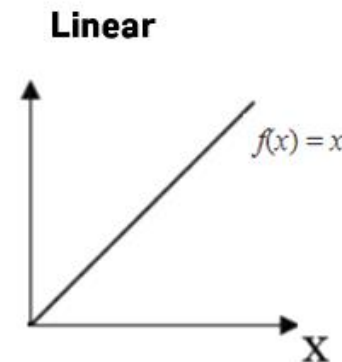
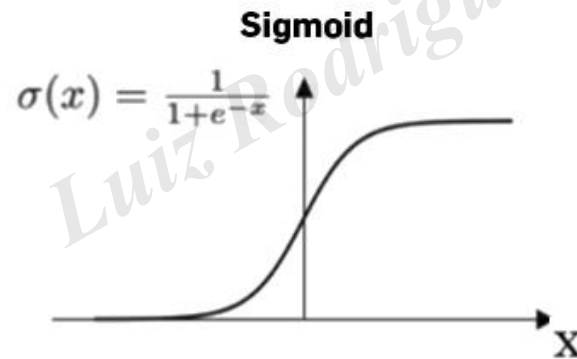
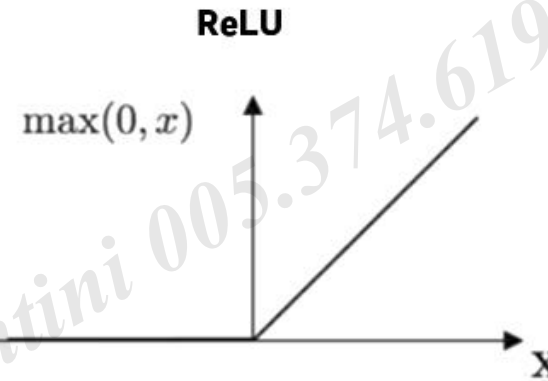
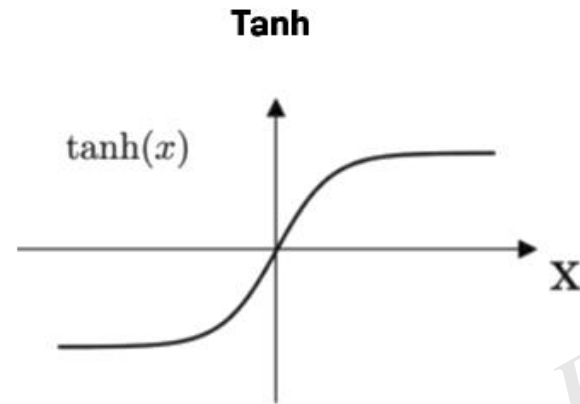


Redes Neurales Artificiales han tenido mucho éxito en problemas con datos poco estructurados como imágenes, audios, textos y videos.

