Algunas ventajas de Deep Learning

Alfonso Ruiz
Escuela de Matemáticas Bourbaki

Sommelier

- Perceptrón
- Árboles de decisión
- Regresión logística
- Programación lineal

Retail e e-commerce.

ELSEVIER

Knowledge-Based Systems
Volume 149, 1 June 2018, Pages 155-168

Applying automatic text-based detection of deceptive language to police reports: Extracting behavioral patterns from a multi-step classification model to understand how we lie to the police

Lara Quijano-Sánchez [a], Federico Liberatore [a, b], José Camacho-Collados [c], Miguel Camacho-Collados [d]

Clasificación de denuncias falsas.

$$d = (x_1, \ldots, x_{88000})$$

$$0 \leq x_1\theta_1 + ... + x_{88000}\theta_{88000} \leq 1$$

$$d = (x_1, \ldots, x_{88000})$$

$$0 \leq x_1\theta_1 + ... + x_{88000}\theta_{88000} \leq 1$$

$$\theta = (\theta_1, ..., \theta_{88000}) = ?$$

$$(d_1, y_1), \ldots, (d_N, y_N)$$

$$\underset{\theta}{argmin}\left(\frac{1}{N}\Sigma\left(d_i\theta_i - y_i\right)^2\right)$$

Problemas:

- Estructura Gramatical
- Antónimos, sinónimos
- Énfasis

Redes neuronales

Nuevos parámetros: mejor organización

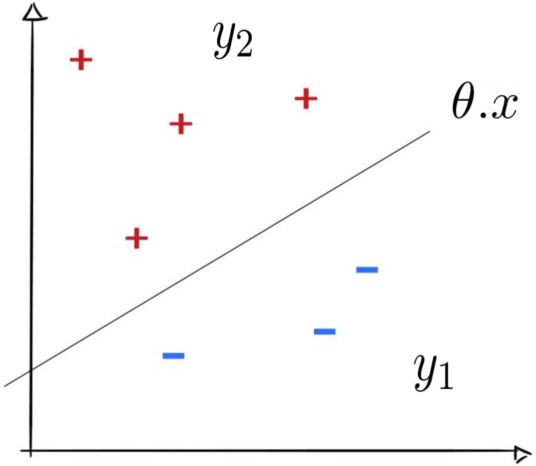
$$\underbrace{\text{hubo}}_{x_1}\ \underbrace{\text{violencia}}_{x_2}$$

$$\alpha = x_1.\theta_1 + x_2.\theta_2$$

$$\gamma = \mathbf{1}_{[x<\frac{1}{2}]}(\alpha)$$

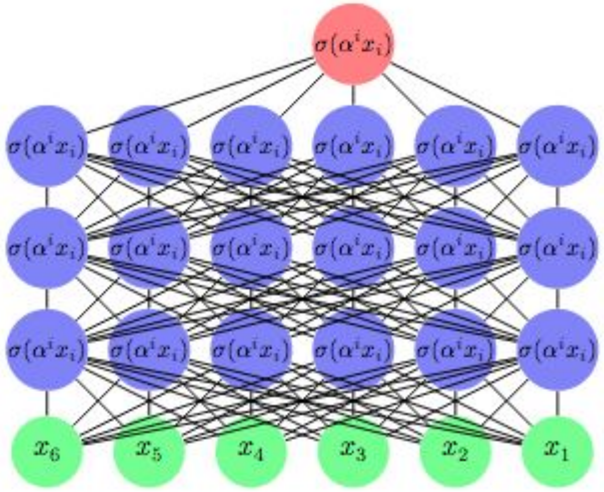
$$\underbrace{\text{no}}_{x_1}\ \underbrace{,}_{x_2}\ \underbrace{\text{hubo}}_{x_3}\ \underbrace{\text{violencia}}_{x_4}$$

