

Neural Networks, Deep Learning and Applications

Alan Gerardo Reyes Figueroa

Centro de Investigación en Matemáticas

CIMAT



CIMAT

Centro de Investigación en Matemáticas, A.C.

Universidad del Valle de Guatemala

Agosto 07, 2018



UVG

UNIVERSIDAD
DEL VALLE
DE GUATEMALA

Inteligencia Artificial



G. Kasparov vs. DeepBlue
(1997)



Lee Sedol vs. AlphaGo
(2016)

Inteligencia Artificial



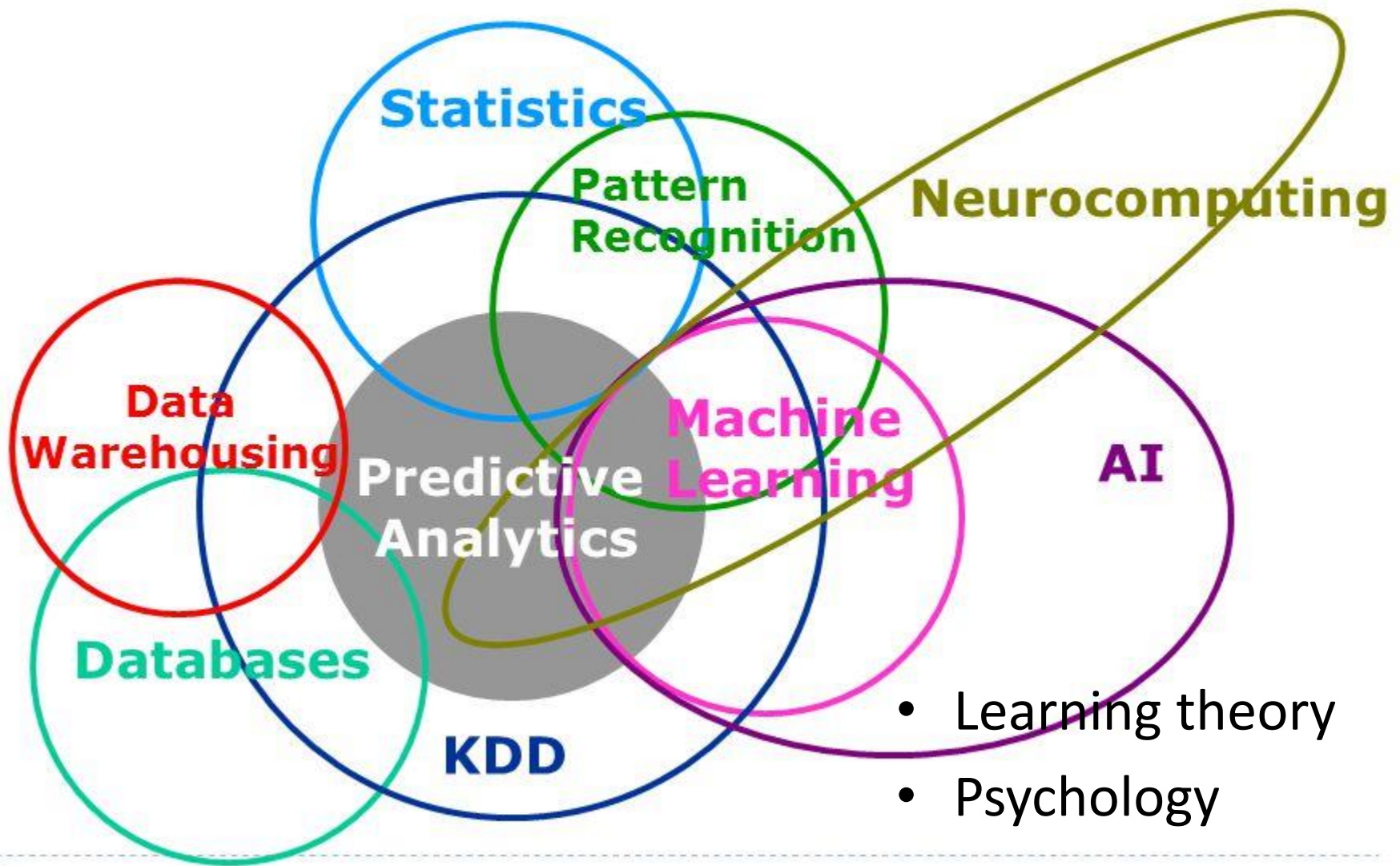
Self-driving cars.



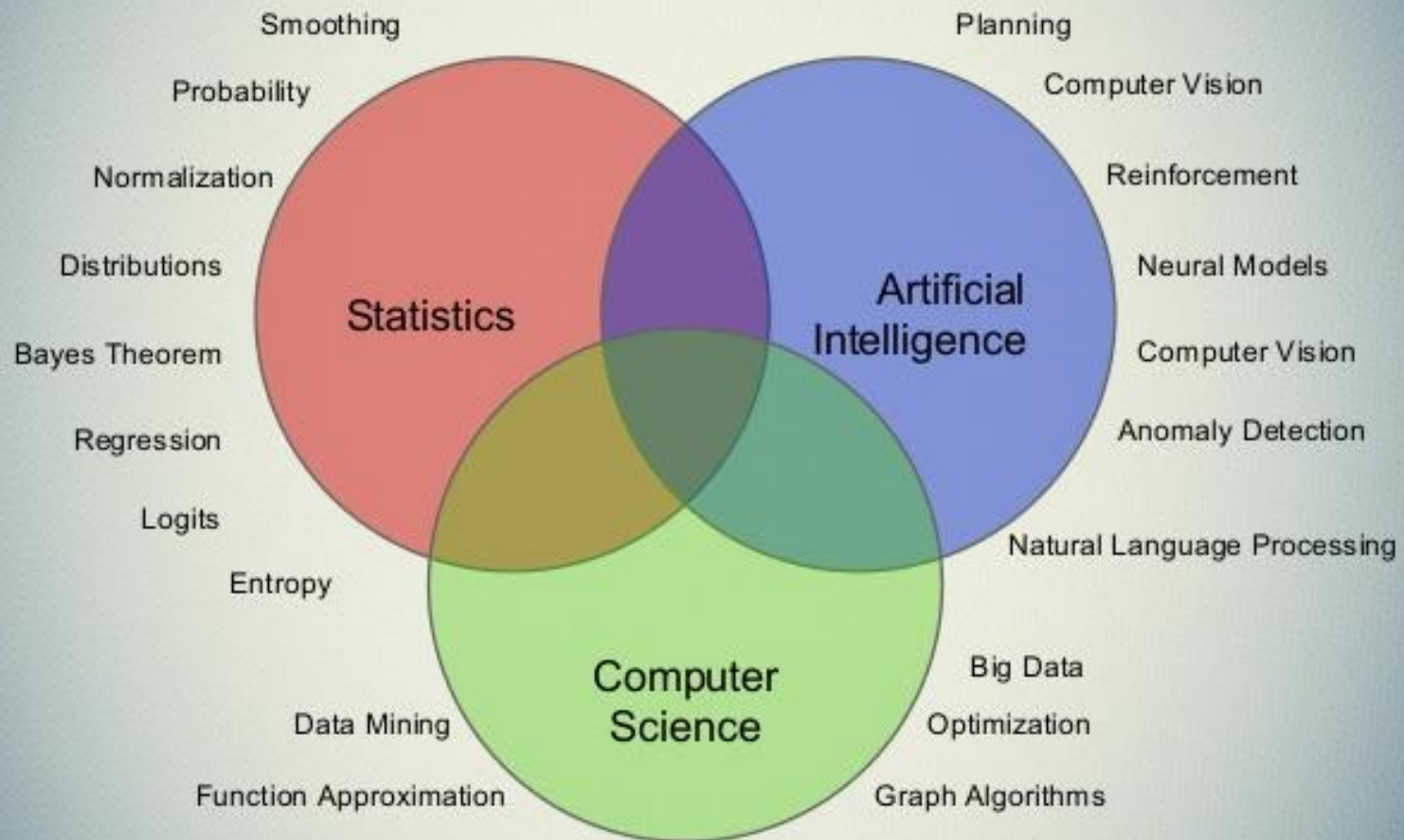
Inteligencia Artificial

- ¿En qué consiste?
- Otros términos:
 - big data
 - machine learning
 - deep learning
 - analytics
 - cognitive systems ...

Inteligencia Artificial



Inteligencia Artificial



Is Machine Learning a one semester course?

Inteligencia Artificial

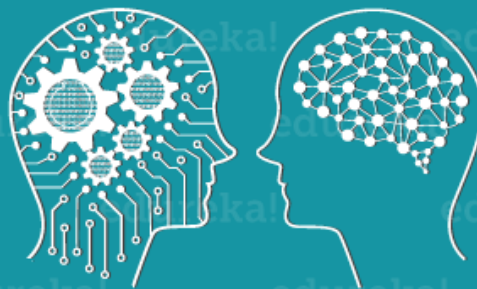
ARTIFICIAL INTELLIGENCE

Engineering of making Intelligent Machines and Programs



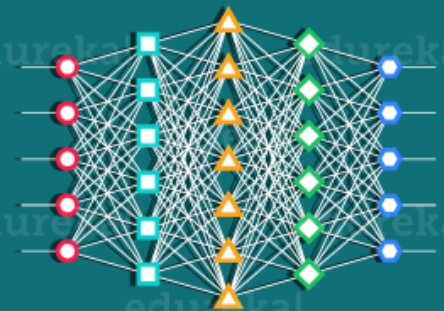
MACHINE LEARNING

Ability to learn without being explicitly programmed



DEEP LEARNING

Learning based on Deep Neural Network



1950's

1960's

1970's

1980's

1990's

2000's

2006's

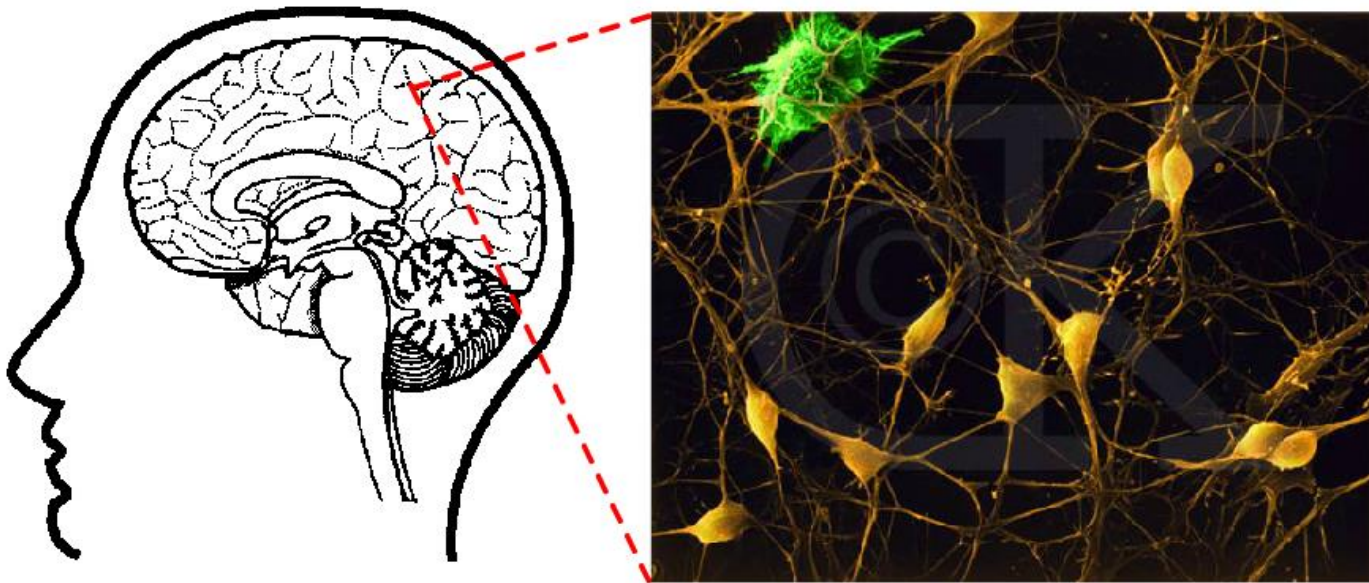
2010's

2012's

2017's

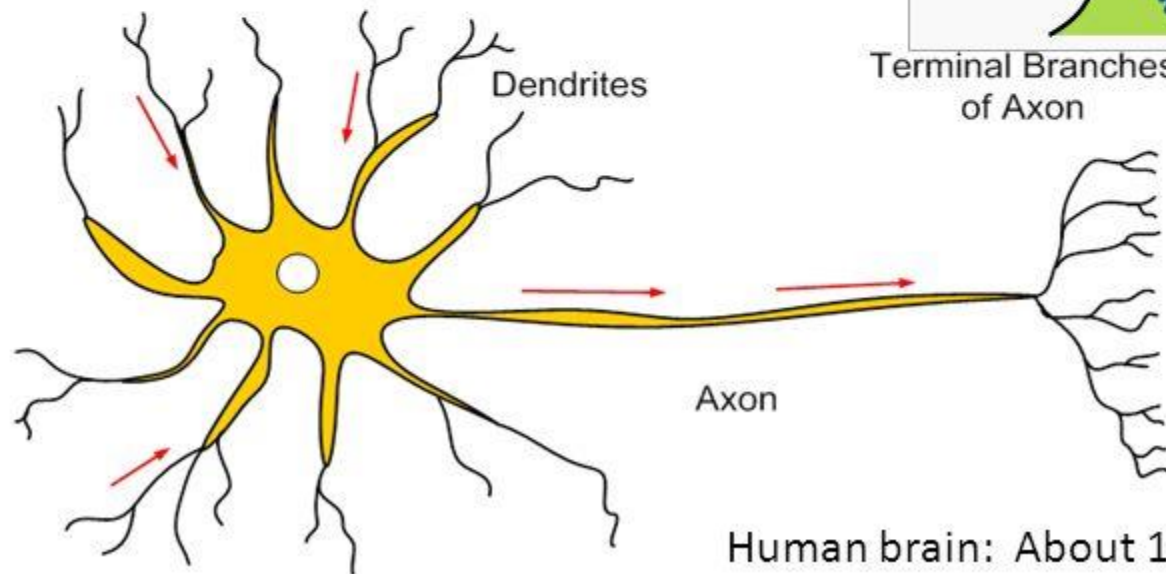
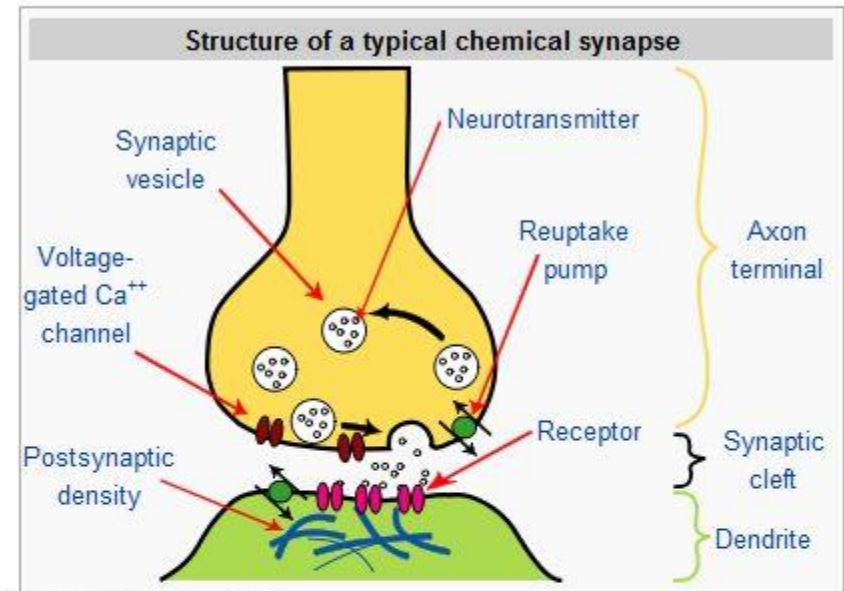
Redes Neuronales

- Algoritmos bioinspirados.
- Idealizan cómo funcionan las conexiones y transmisiones dentro del cerebro biológico.



Biologically Inspired.