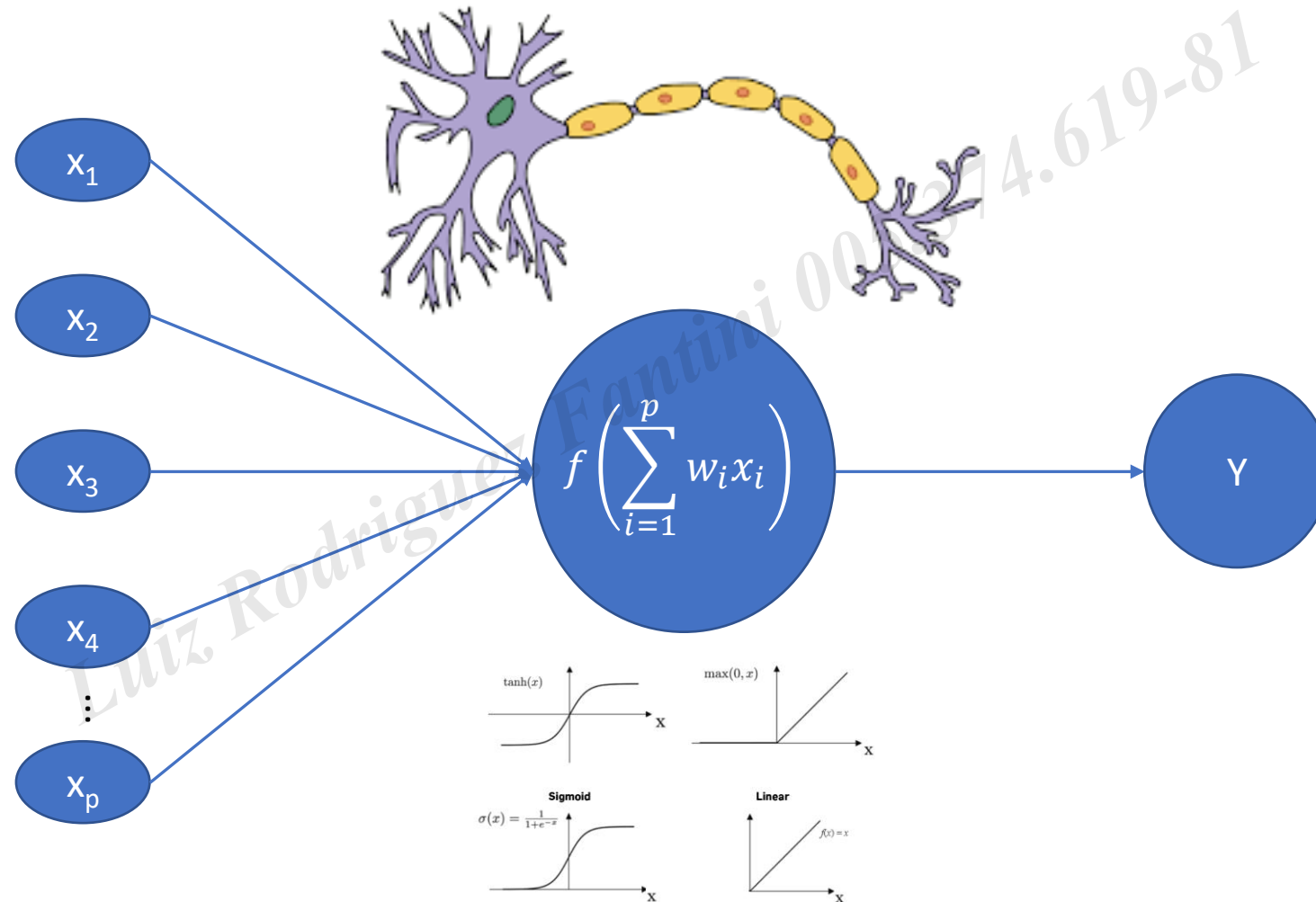
Neurona de McCulloch-Pitts



Funciones de activación



Perceptrón



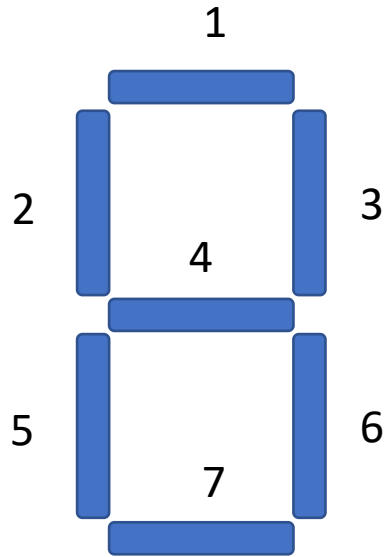
OCR – Optical Character Recognition



Vamos a pensar en una versión bien simple del problema. Dígitos de un radio reloj antiguo poseen una estructura bien simple.

Luiz Rodriguez Fuentetaja 005.374.619-81

OCR – Optical Character Recognition



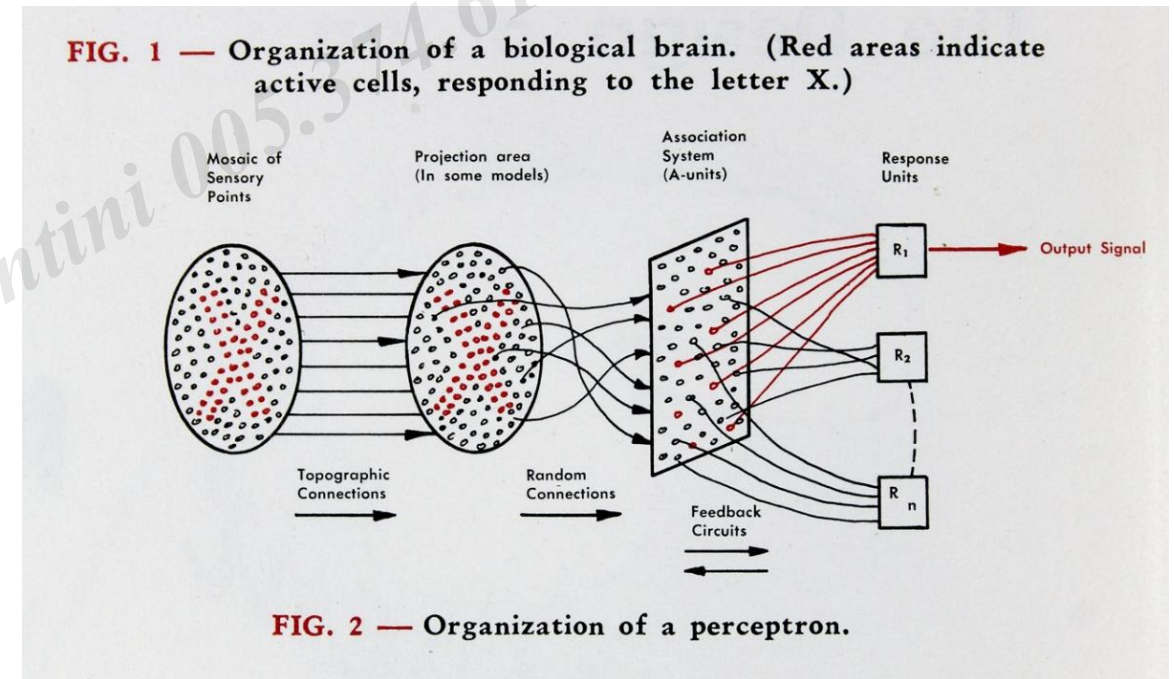
Hay 7 regiones básicas, que pueden estar activas o inactivas, y definen un dígito.

Por ejemplo, si solamente las regiones 1, 3 y 6 están activadas, tenemos el número 7.