Funciones no lineales



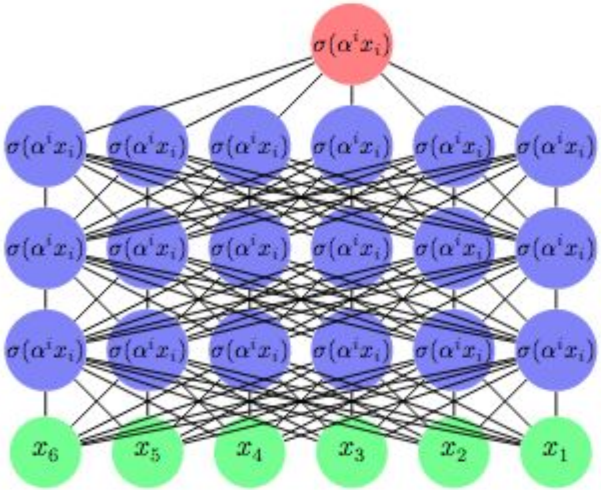
$$y_i = \text{sign}(\theta.x)$$

- Regresión logística
- RELU
- Sign
- Sigmoid
- tanh



Definición

1. Arquitectura
2. Funciones de activación
3. Parámetros



$$(V, E, \theta, \sigma)$$

Estudio Matemático de las redes Neuronales

1. Aproximación
2. Aprendizaje
3. Generalización

Aproximación

$$f(x_i) = y_i$$

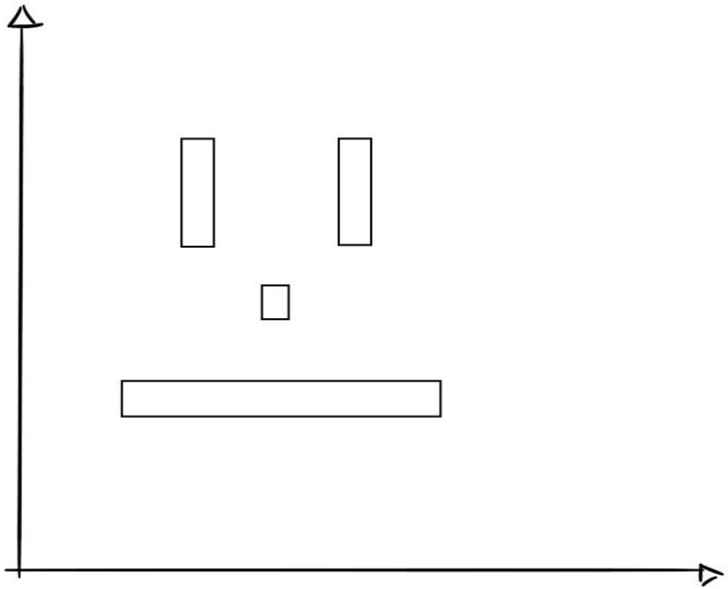
Funciones

1. Clasificación de textos, imágenes o audio.
2. Gasto de energía de acuerdo a la temperatura.

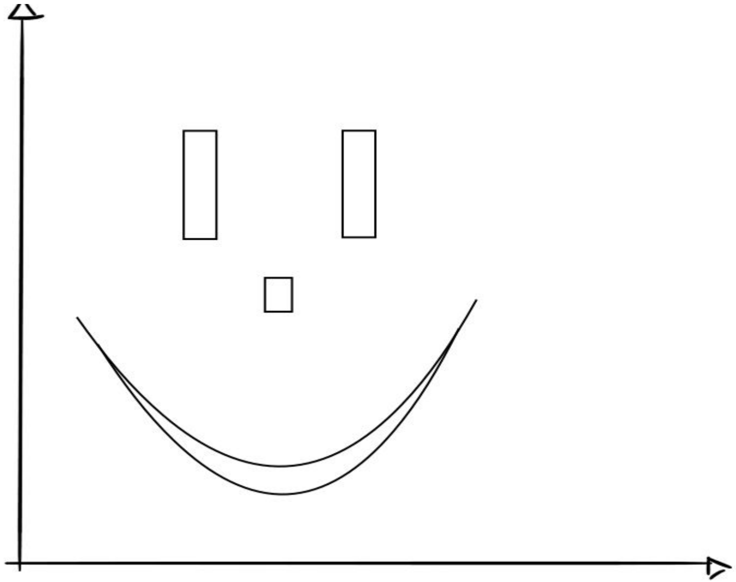
$$\{0,1\}^d \to \{0,1\}$$

Discretización del problema

1. Bits
2. Pixeles
3. Intervalos de tiempo



Primer teorema de aproximación.