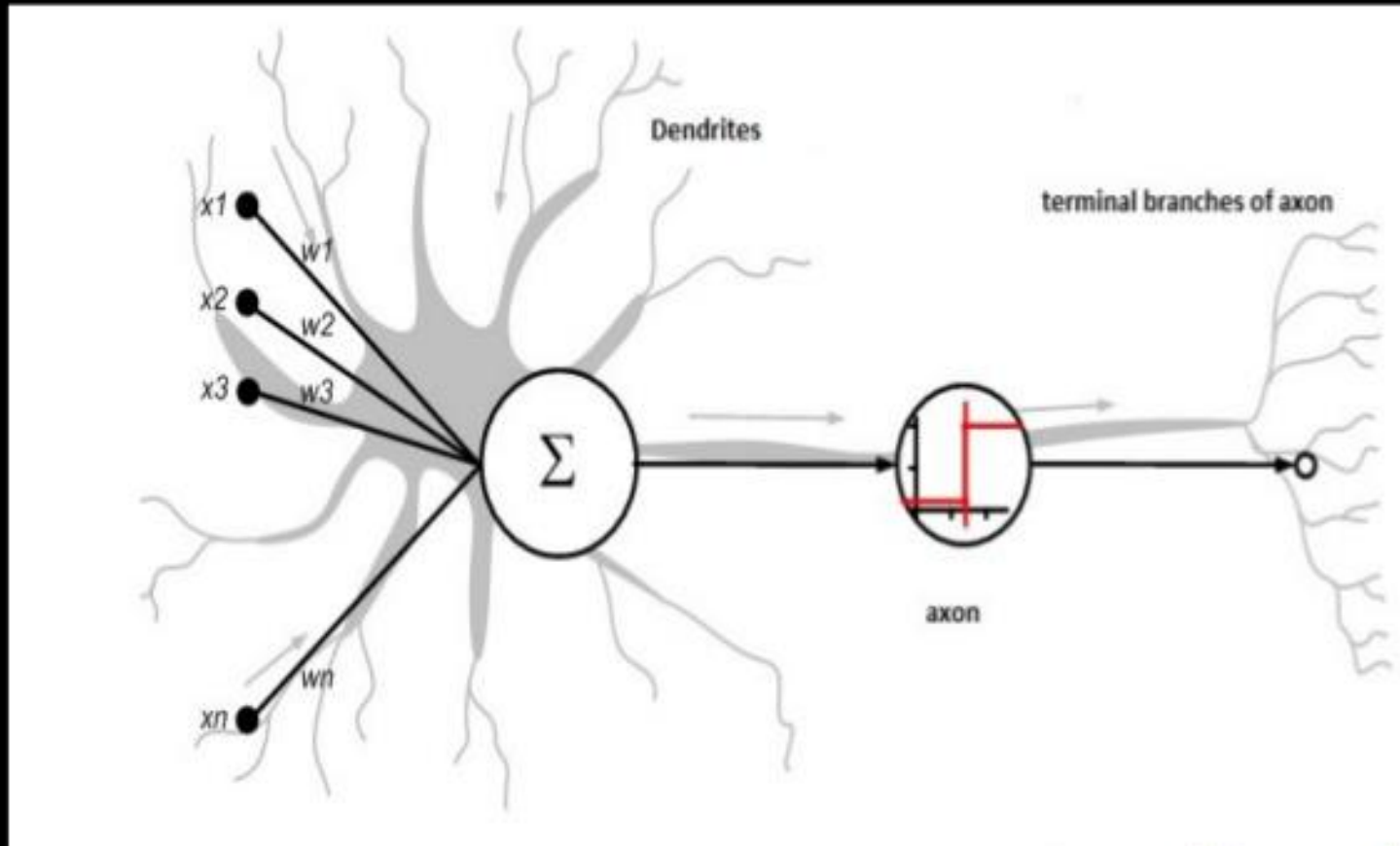
- Electro-chemical signals
- Threshold output firing



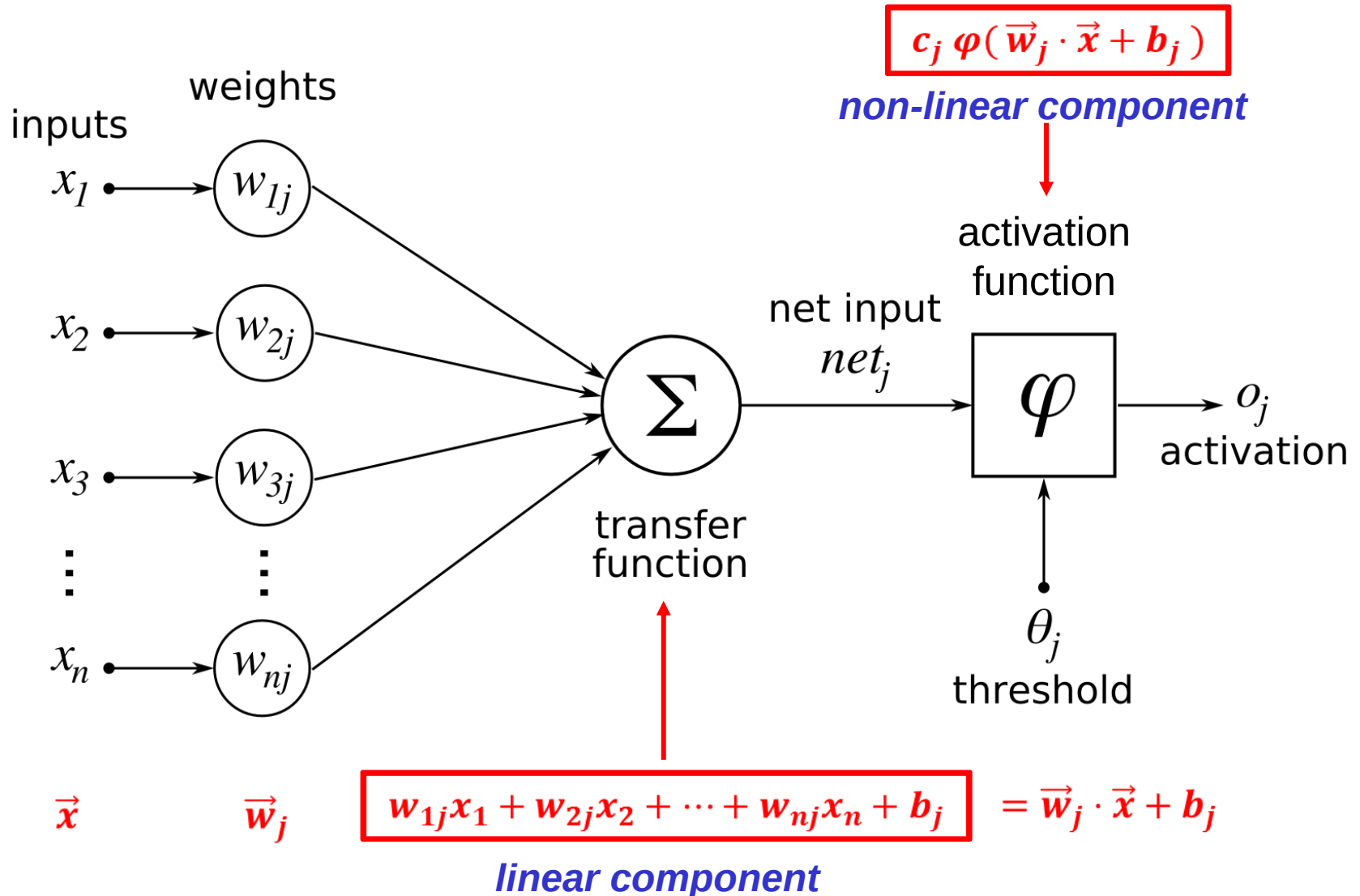
Human brain: About 100 billion (10^{11}) neurons and 100 trillion (10^{14}) synapses

Redes Neuronales Artificiales

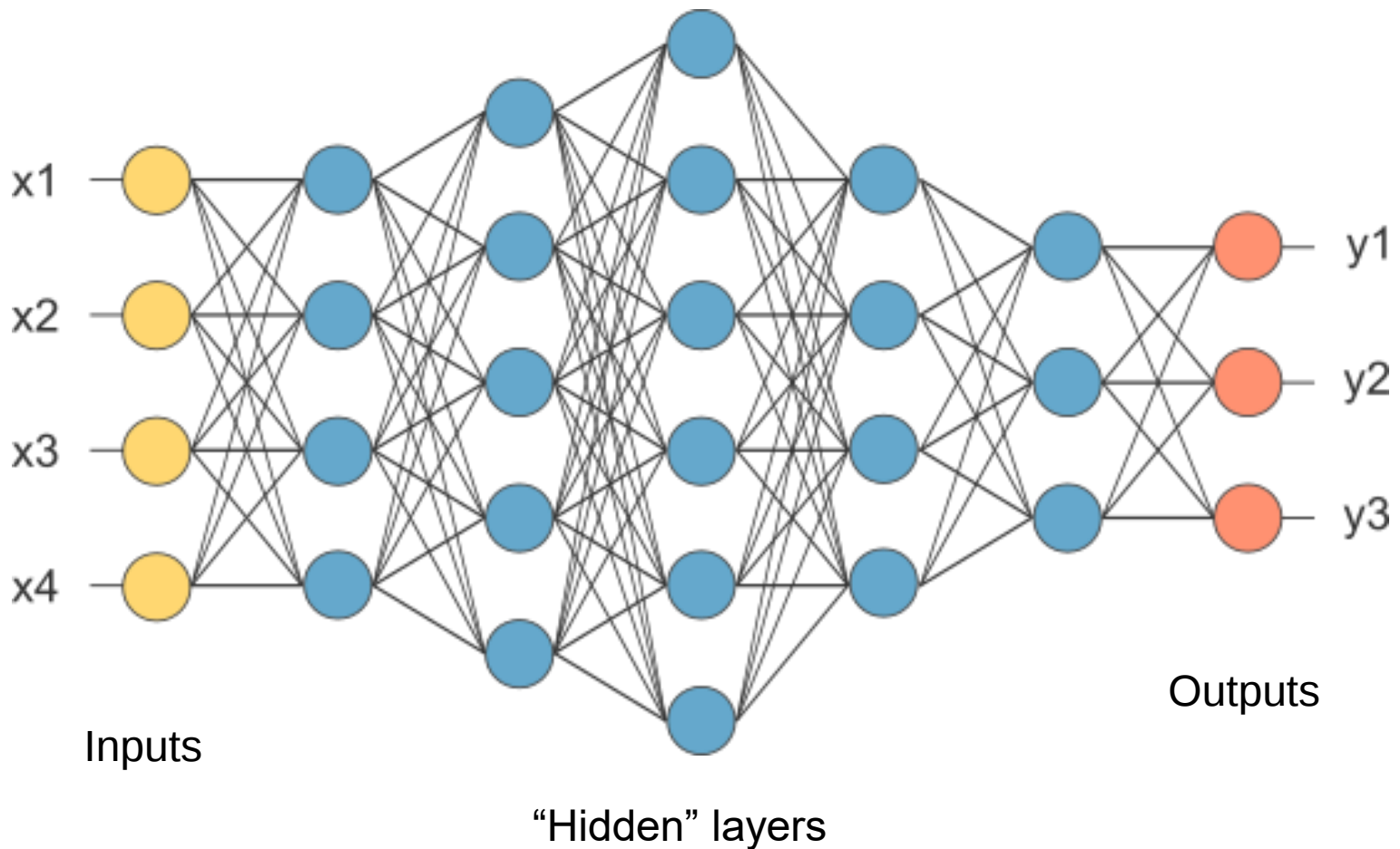
Biologically Inspired Neuron



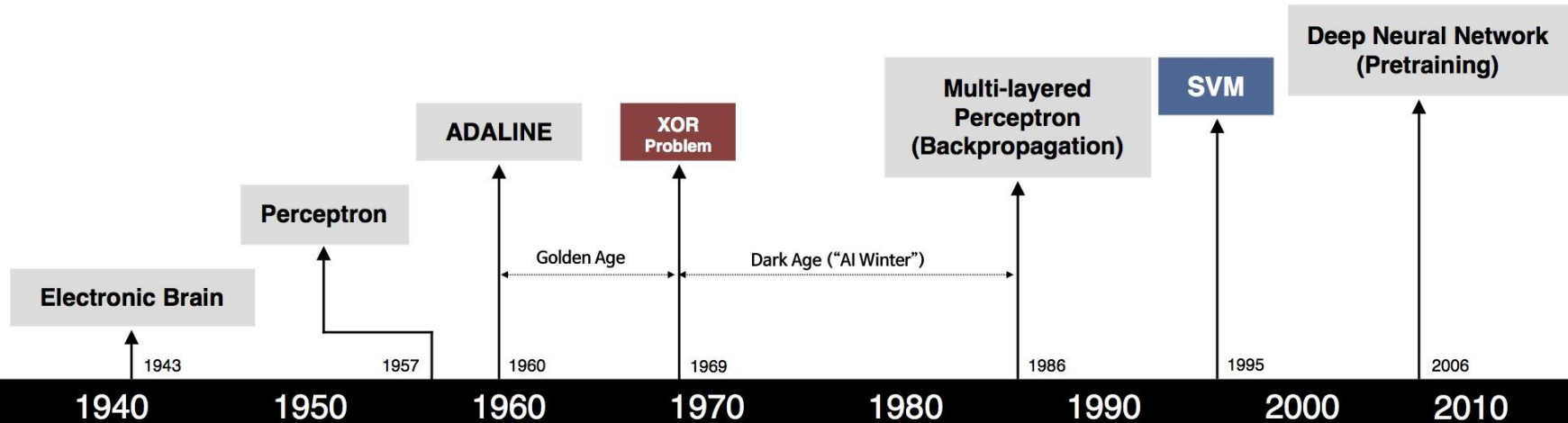
Redes Neuronales Artificiales



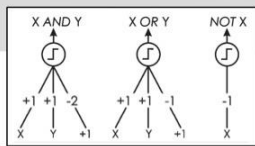
Redes Neuronales Artificiales



Desarrollo Redes Neuronales



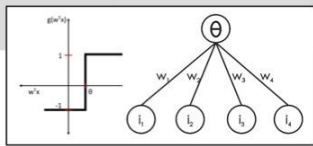
S. McCulloch - W. Pitts



- Adjustable Weights
- Weights are not Learned



F. Rosenblatt



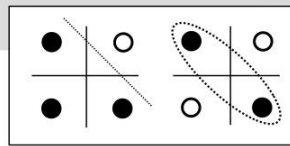
- Learnable Weights and Threshold



B. Widrow - M. Hoff



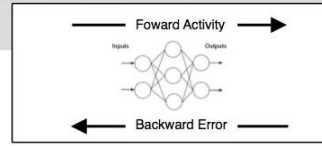
M. Minsky - S. Papert



- XOR Problem



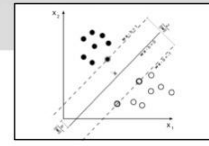
D. Rumelhart - G. Hinton - R. Williams



- Solution to nonlinearly separable problems
- Big computation, local optima and overfitting



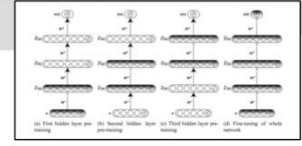
V. Vapnik - C. Cortes



- Limitations of learning prior knowledge
- Kernel function: Human Intervention

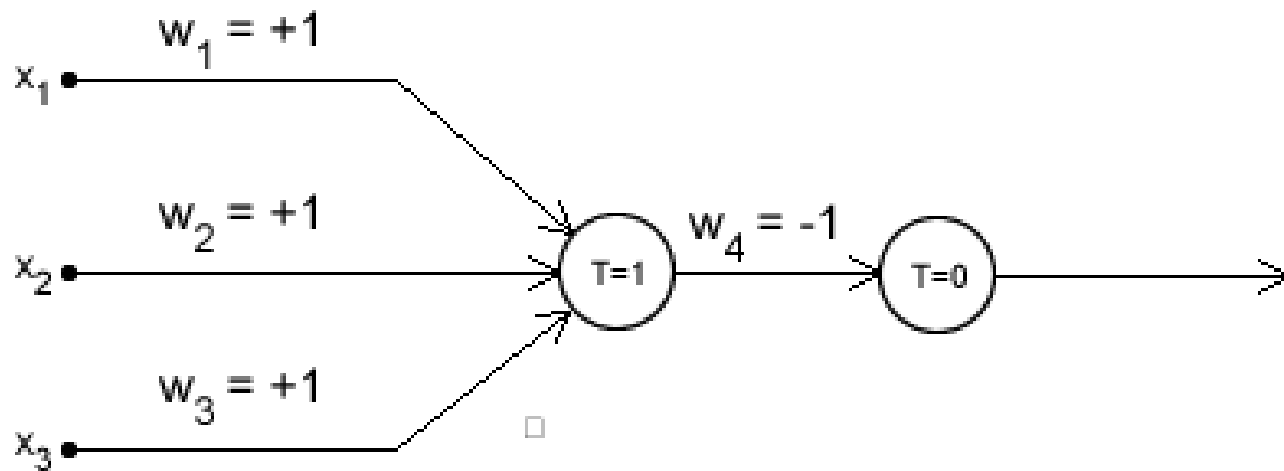


G. Hinton - S. Ruslan



- Hierarchical feature Learning

McCulloch-Pitts Model (1940's)



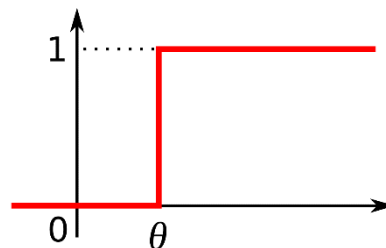
McCulloch-Pitts Model (1940's)

- Llamado *Logical Threshold Unit*.
- Inputs $x_j = 0$ ó $x_j = 1$.
- Pesos w_{ij} fijos.