Luiz Rodriguez Fortini 003.274.619-81

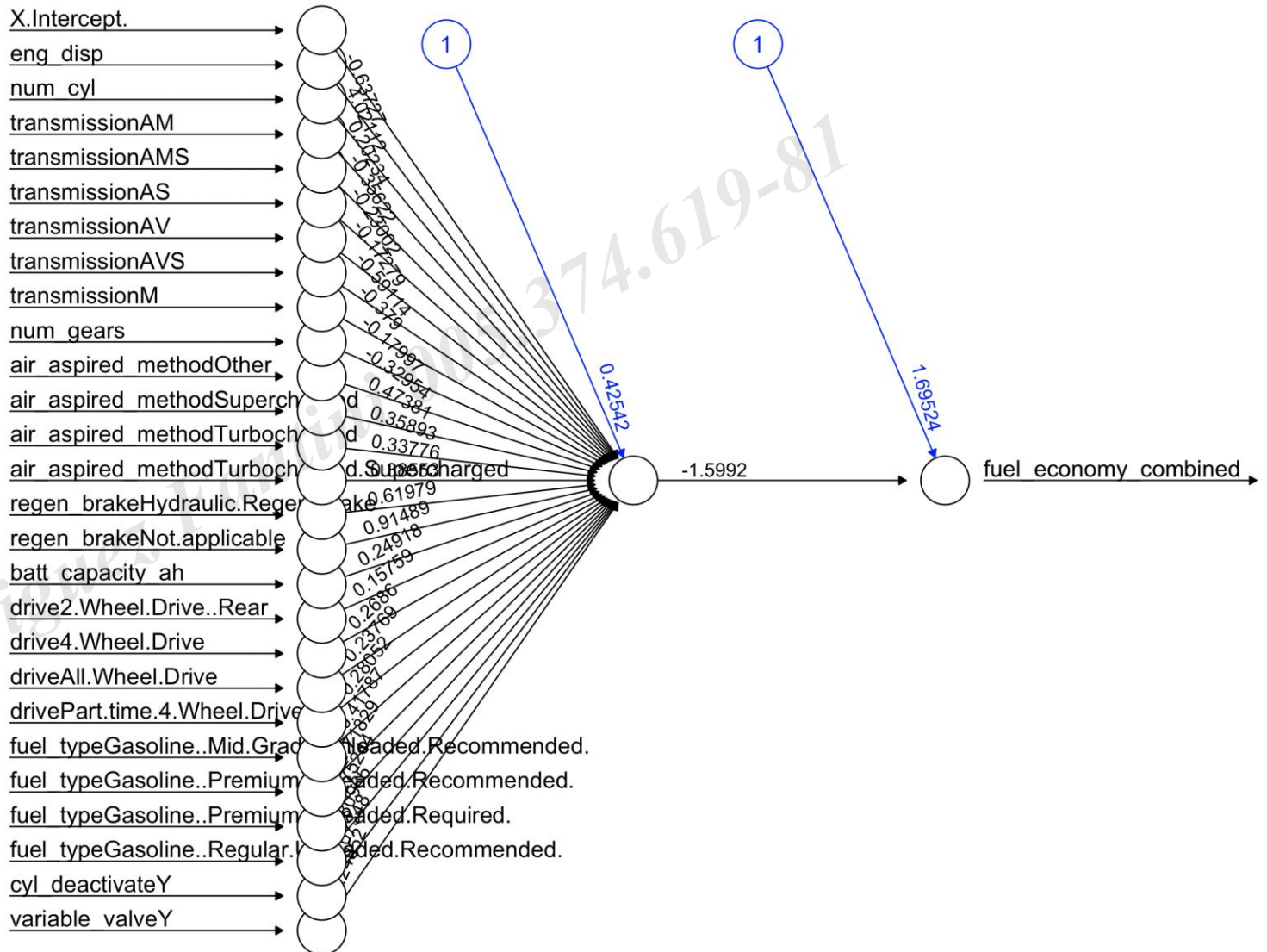
Perceptrón de Rosenblatt

- El Perceptrón de Rosenblatt (~1950-1960) tiene esa idea, sólo que con propósito más general
- fue construido para hacer OCR (optical character recognition)
- Para eso, mapea regiones de una imagen “activadas” y “no activadas”
- Cada unidad es una neurona de McCulloch-Pitt



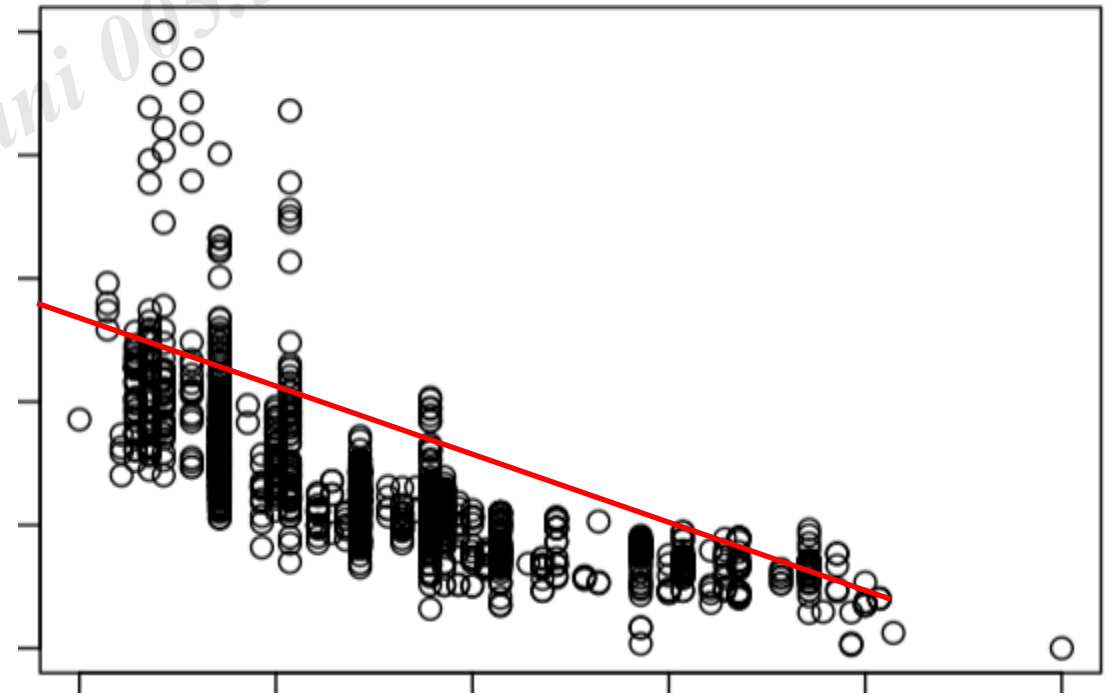
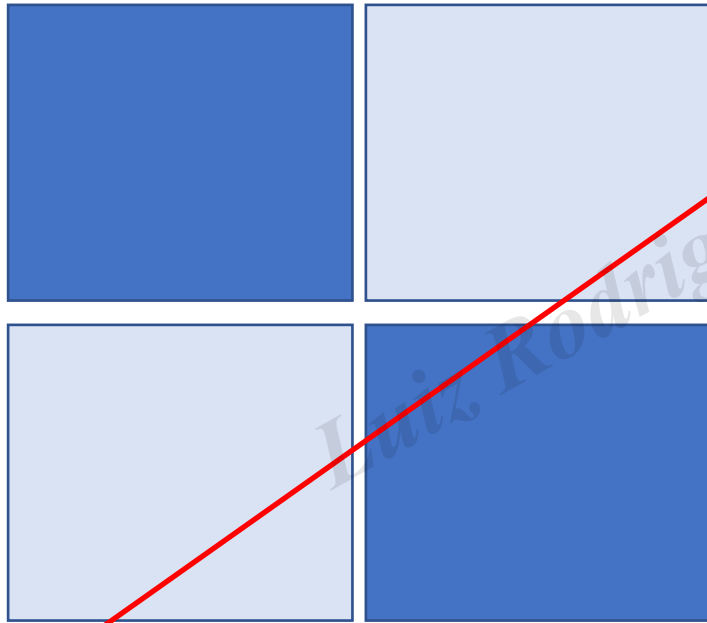
Perceptron Linear

- Posee la misma estructura de una regresión lineal con la función de activación indicada.