Caso continuo

Crecimiento exponencial

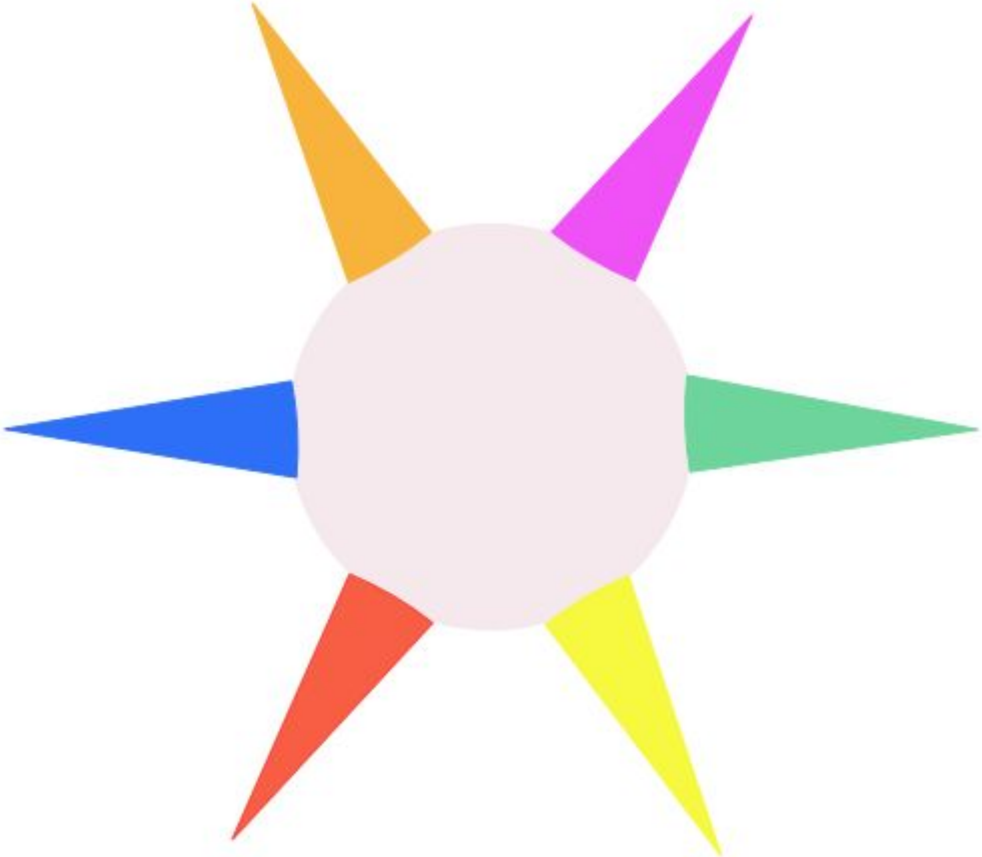
x	2^x	x^2
1	2	1
2	4	4
3	8	9
10	1024	100
100	1.26765E+30	10000
10000000000	#NUM!	1E+20