$w_{ij} = 1$: excitador

$w_{ij} = -1$: inhibidor

- Función de activación tipo umbral:



Rosenblatt's Model (1960's)

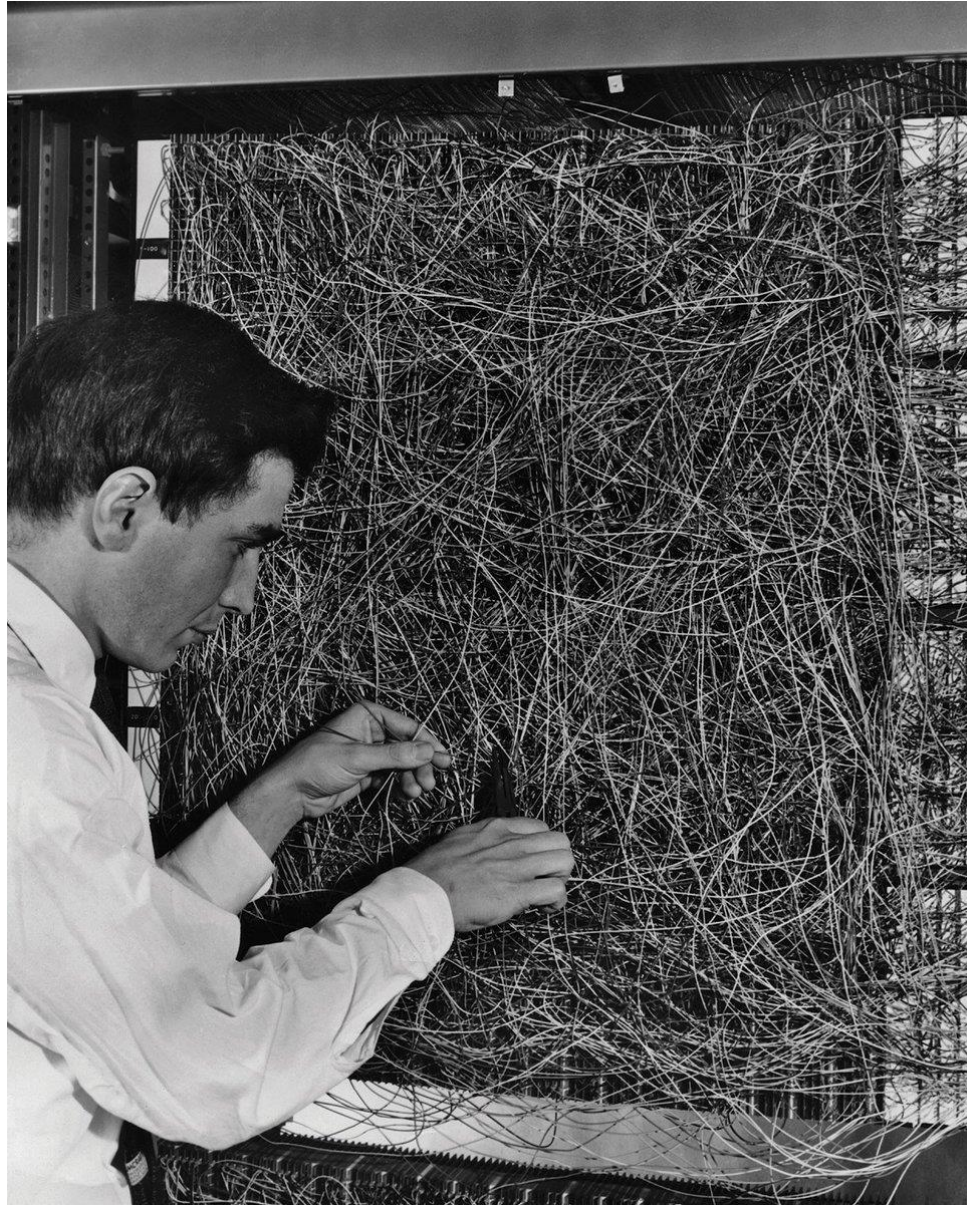
- Llamado *Perceptron*.
- Inputs variables.
- Pesos w_{ij} variables:

Los pesos pueden ser aprendidos
(mediante algoritmos).

- Función de activación tipo umbral
- Clasificador binario

$$\varphi(x) = \text{sign}(w \cdot x + b)$$

Rosenblatt's Model (1960's)



Rosenblatt's Model (1960's)

Algoritmo iterativo:

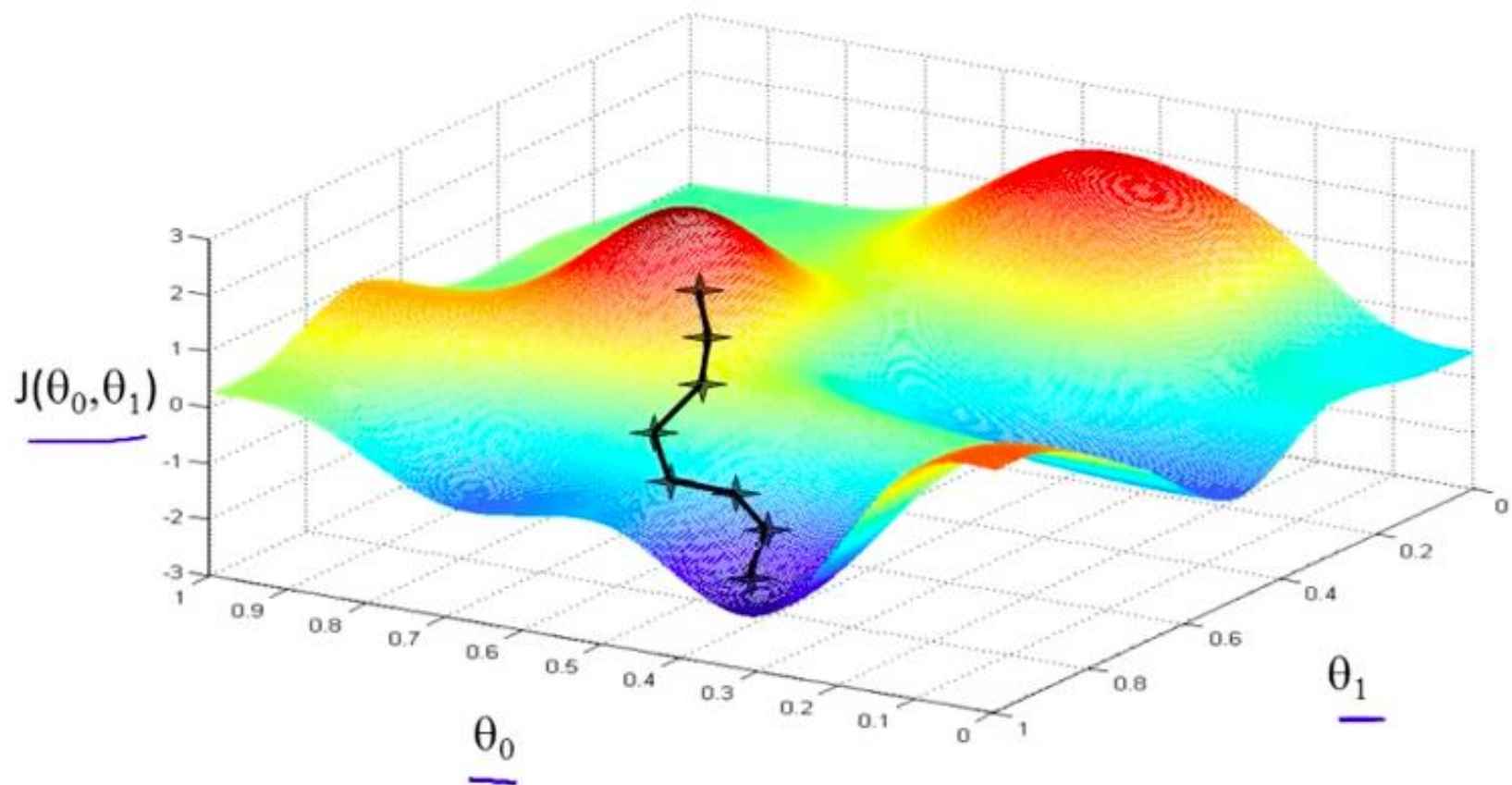
- Dados (x_i, y_i) y pesos iniciales w_j, b aleatorios, se calcula

$$\hat{y}_i = \varphi(w \cdot x_i + b)$$

- Los pesos w_{ij}, b se actualizan usando descenso gradiente

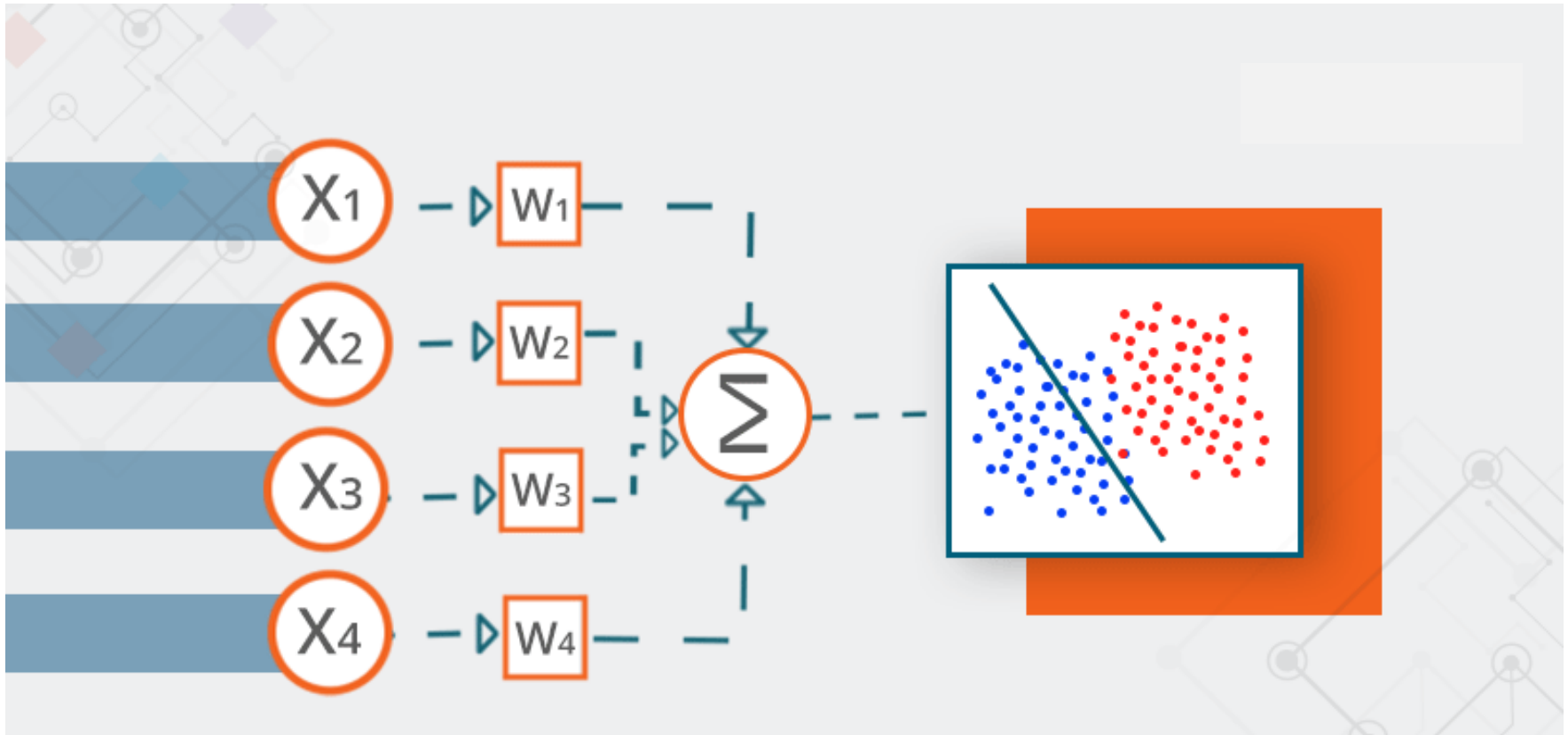
$$w_j^{(t+1)} = w_j^{(t)} + \alpha \sum_{i=1}^n (y_i - \hat{y}_i) x_{ij}$$

Gradient Descent



$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha \nabla J(\mathbf{x}_k)$$

Rosenblatt's Model (1960's)

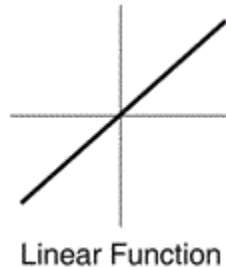


Perceptron Learning Algorithm

Widrow-Hoff Model (1960's)

Similar al modelo de Rosenblatt, pero

- Función de activación lineal

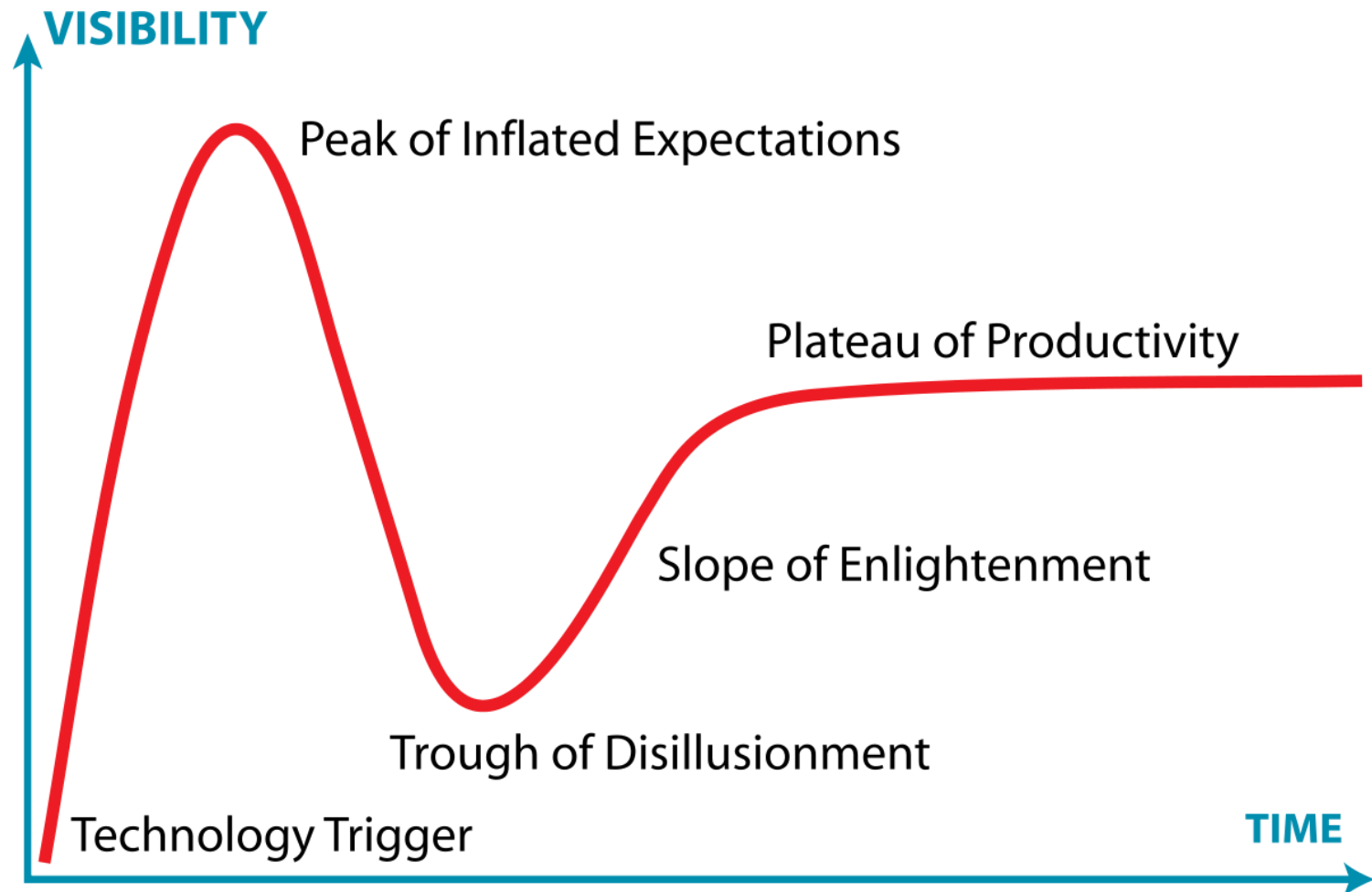


$$\hat{y}_i = \varphi(w \cdot x_i + b) = w \cdot x_i + b$$

- Algoritmo de actualización ADALINE
(**Adaptive Linear Neuron**)