Limitaciones del perceptrón lineal

- El perceptrón lineal sólo captura patrones lineales

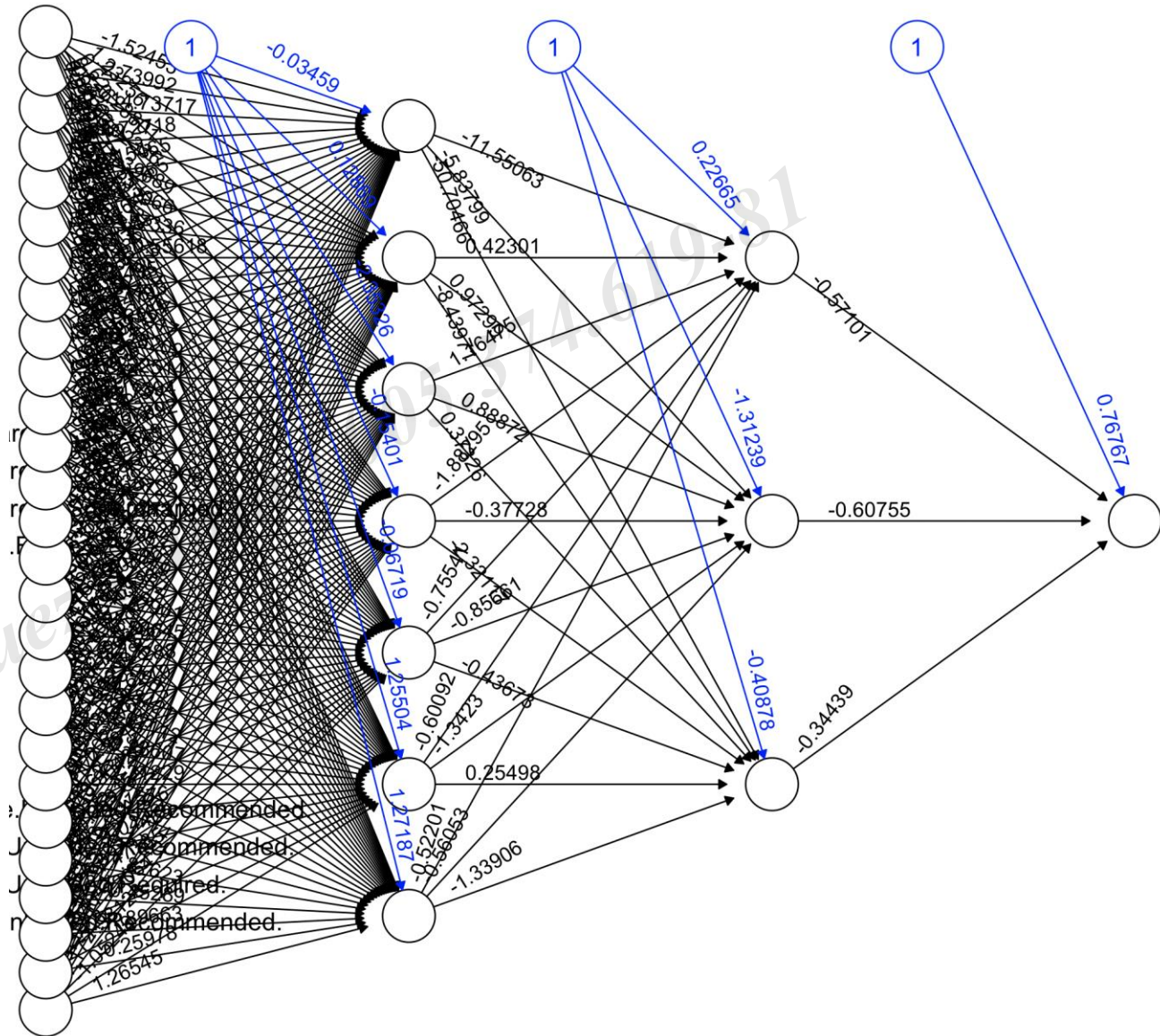


Luiz Rodriguez Fantini 005.374.619-81



Perceptrón multicapa

- Posee capas “ocultas” intermediarias
- Captura patrones no lineales
- Puede aprovecharse del procesamiento paralelo de GPUs
- No es “interpretable” como la regresión