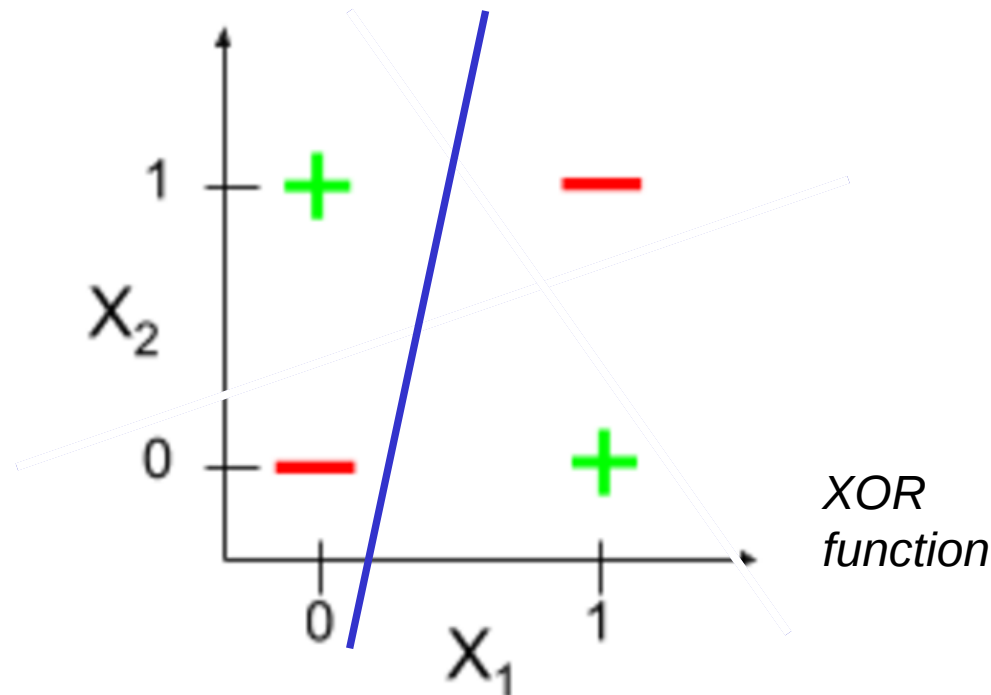
The Golden Age (1960's)

- Curva de Gartner para una tecnología



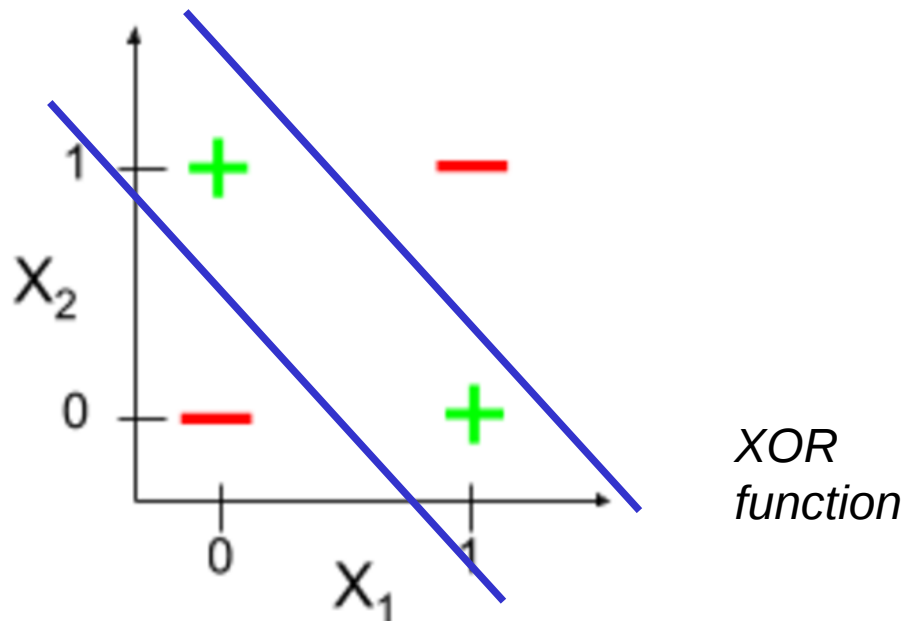
Dark Ages (1970's)

- Minsky y Papert (1969) publican ***Perceptrons: an introduction to computational geometry.***
- Crítica y limitaciones al modelo perceptrón.
- “Invierno” de la inteligencia artificial.

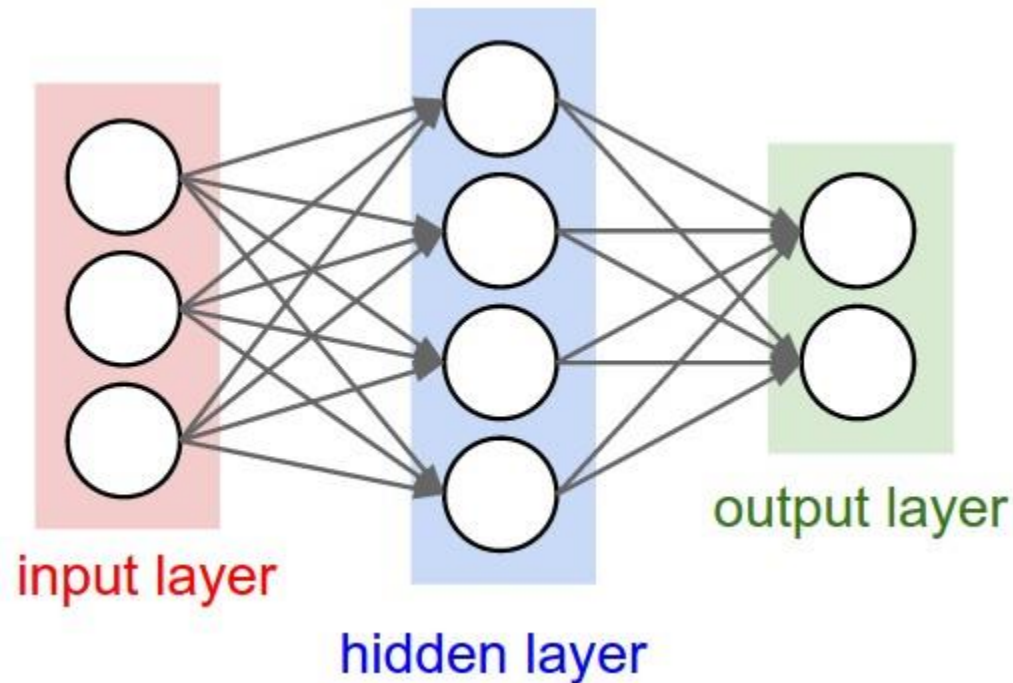


Segunda Era (1980's)

- Rumelhart, Hinton y Williams (1986)
Nature: Learning representations by back-propagating errors.
- Solución al problema de la separabilidad.



Segunda Era (1980's)

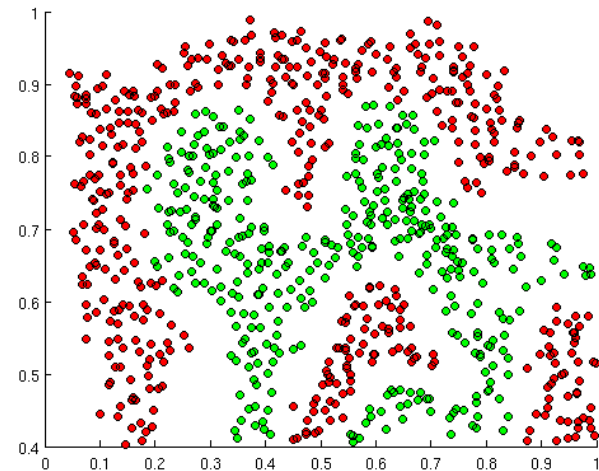
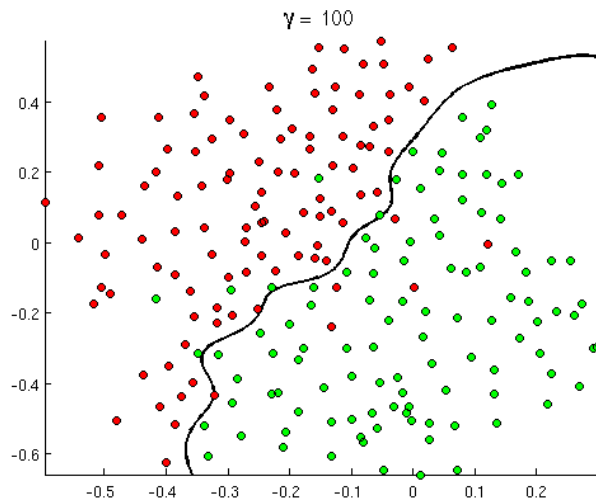
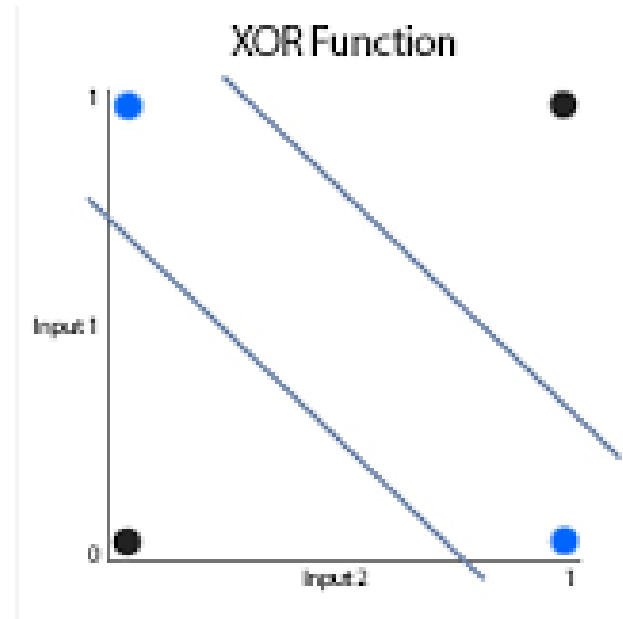
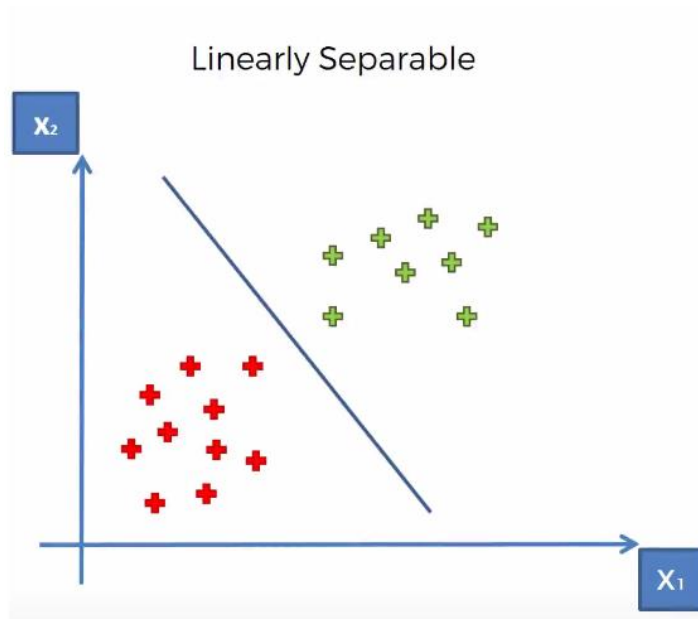


- Redes multicapa.
- Algoritmo de ***back-propagation***.

Segunda Era (1980's)

- K. Fukushima (79')
NeoCognitron.
Trabajos en redes multicapa.
- Teoría matemática de redes neuronales.
- G. Cybenko (1989).
Teorema de representación universal.

Aproximadores universales



Teorema de Cybenko

- **Teorema de Aproximación Universal (Cybenko '89).**

Sea $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ continua, no-constante, limitada, y monótona creciente. Sea $I_m = [0,1]^m$ el cubo m -dimensional, y $C(I_m)$ el espacio de funciones continuas en I_m . Dado $\varepsilon > 0$ y cualquier función $f \in C(I_m)$, existe un entero N , constantes $c_i, b_i \in \mathbb{R}$, y vectores $\mathbf{w}_i \in \mathbb{R}^m$, tales que la función

$$F(x) = \sum_{i=1}^N c_i \varphi(\mathbf{w}_i \cdot x + b_i)$$

satisface $|F(x) - f(x)| < \varepsilon$, para todo $x \in I_m$.

Teorema de Aproximación

- Si la función de activación φ satisface “ciertas condiciones”, las redes multi-capa son capaces de aproximar cualquier función.
(X compacto)
- Las redes neuronales se pueden adaptar a cualquier conjunto (finito) de datos.

Teorema de Aproximación

- Composición de transformaciones lineales:

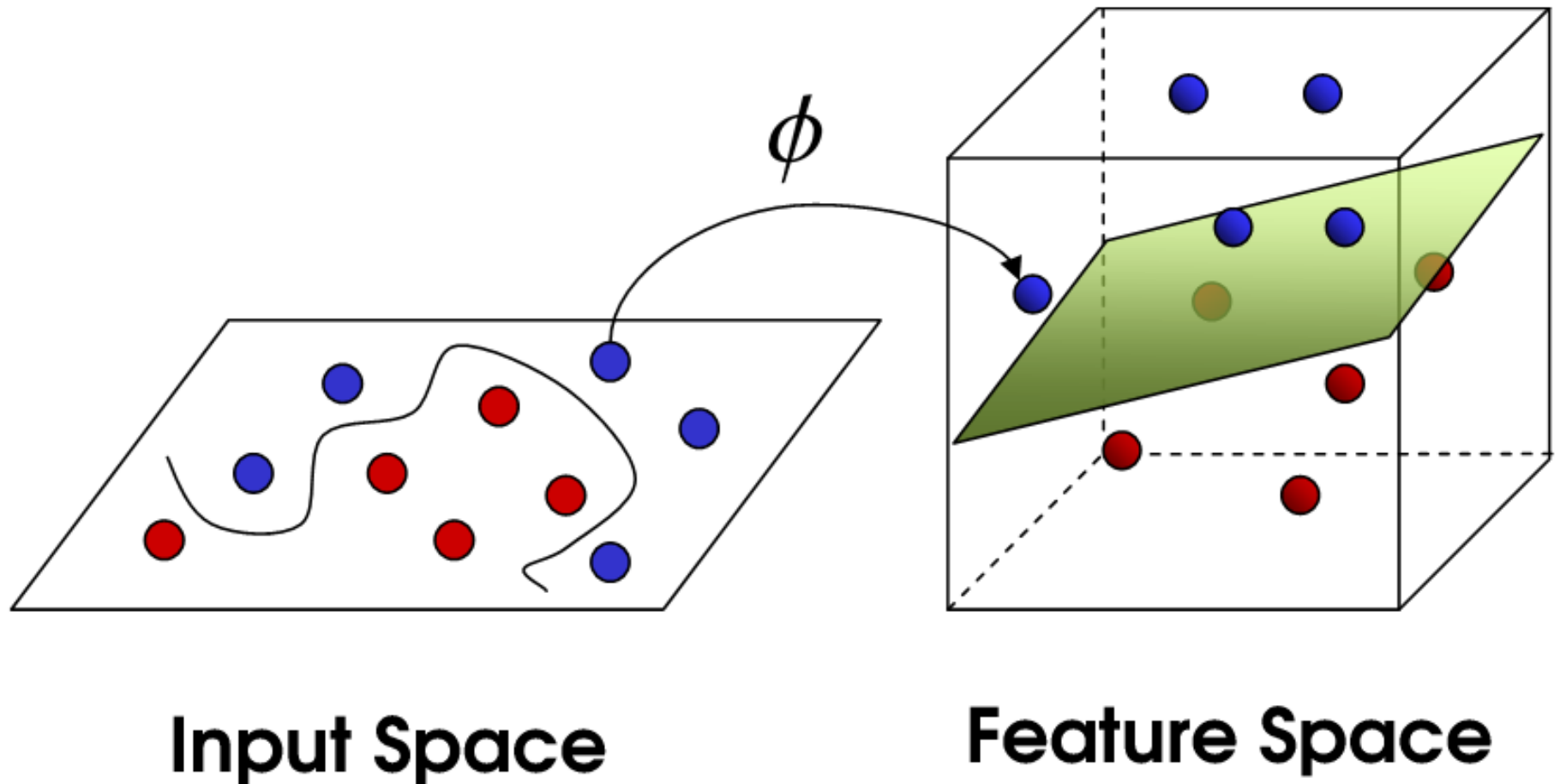
$$T_1(\vec{x}) = W_1\vec{x} + \vec{b}_1$$

$$T_2(\vec{x}) = W_2\vec{x} + \vec{b}_2$$

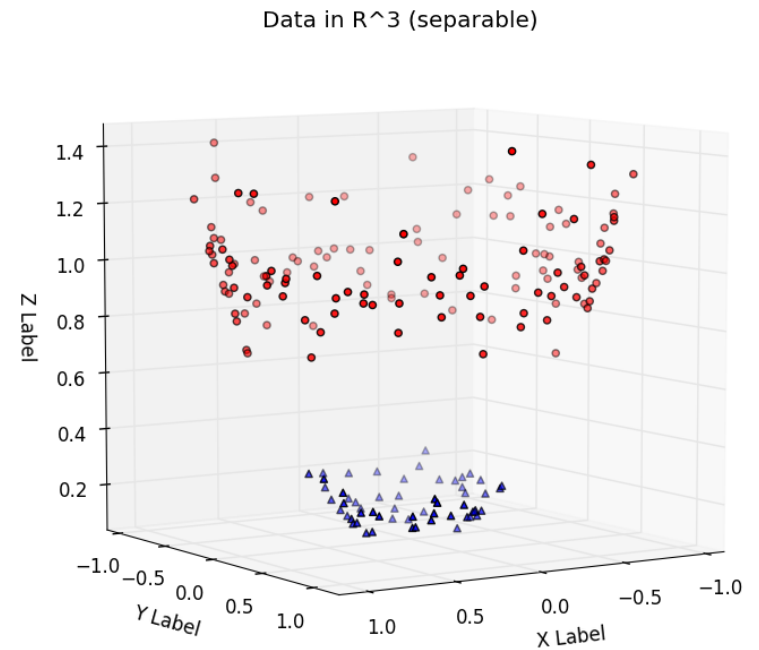
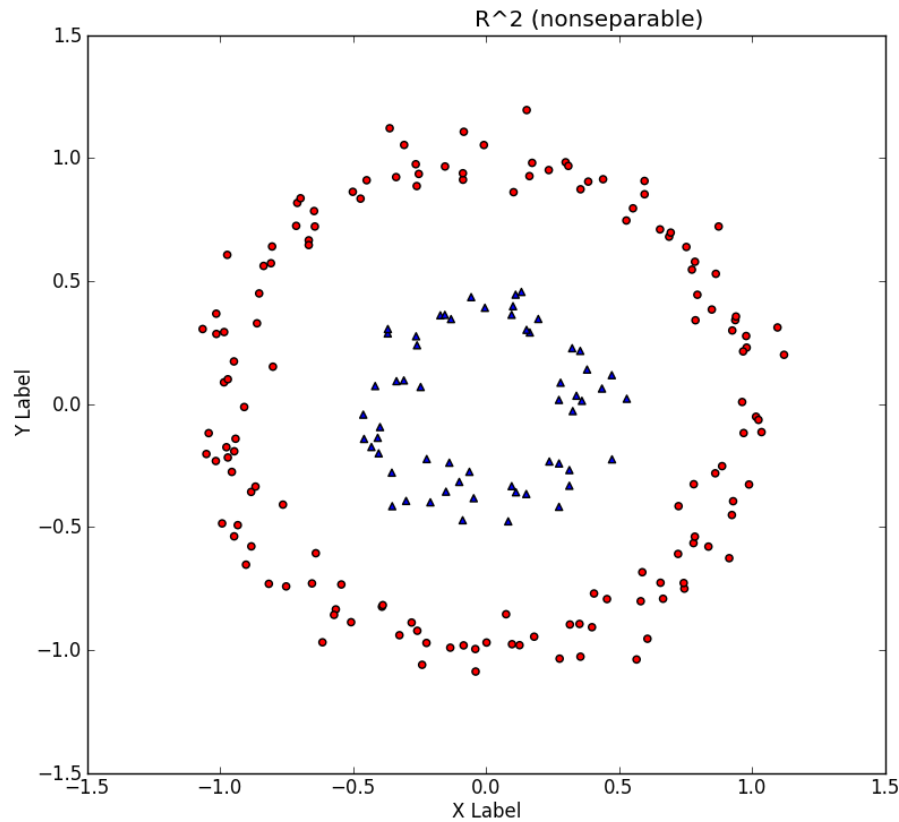
$$\begin{aligned}(T_2 \circ T_1)(\vec{x}) &= W_2(W_1\vec{x} + \vec{b}_1) + \vec{b}_2 \\ &= \underbrace{(W_2W_1)\vec{x} + (W_2\vec{b}_1 + \vec{b}_2)}_{\text{linear}}\end{aligned}$$

- Importante!:** composición del componente no-lineal.

Geometría de las redes neuronales

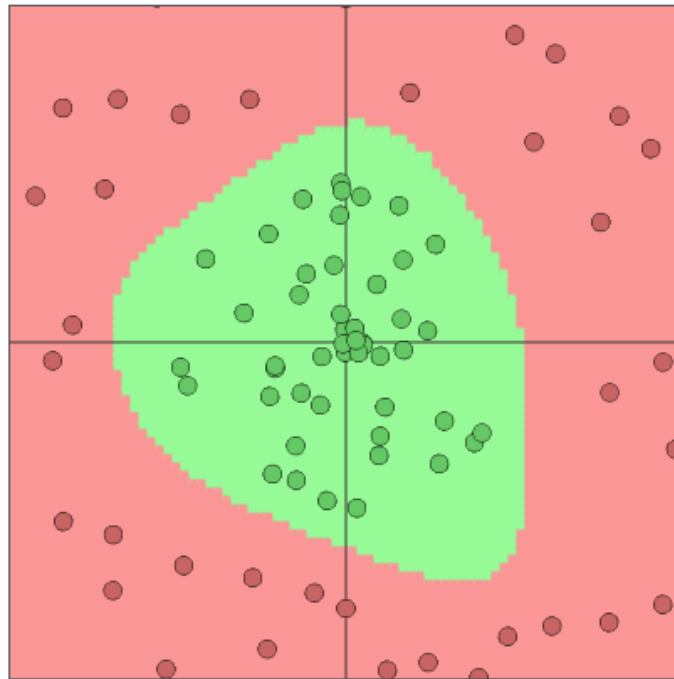


Ejemplo: Kernel Trick



Ejemplo

- Geometría de las redes neuronales.



simple data

circle data

spiral data

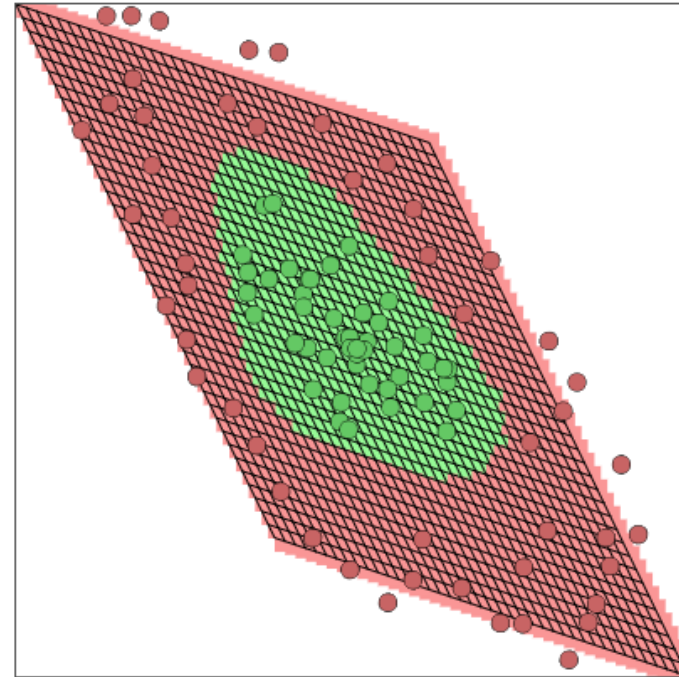
random data

Controls:

CLICK: Add red data point

SHIFT+CLICK: Add green data point

CTRL+CLICK: Remove closest data point



drawing neurons 0 and 1 of layer with index 1 (fc)

fc(6)

tanh(6)

fc(2)

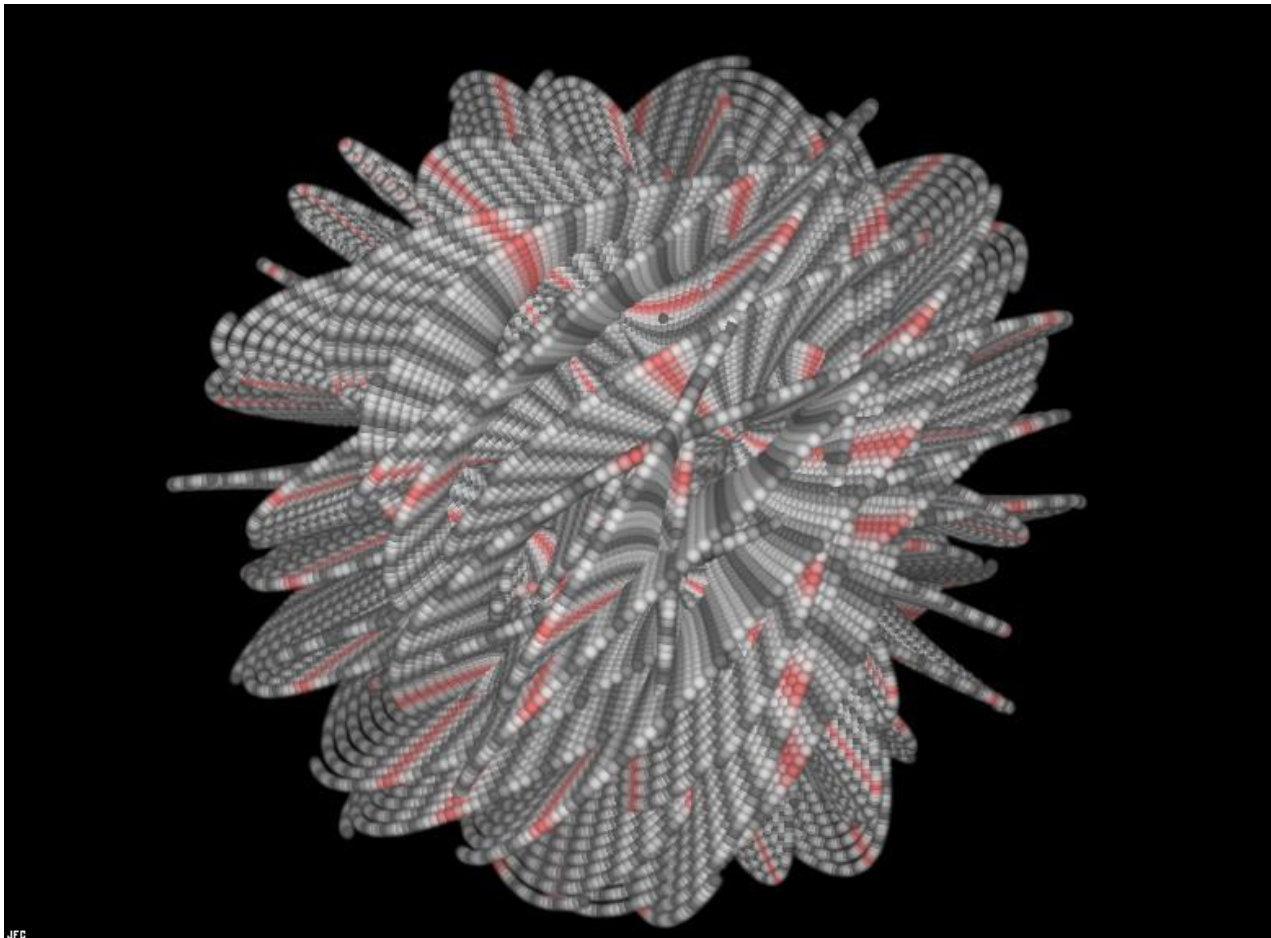
tanh(2)

fc(2)

cycle through visualized neurons at selected layer (if more than 2)

Manifold Hypothesis

- Datos - - - Estructura de variedad M .
- Red Neuronal - - - transforma $\mathbb{R}^m$ en M .



Second winter (1990's)

- Financiamiento sector militar.
- Sobre-expectativa en proyectos.
 - autos autónomos
 - máquinas inteligentes
- Reducción en el apoyo a proyectos del área “Inteligencia Artificial”.
- Cambio de nombre: [informatics](#), machine learning, analytics, [knowledge-based systems](#), [business rules management](#), [cognitive systems](#), intelligent systems, [intelligent agents](#) or [computational intelligence](#)

Siglo XXI (2000 – hoy)

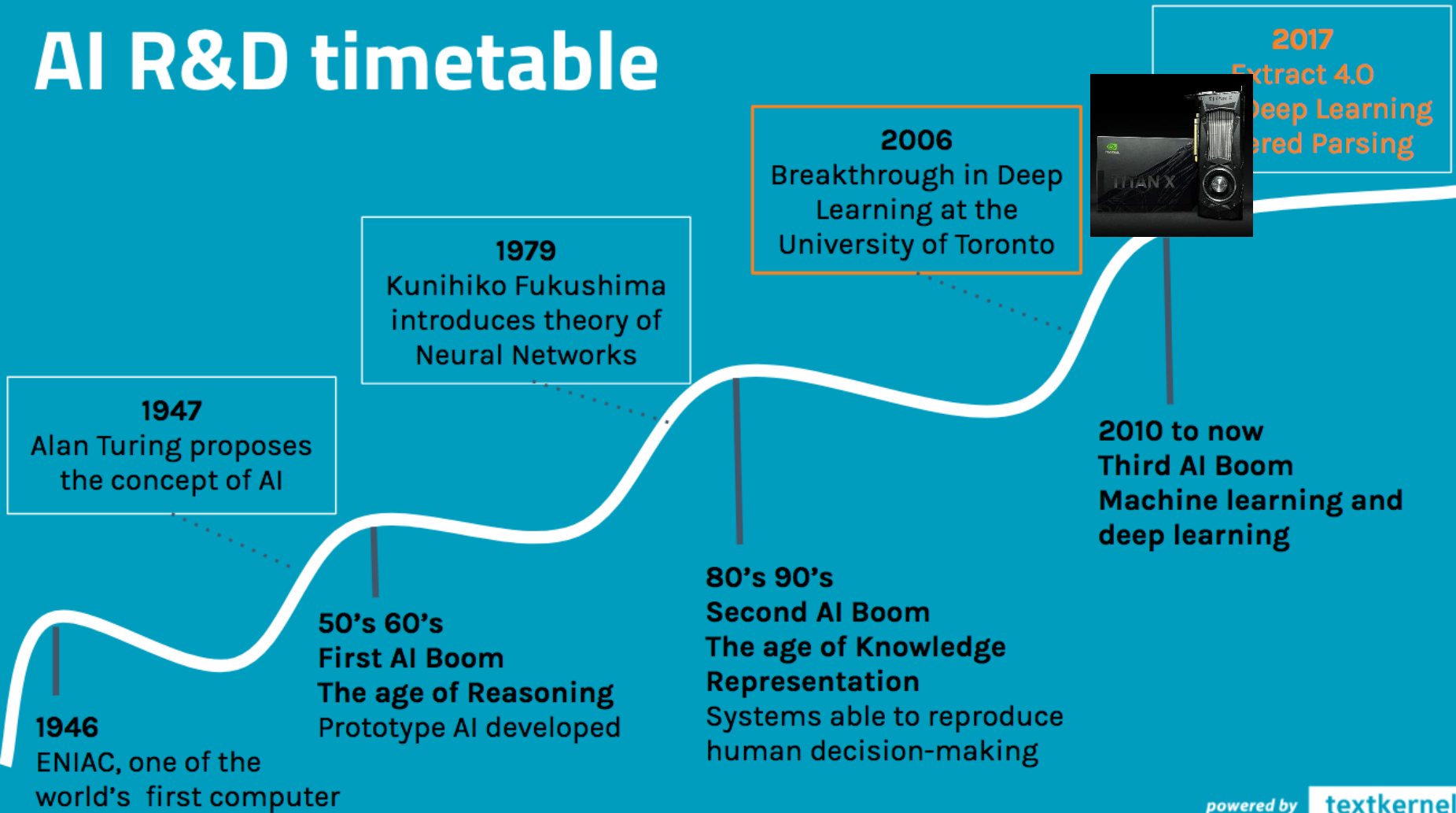
- Era de la información: redes cada vez mayores.
- Industria de cómputo (*gaming*) y de cómputo distribuido



- Inversión de grandes compañías (Nvidia, Google, Facebook, Amazon, Baidu, ...)