Luiz Rodriguez Fantini 005.374.619-81



Funciones de pérdida

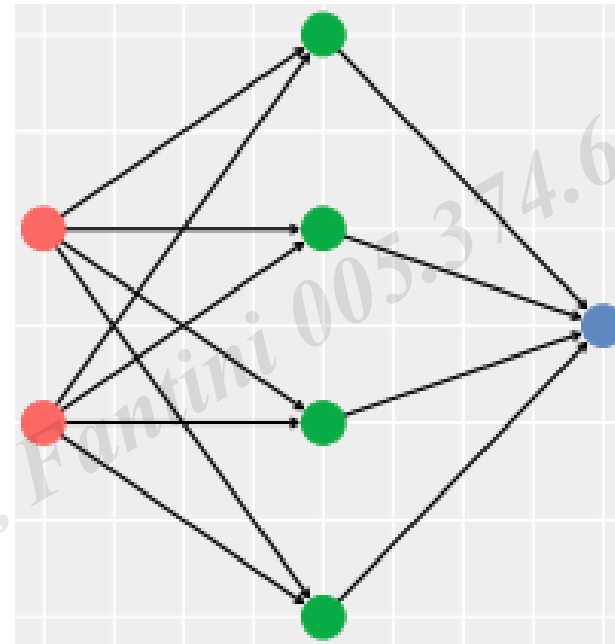
Variables Continuas
SQE

$$SQE = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

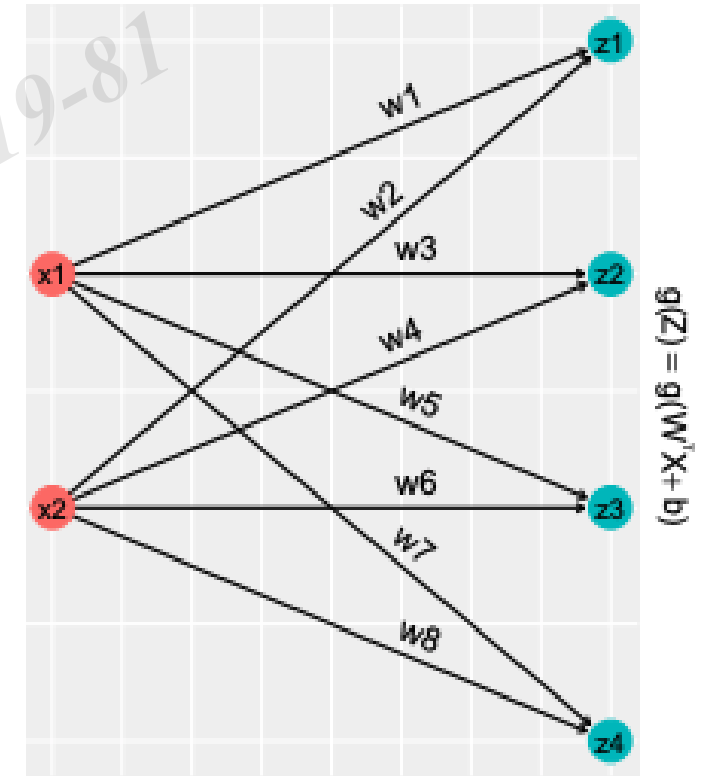
Variables binárias
Cross-Entropy

$$L = y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)$$

Redes Neurales Artificiales



layer ● Input ● Hidden ● Output



Deep learning with R - Abhijit Ghatak, ed. Springer, 2019

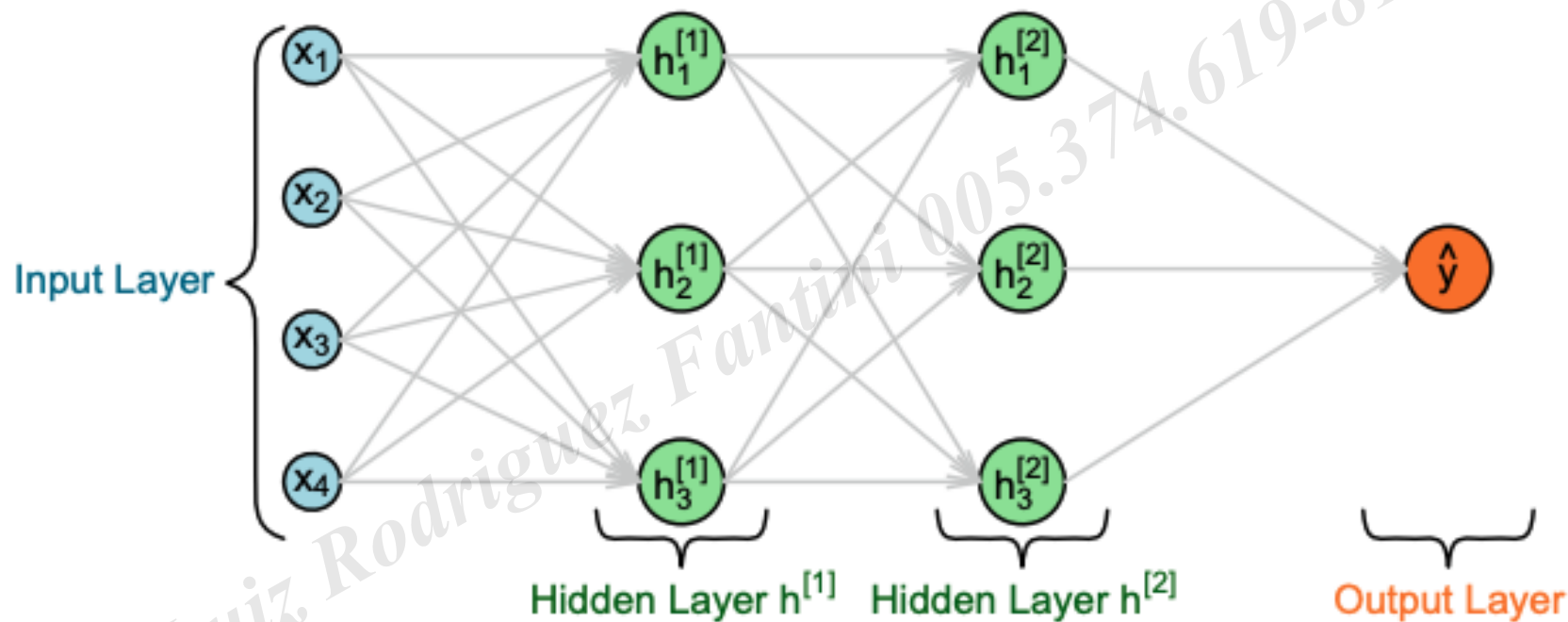
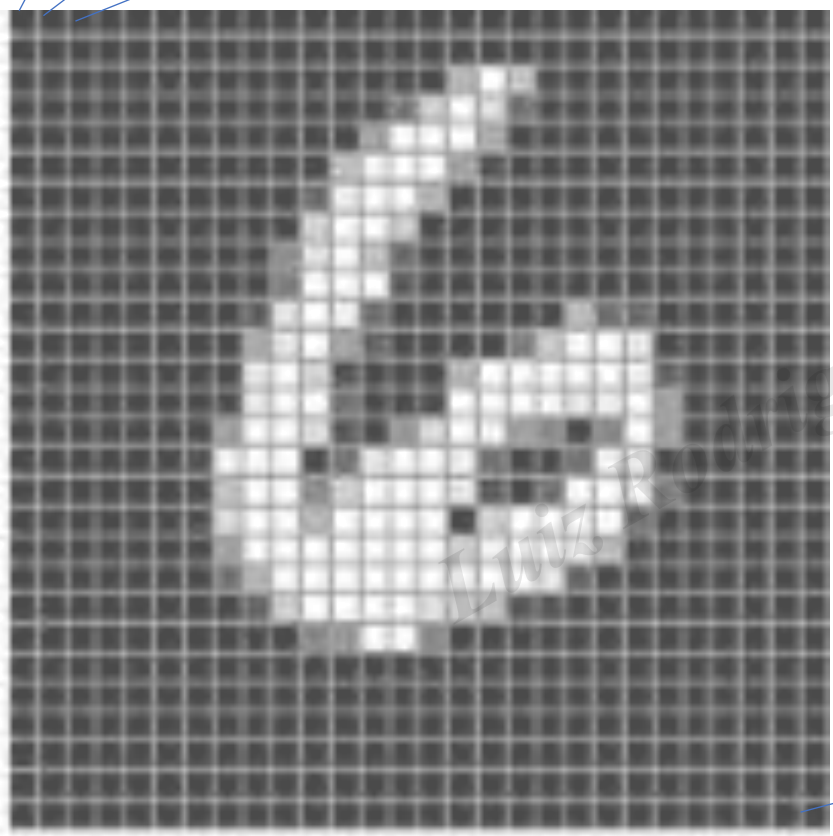