Estado actual de la IA

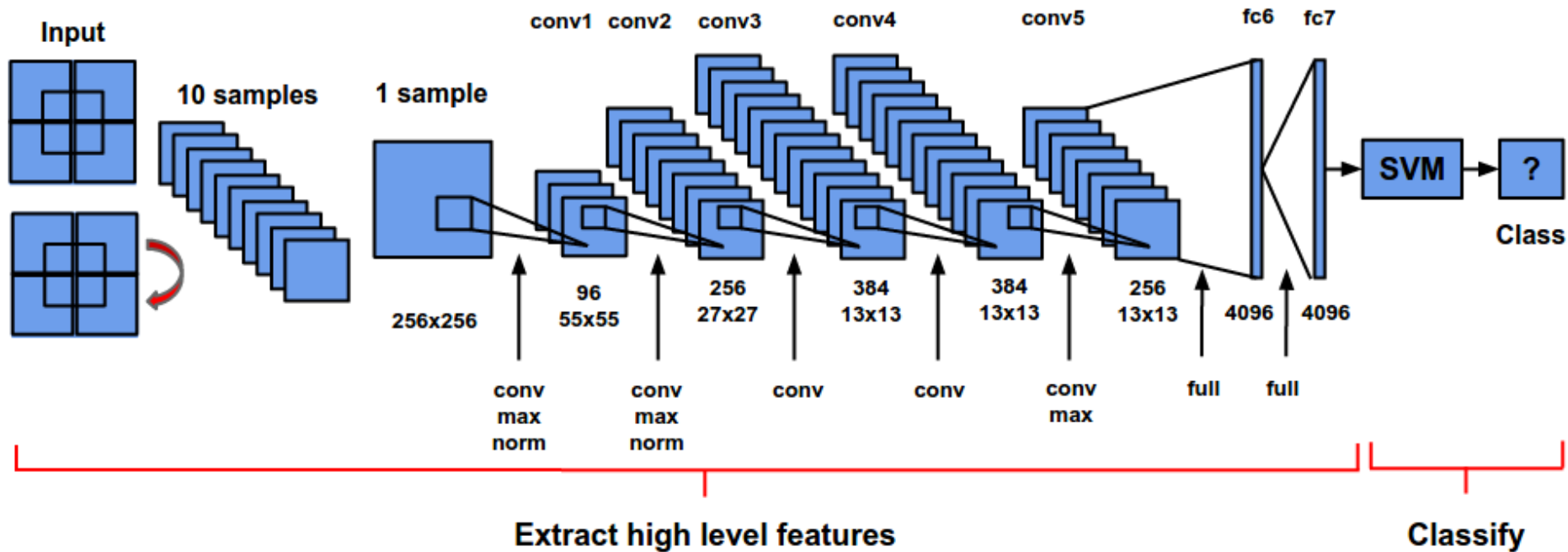
AI R&D timetable



Deep Learning (2012 -)

- ImageNET Challenge (2010)
 - clasificar 14 millones de imágenes
- AlexNet (2012)
 - breakthrough de las redes neuronales

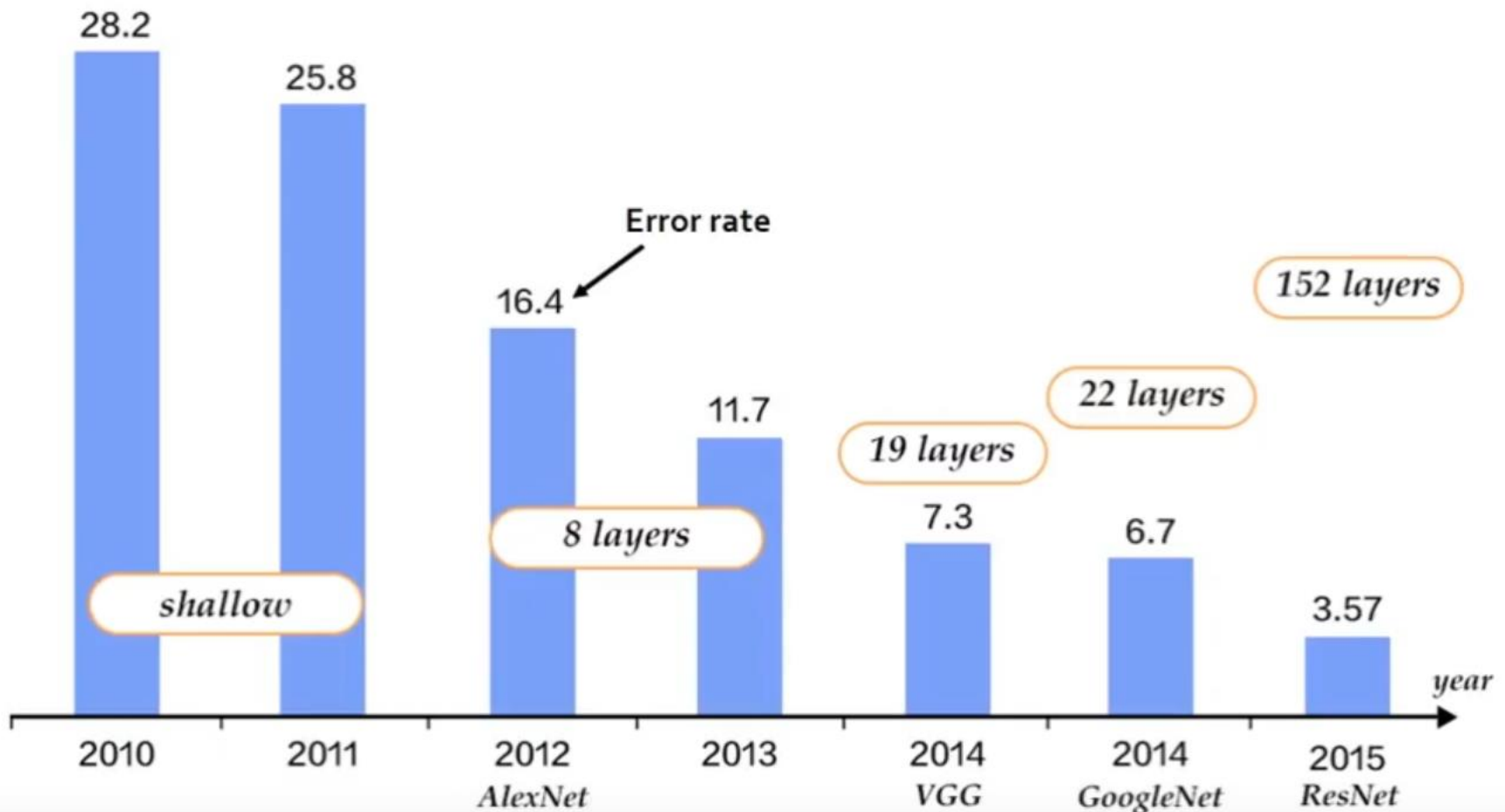
IMAGENET



Deep Learning (2012 -)

- ILSVRC (ImageNet Large Scale Visual Recognition Challenge) Winners

: After 2012, All winners of image recognition challenge used 'Deep convolutional networks'.

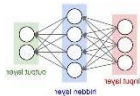


Deep Learning (2012 -)

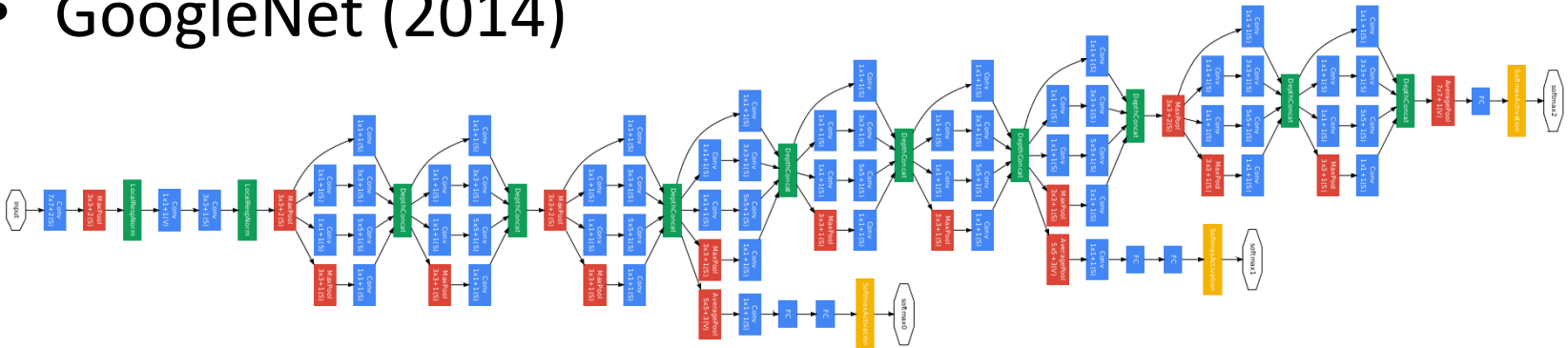
- Perceptron



- Two-layer network



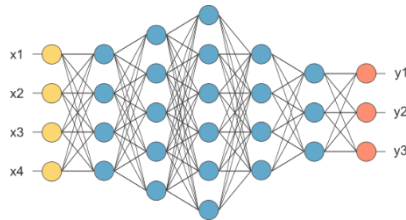
- GoogleNet (2014)



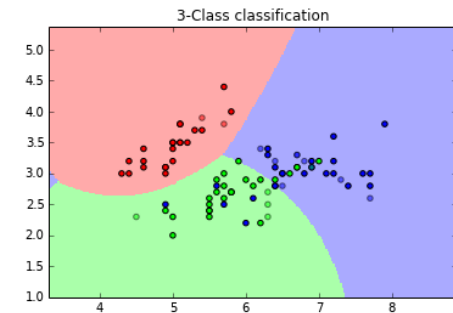
Deep Learning (2012 -)

Topologías especializadas (bloques básicos)

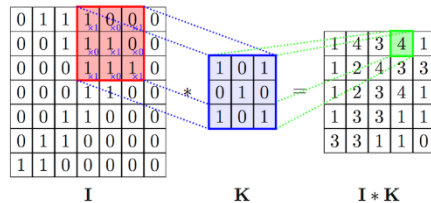
- Fully-connected



- clasificación



- Convolutional

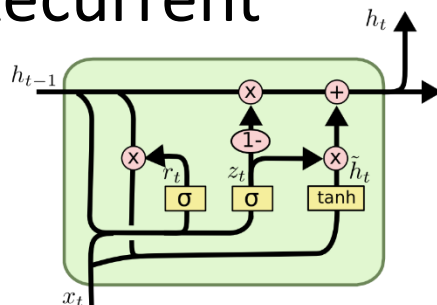


- filtrado

- visión

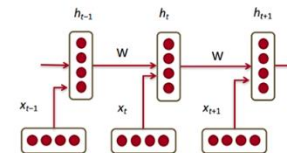


- Recurrent



- evolución

- lenguaje

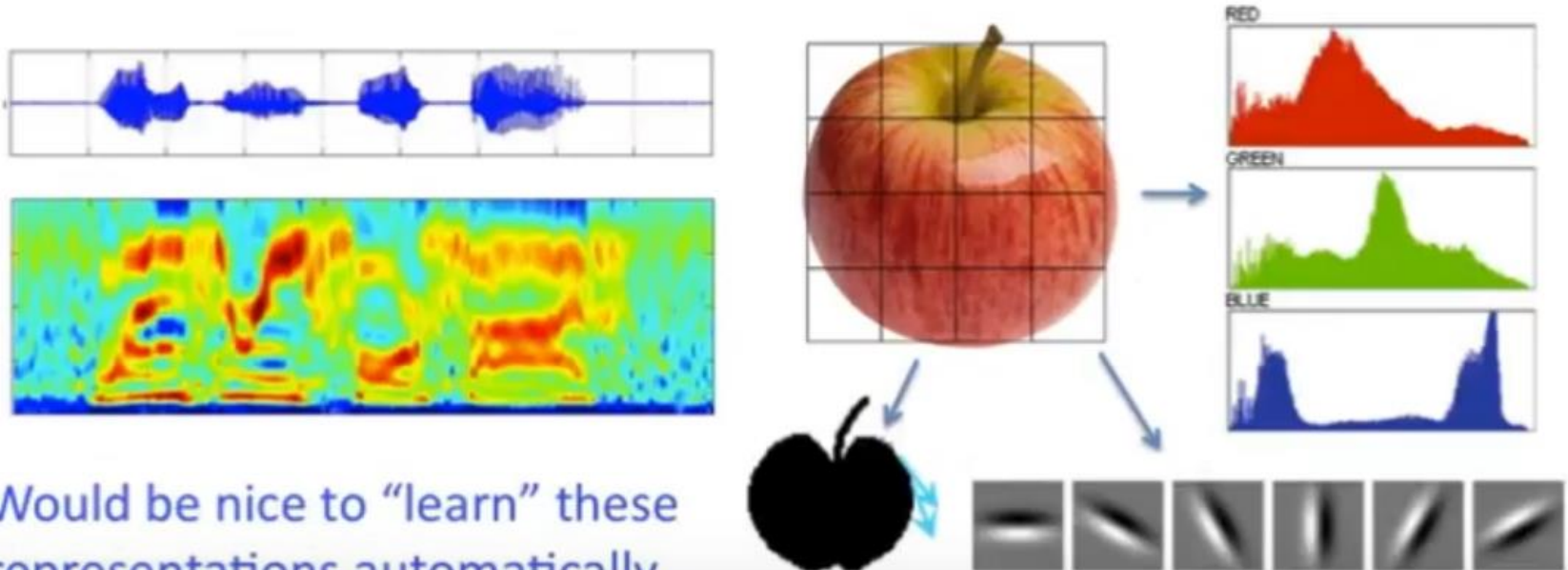


$$\begin{aligned}
 x_{\text{shirt}} - x_{\text{clothing}} &\approx x_{\text{chair}} - x_{\text{furniture}} \\
 x_{\text{king}} - x_{\text{man}} &\approx x_{\text{queen}} - x_{\text{woman}}
 \end{aligned}
 \quad
 \log p(o|c) = \log \frac{\exp(u_o^T v_c)}{\sum_{w=1}^W \exp(u_w^T v_c)}$$

Era pre – Deep Learning

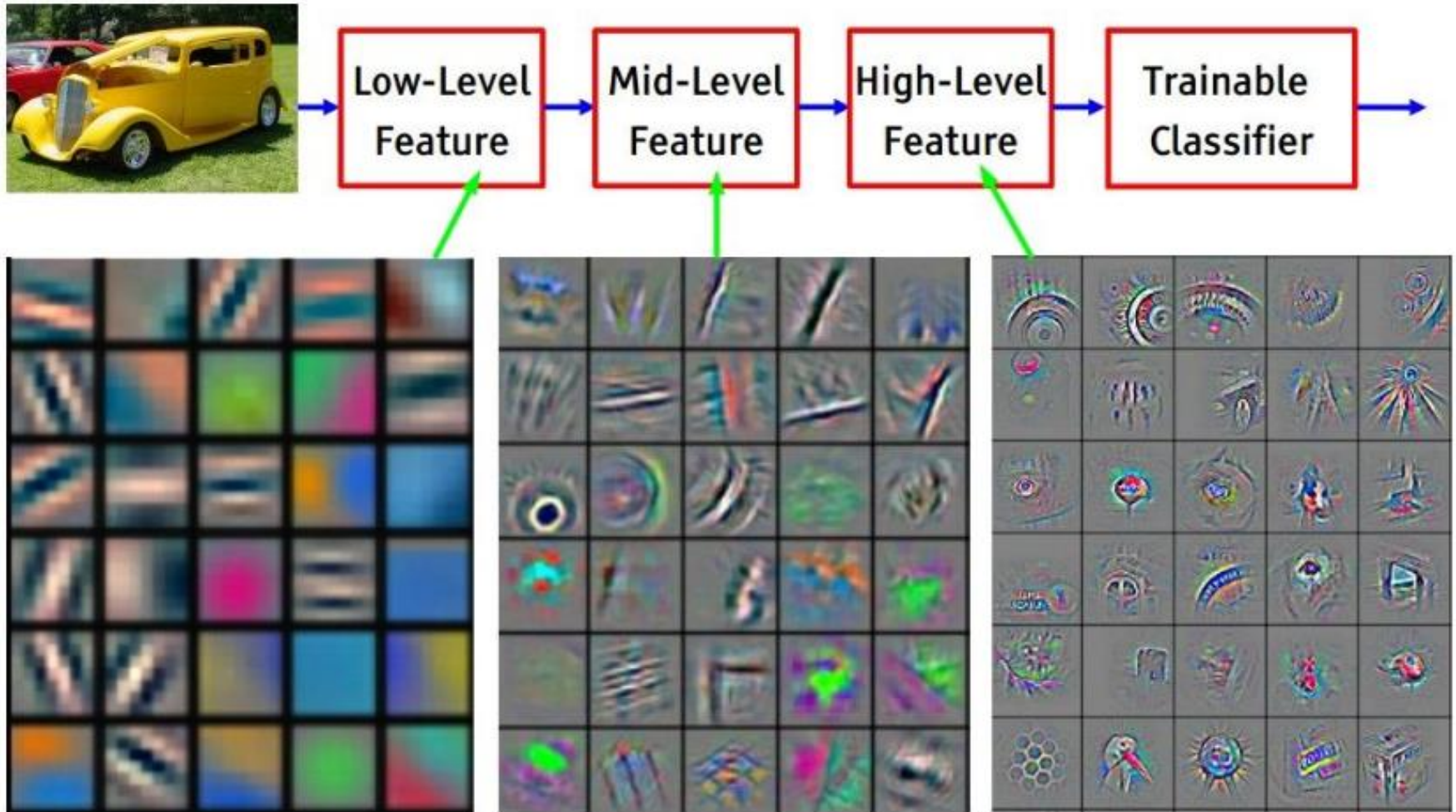
Features in Machine Learning

- 1-line summary of ML: $y = \text{sgn}(\mathbf{w}^T \mathbf{x} + b)$
 - SVM can learn very effective weights $\mathbf{w}$
 - ... if you use the right representation $\mathbf{x}$



Would be nice to “learn” these representations automatically...

Era post – Deep Learning



Aplicaciones

Aplicaciones

- Clasificación
- Predicción
- Generación

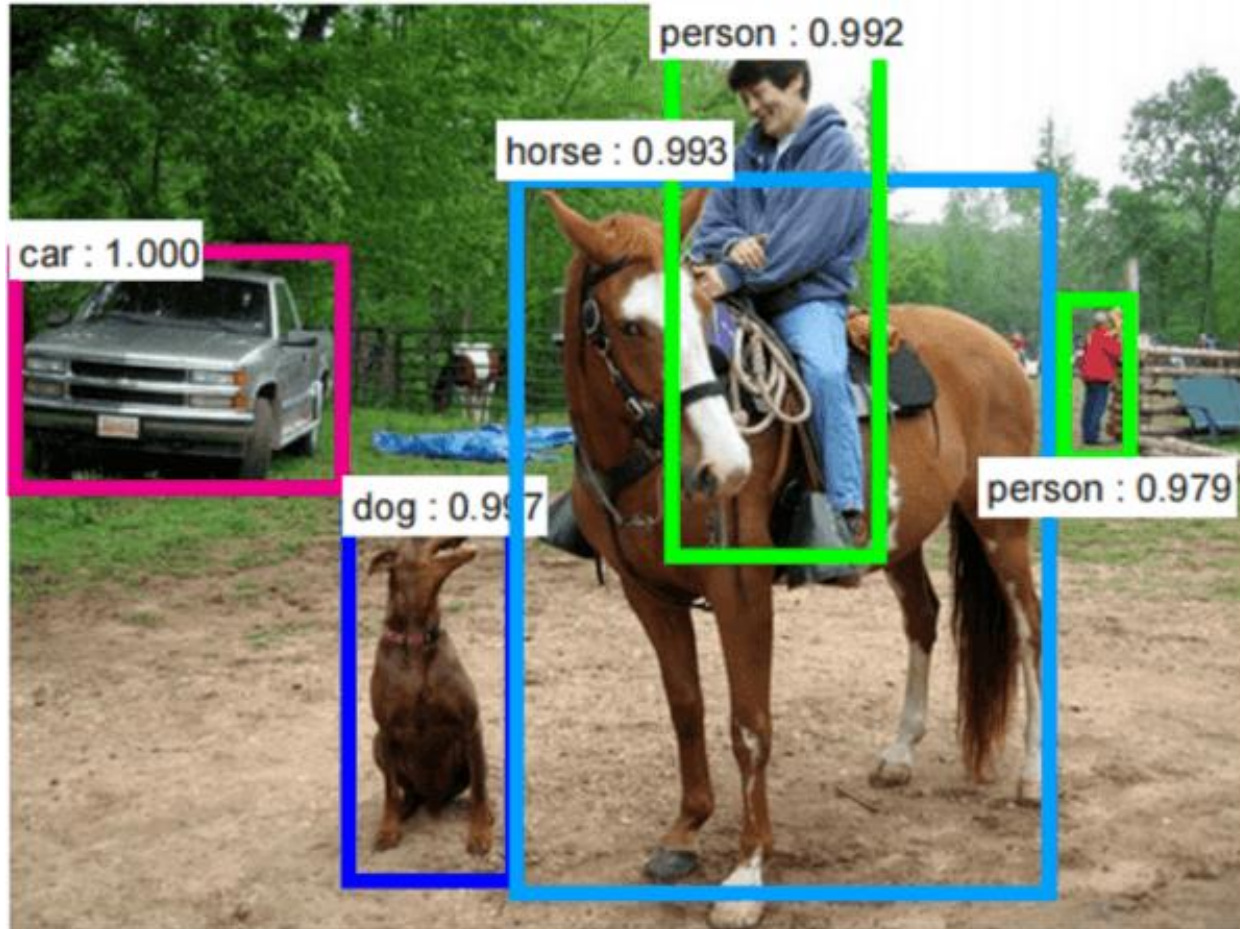
- Visión Computacional
- Procesamiento de Lenguaje
- Diagnóstico Médico
- Prevención de desastres
- Robótica

Computer Vision



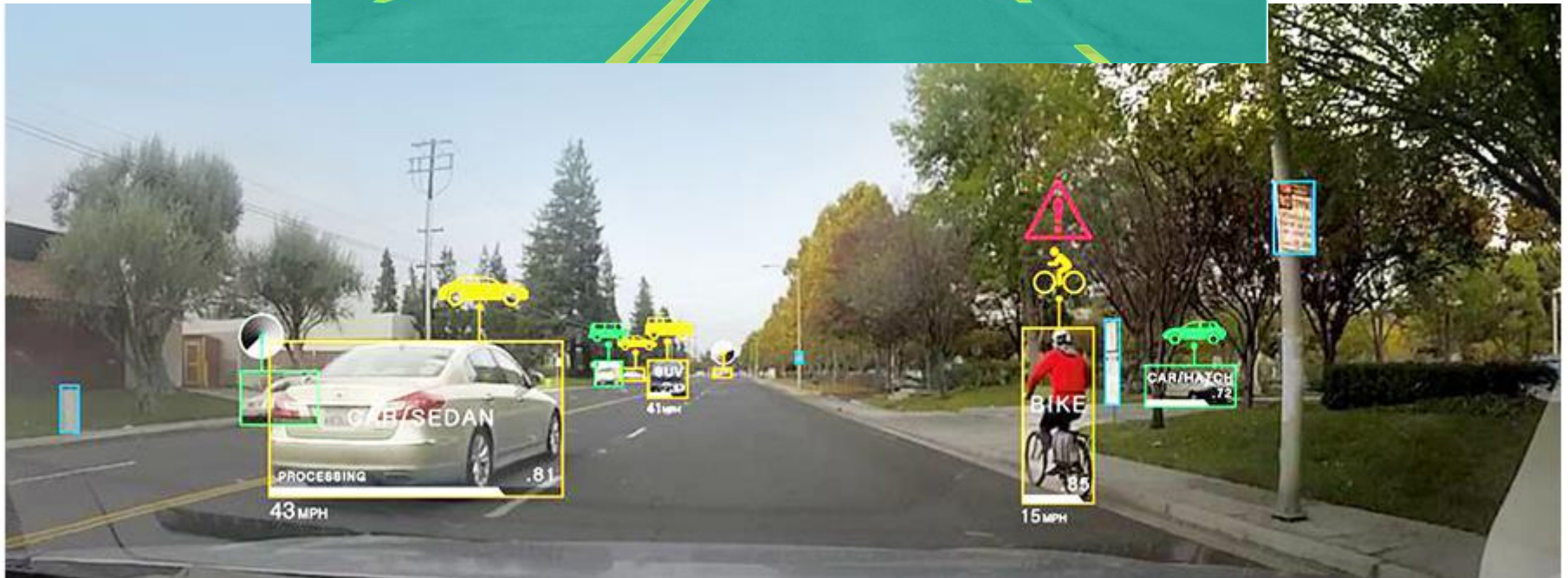
Image Segmentation

Computer Vision

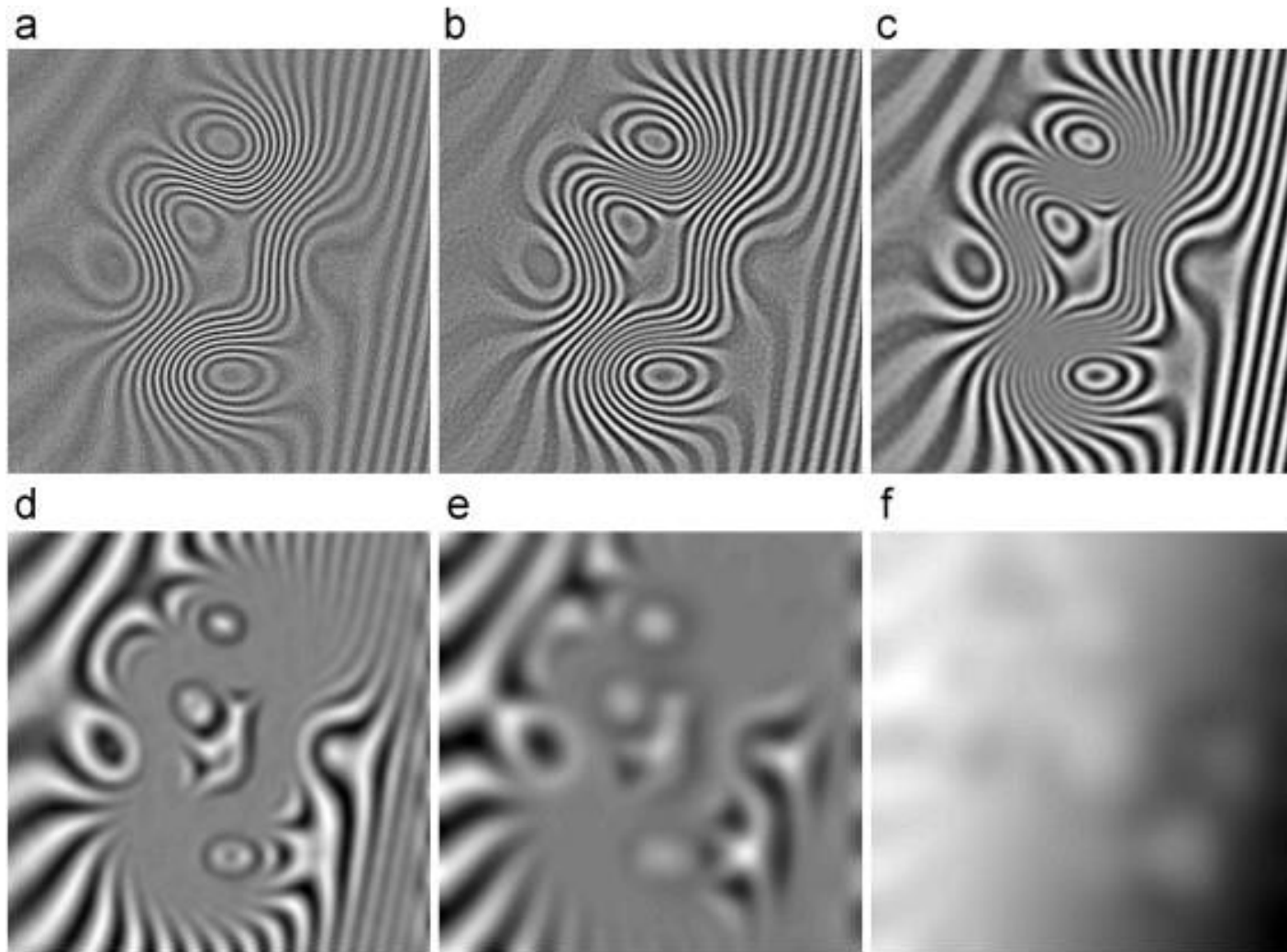


Object Detection

Self-Driving Cars



Noise Removal



Speckle noise removal in fringe patterns.

Noise Removal



Noisy input



Our result



Ground truth

Noise removal from images using Deep neural networks

Restoration



Automated colorization of images.

Restoration

The Lost World (1925)



Automated colorization of films.

Restoration



Inpainting and art restoration (Gent Altarpiece).

Super-Resolution

original



bicubic
(21.59dB/0.6423)

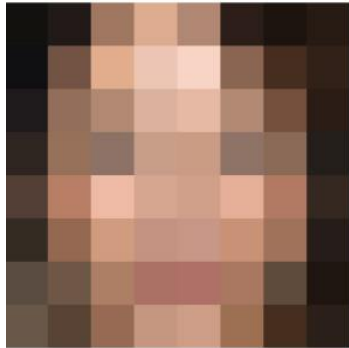


SRGAN
(20.34dB/0.6562)



Sistemas de seguridad

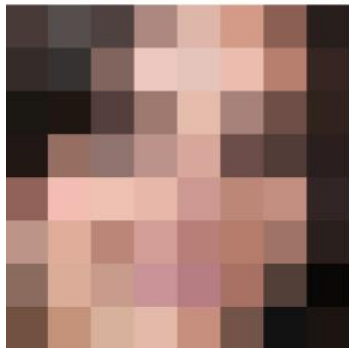
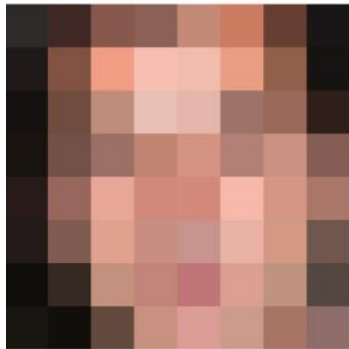
8×8 input



32×32 samples



ground truth



Healthcare

python /tf_files/label_image.py /tf_files/diagnose/leg005.jpg

surgery 0
clinical test
medications
blood pressure
lab test 52%
vaccination 82%
BMI normal

10-m

Use
Deep
Learning
to Help
Diagnose
Injury...



GraphDef version 9. Use tf.nn
brokenleg (score = 0.91144)
healthyleg (score = 0.08856)

python /tf_files/label_image.py /tf_files/diagnose/leg009.jpg

gender ♂
age 23
HR 95 bpm
120/60
ECHO D
CD PWR<500
Frq 2.0 MHz
1800 mm
AO 100%

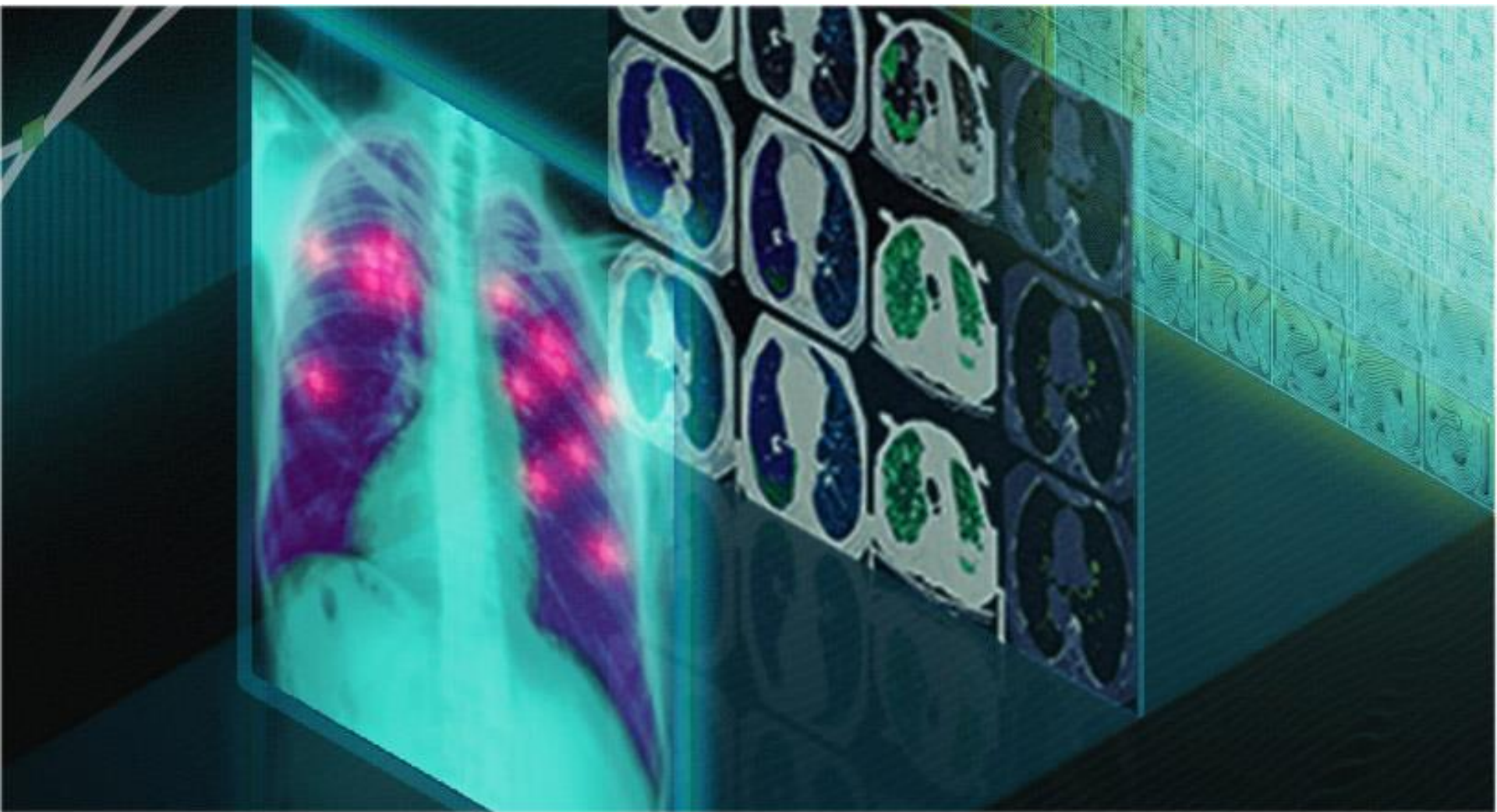


GraphDef version 9. Use tf.nn
healthyleg (score = 0.76665)
brokenleg (score = 0.23335)