Fig. 2.3 A representation of a neural network with four input features, two hidden layers with three nodes each, and an output layer

Tratamiento inicial de los datos

Pixel 1
Pixel 2
Pixel 3



60.000 de esas

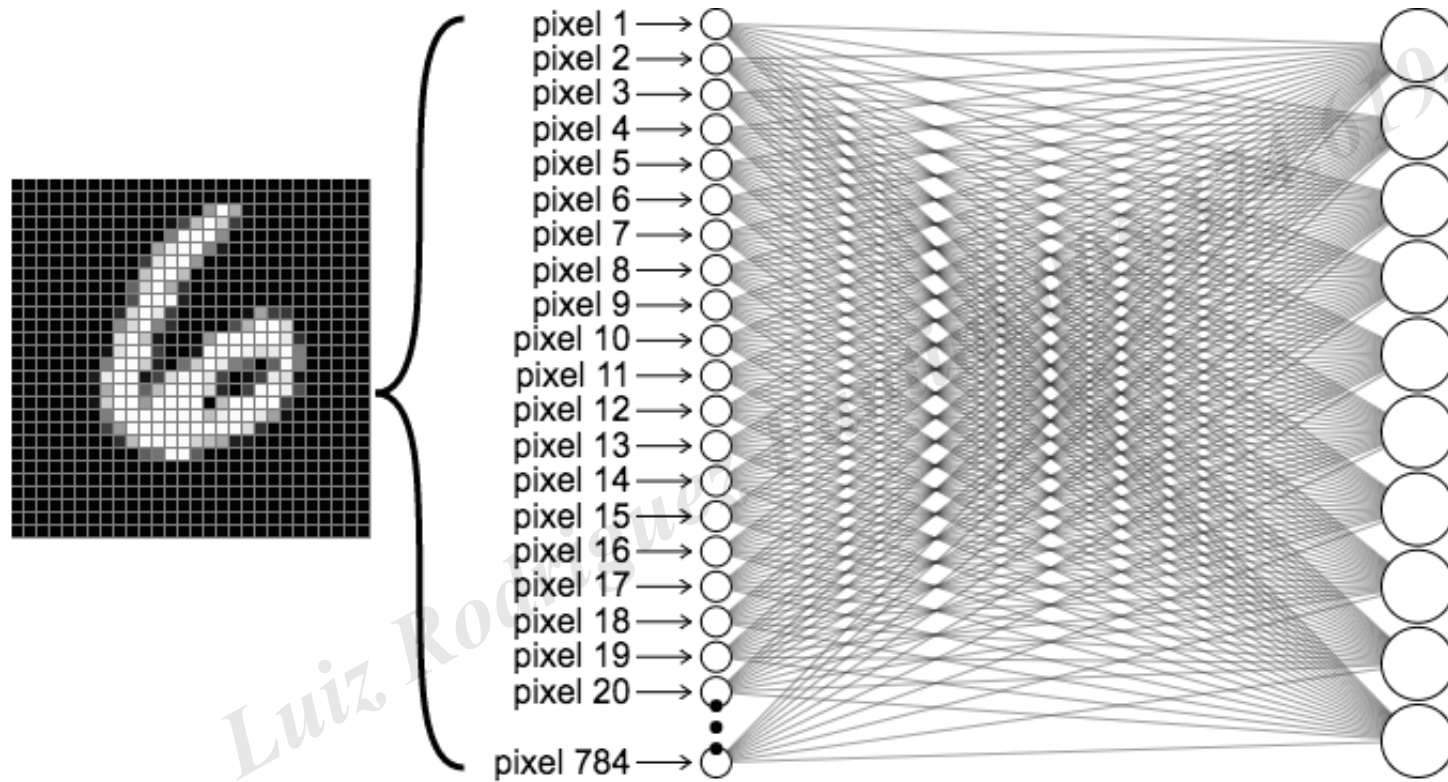


Se convierten
en una tabla así

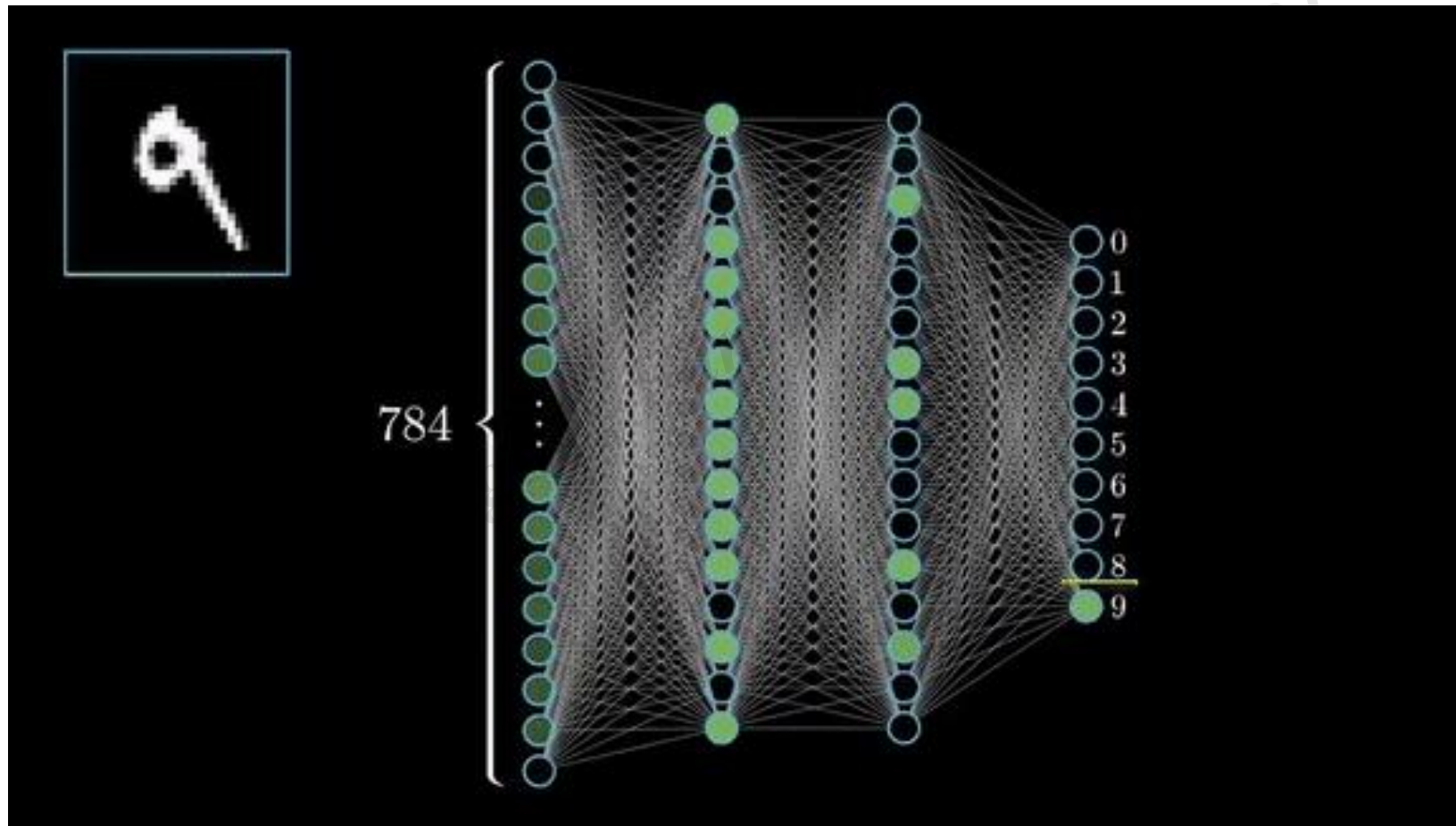
Pixel 784

	Pixel 1	Pixel 2	Pixel 3	⋮	Pixel 784	Rótulo
Imagen 1	0	0	0...		0	4
Imagen 2	0	0	0...		0	3
Imagen 3	0	0	0...		0	9
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Imagen 60.000	0	0	0...		0	5

Red Neural MNIST



¡Con apenas una capa ya tenemos $784 \times 10 = 7.840$ parámetros!



3blue1brown - <https://www.youtube.com/watch?v=aircAruvnKk>