+

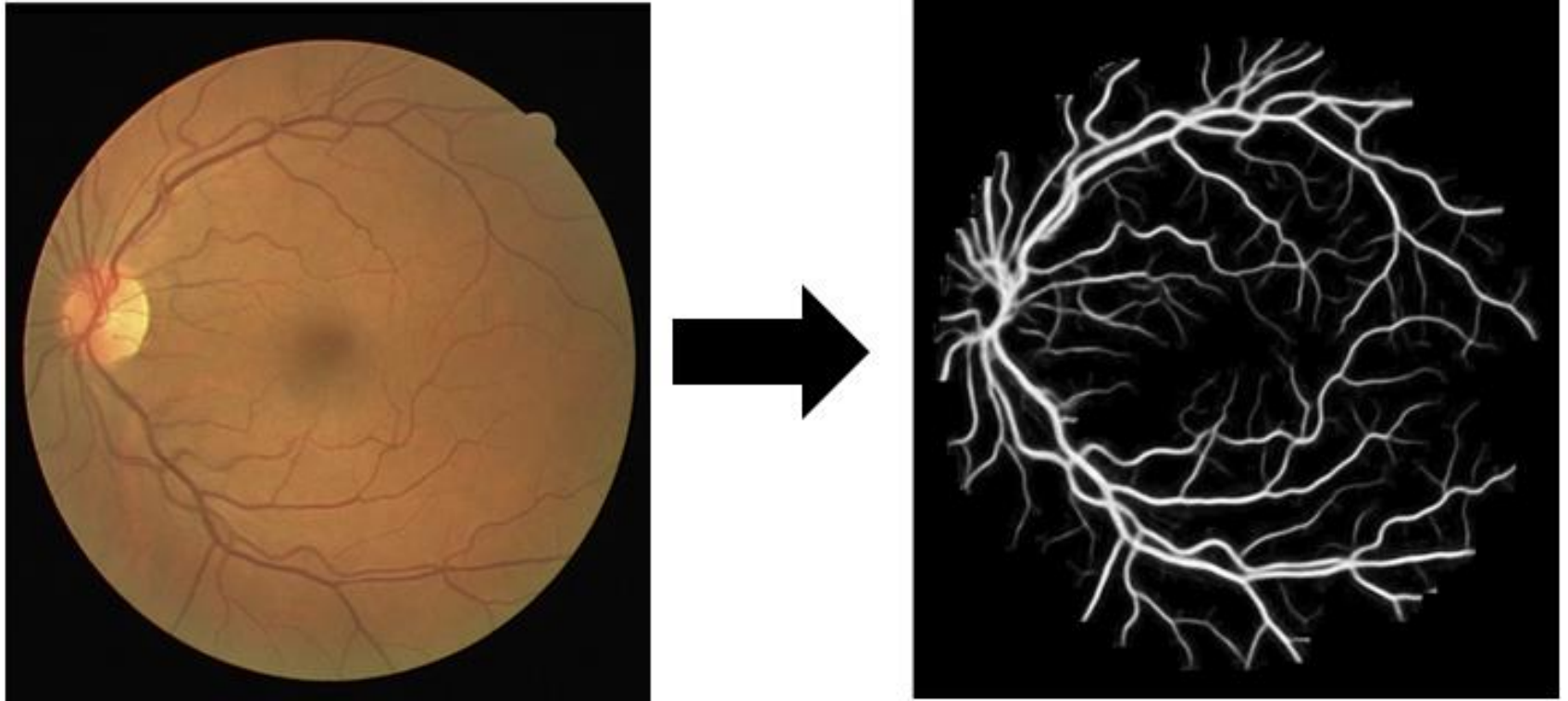
Assisted Diagnosis

Automated Diagnosis



Early detection of Health Anomalies

Automated Diagnosis

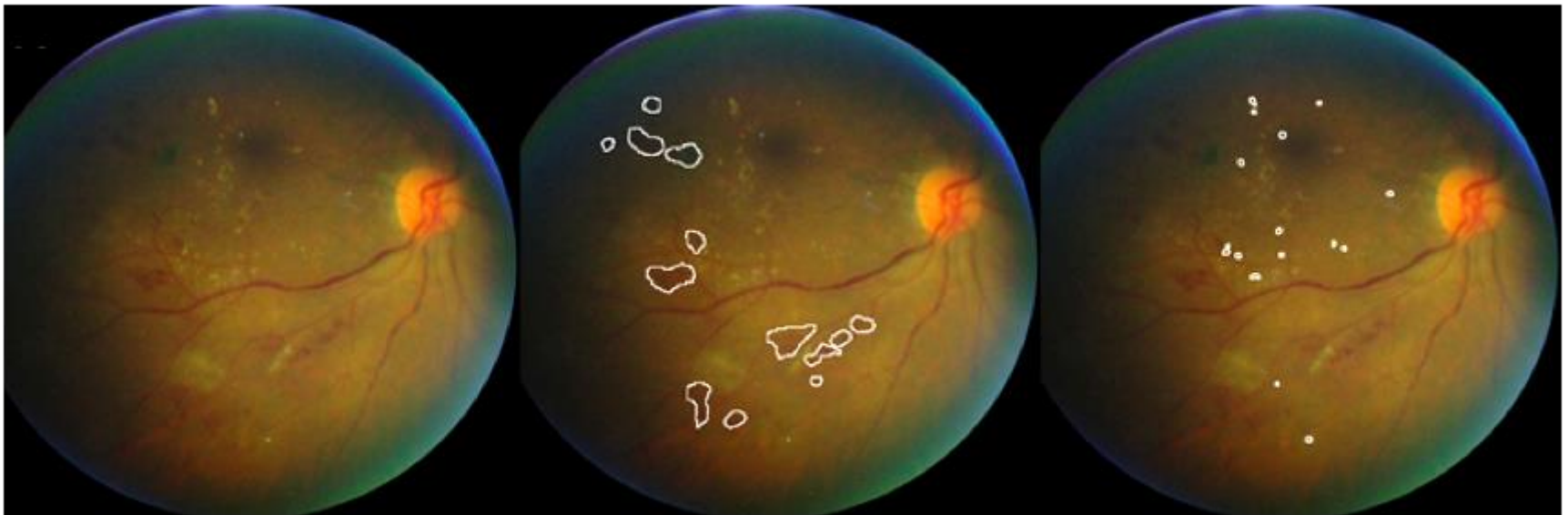


Retinal vessel segmentation using deep learning

Automated Diagnosis

Lesion localization

Robust lesion localization achieved using multi-scale pixel and lesion-level descriptors



Original Image

Red lesion detected

Bright lesions detected

Remote Sensing



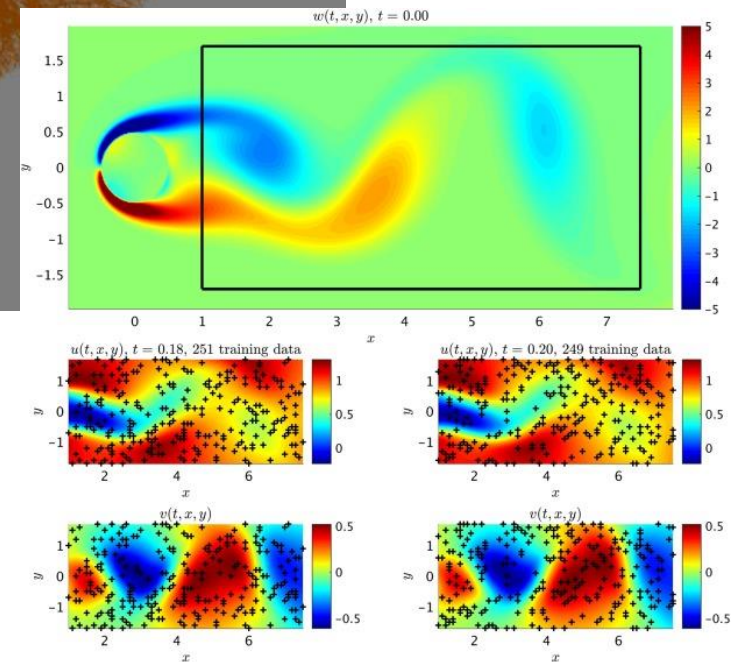
- Clasificación del uso del suelo
- Detección de área en riesgo
Incendios
Deslaves
- Prevención de desastres

Applications in Physics

TRACKML PARTICLE TRACKING CHALLENGE

Can machine learning assist high energy physics in discovering and characterizing new particles?

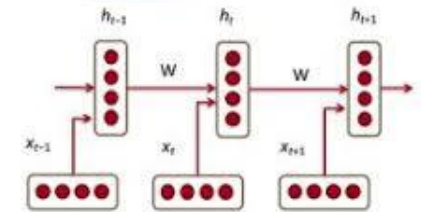
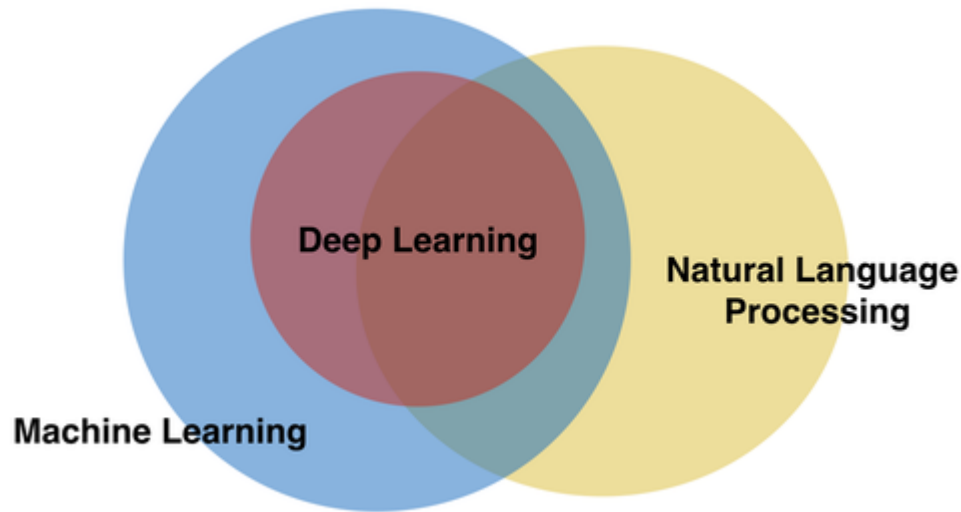
CERN Tracking Challenge



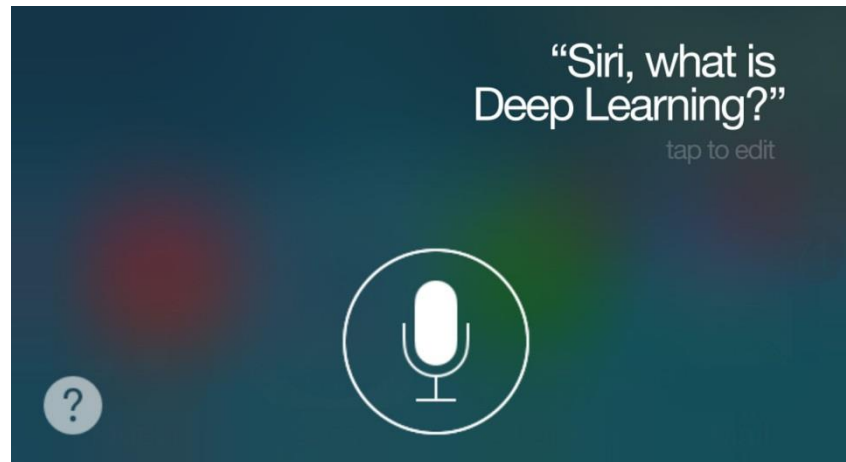
Solution of PDE's

Correct PDE	$u_t + (uu_x + vv_y) = -p_x + 0.01(u_{xx} + u_{yy})$ $v_t + (uv_x + vv_y) = -p_y + 0.01(v_{xx} + v_{yy})$
Identified PDE (clean data)	$u_t + 0.983(uu_x + vv_y) = -p_x + 0.00826(u_{xx} + u_{yy})$ $v_t + 0.983(uv_x + vv_y) = -p_y + 0.00826(v_{xx} + v_{yy})$
Identified PDE (1% noise)	$u_t + 0.849(uu_x + vv_y) = -p_x + 0.01399(u_{xx} + u_{yy})$ $v_t + 0.849(uv_x + vv_y) = -p_y + 0.01399(v_{xx} + v_{yy})$

Natural Language Processing

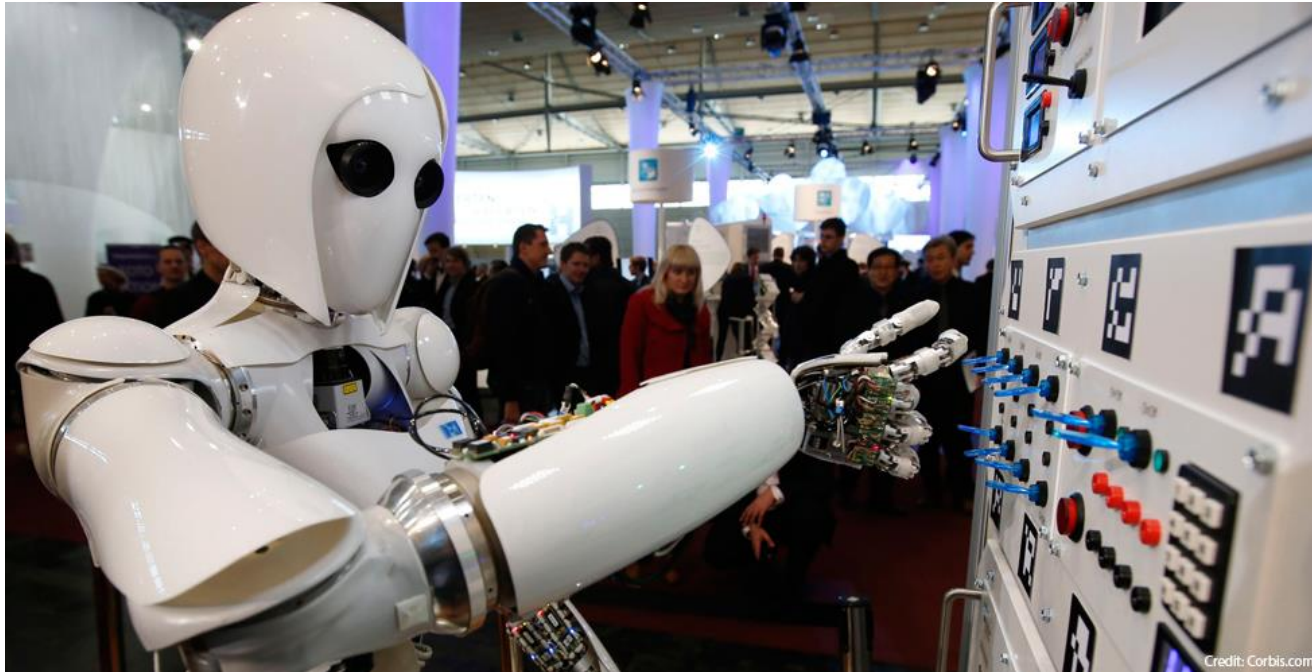


amazon alexa

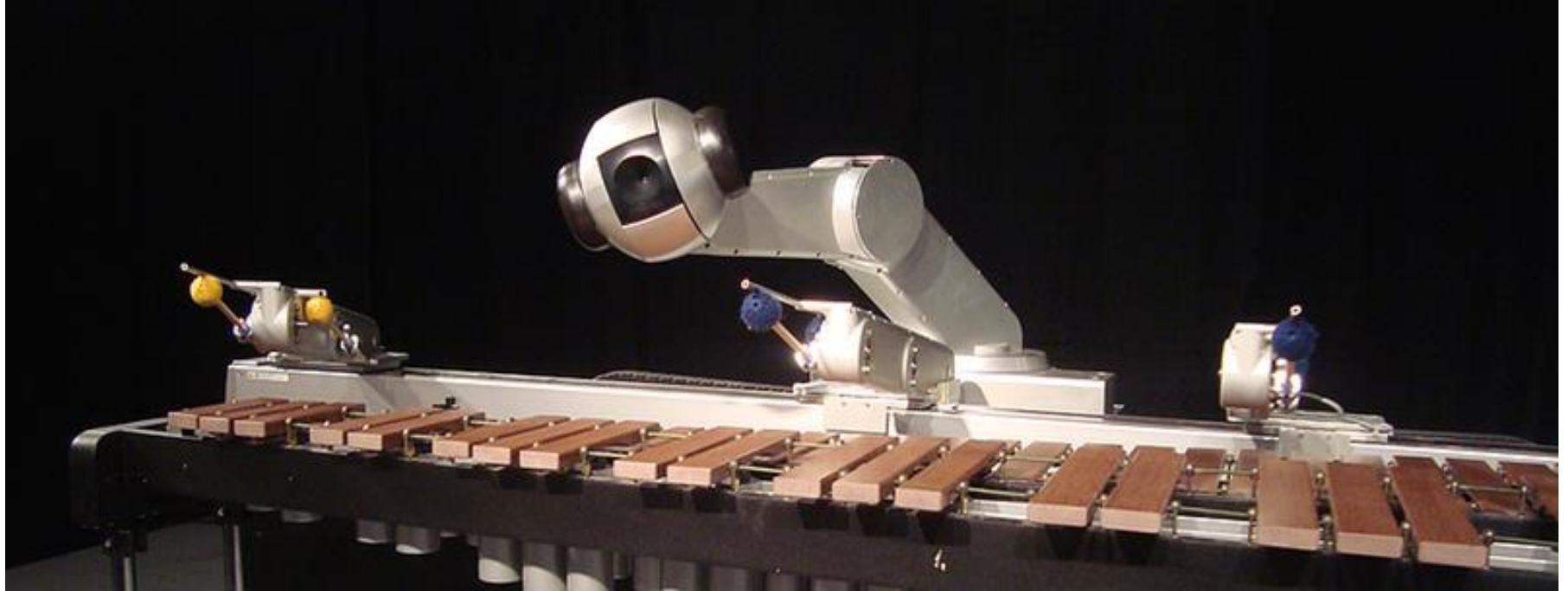


- Traducción
- Transliteración audio-texto
- Clasificación de contenidos
- Asistentes virtuales de voz

Robótica



Generación de Arte



Art style transfer



Art style transfer



Art style transfer



Automated art generation



- Generación de textos literarios, poemas, ...
- Transliteración música y partituras
- Generación de pinturas

Futuro de la IA



Gracias!