Gradiente descendente

Es el algoritmo más popular para entrenar redes neurales artificiales por presentar algunas características:

- Puede cambiar las estimaciones con pequeños subconjuntos de puntos a cada iteración (en el límite 1 único punto)
- No depende de invertir matriz
- Funciona con una base de datos bien grande
- Puede ser procesado en paralelo con GPU
- Permite interrumpir el algoritmo a cierto punto y continuar más tarde o en otro problema similar (*transfer learning*)

Redes Gradiente Descendente

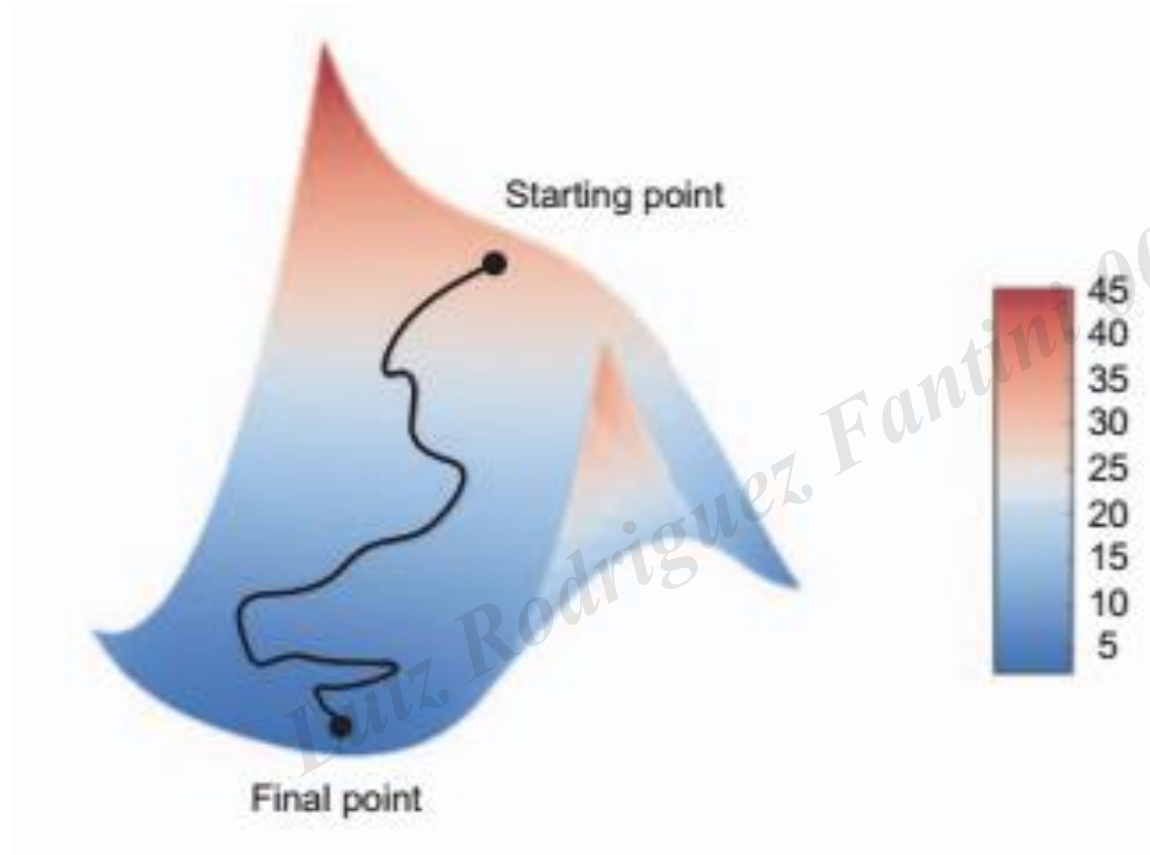


Figure 2.12 Gradient descent down a 2D loss surface (two learnable parameters)

Deep learning with python – François Chollet

Gradiente descendente

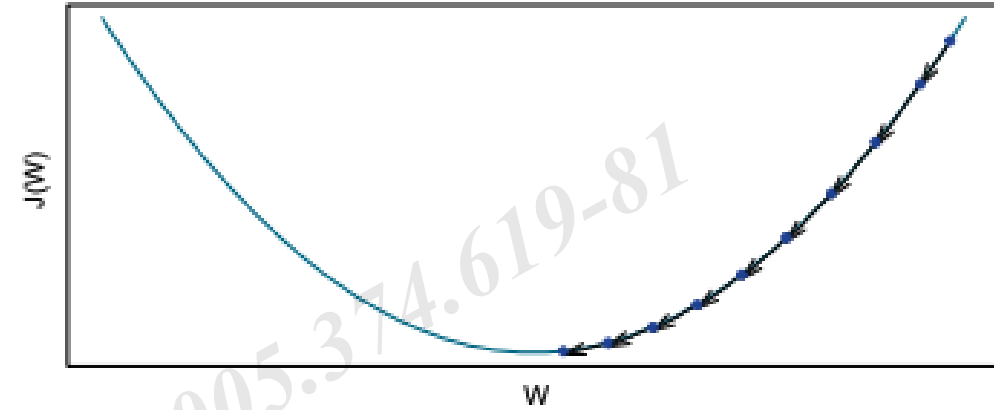


Fig. 1.4 Gradient descent: Rolling down to the minima by updating the weights by the gradient of the loss function

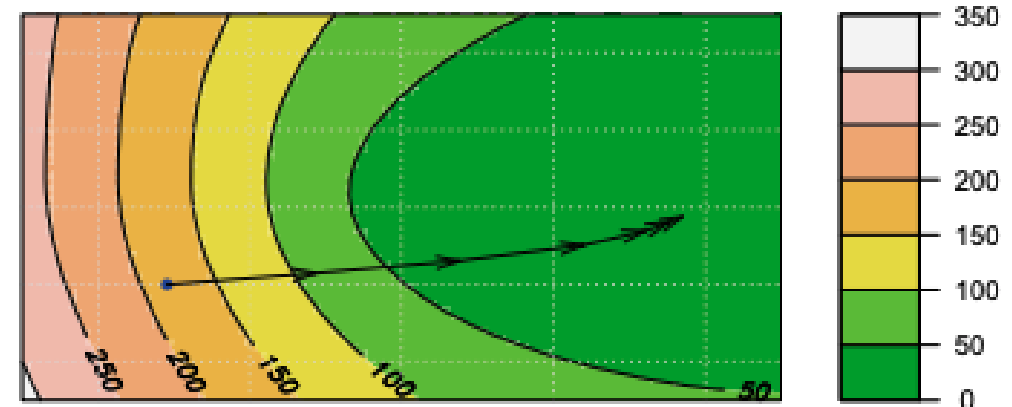


Fig. 1.5 A contour plot showing the cost contours of a sigmoid activation neural network and the cost minimization steps using the gradient descent optimization function

Previsión de consumo de vehículo

- Tamaño del motor
- Combustible
- Número de cilindros
- Marca
- Potencia
- Tracción



Luiz Rodriguez Fantini 005.374.619-81



Juego

La contribución de la industria del Juego en las Redes Neurales



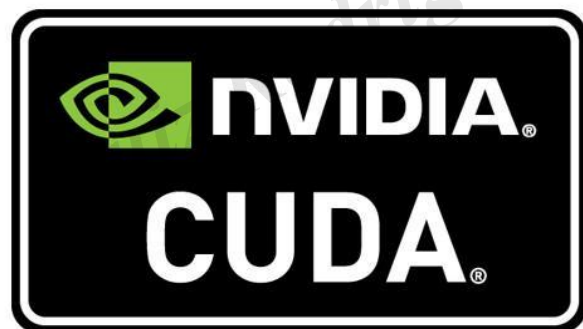
Procesadores

- Distancia entre transistores: 14 nm
- Cabello humano: 80.000 nm
- Diámetro del átomo de oro: 0,3 nm

A close-up, macro photograph of a GPU (Graphics Processing Unit) chip mounted on a printed circuit board (PCB). The chip is a dark, square component with a grid of gold-plated pins along its edges. The PCB is populated with various other components, including capacitors, resistors, and smaller integrated circuits, all in a dense arrangement. The lighting is soft, highlighting the metallic surfaces and the intricate patterns of the circuitry. The background is blurred, emphasizing the central GPU chip.

GPU

Procesamiento con GPU



AMD
ROCm

The background of the slide is a detailed, close-up photograph of a printed circuit board (PCB). The board is a deep blue color, and it is covered with a complex network of fine, gold-colored conductive traces. These traces form various patterns, including straight lines, curves, and dense grids. Scattered throughout the board are numerous small, circular components that appear to be glowing with a bright blue light, giving the image a high-tech, futuristic feel. The lighting is focused, creating highlights on the metallic surfaces and deep shadows in the recesses of the board.

TPU

Regularización L2

Variables Continuas
SQE

$$SQE = \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \sum \beta_i^2$$

Variables binárias
Cross-Entropy

$$L = \sum y_i \log(\hat{y}_i) + \lambda \sum \beta_i^2$$



Reconocimiento de actividad humana con smartphone

Luiz Rodriguez Fantini 005.374.619-81



Conclusiones

- Redes Neurales son la introducción al Deep Learning (que es un ramo muy promisor)
- Son poderosas y flexibles
- Requieren poder computacional especial (GPU / TPU)
- Son famosas en datos menos estructurados (ej. imágenes, audio)





“El mundo está en constante cambio, por eso no se apegue a nada.”

“Cuando evolucionas el mundo evoluciona”

Sidarta Gautama



Por hoy es sólo
eso ;)



[linkedin.com/in/joao-serrajordia](https://www.linkedin.com/in/joao-serrajordia)