$$|V| = e^d$$

Tamaño computacional: vacuidad

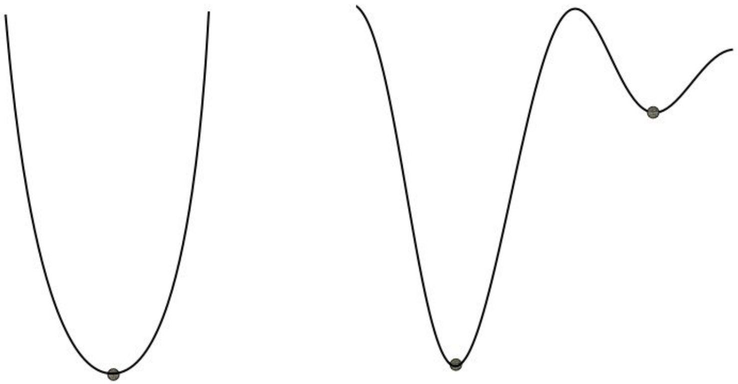
Aprendizaje (cálculo de los parámetros)

La maldición de la dimensión



$$\underset{\theta}{argmin}\left(\sigma(\theta \cdot x)\right)$$

Método del gradiente



Método del gradiente, saddle points

Backpropagation: gran eficacia estadística y computacional.

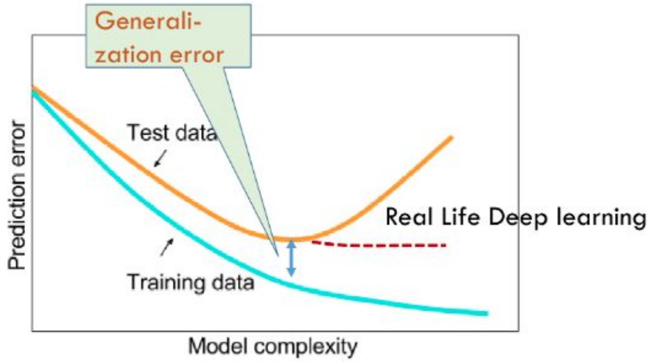
Generalización: exceso de parámetros

Capacidad:

Teorema de aproximación v.s. Overfitting

$$Q = mc\theta$$

m = mass		(kg)
c = specific heat capacity		($J\ kg^{-1}\ {}^{o}C^{-1}$)
θ = temperature change		(${}^{o}C$)



¿Overfitting?

UNDERSTANDING DEEP LEARNING REQUIRES RE-THINKING GENERALIZATION

Chiyuan Zhang[*]
Massachusetts Institute of Technology
chiyuan@mit.edu

Samy Bengio
Google Brain
bengio@google.com

Moritz Hardt
Google Brain
mrtz@google.com

Benjamin Recht[†]
University of California, Berkeley
brecht@berkeley.edu

Oriol Vinyals
Google DeepMind
vinyals@google.com

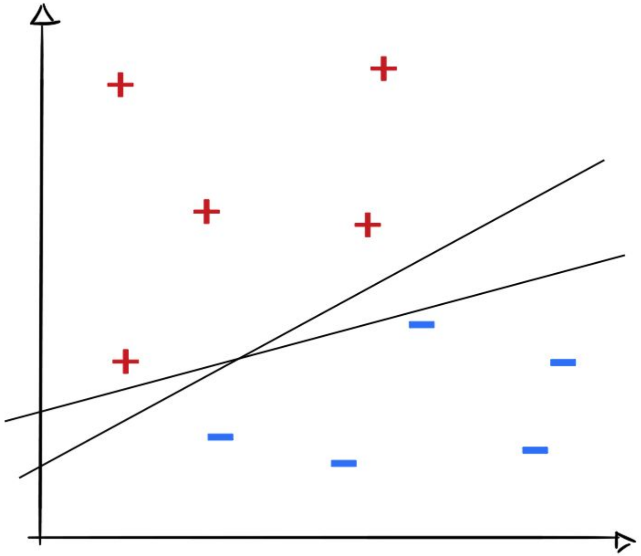
ABSTRACT

Despite their massive size, successful deep artificial neural networks can exhibit a remarkably small difference between training and test performance. Conventional wisdom attributes small generalization error either to properties of the model family, or to the regularization techniques used during training.

Through extensive systematic experiments, we show how these traditional approaches fail to explain why large neural networks generalize well in practice. Specifically, our experiments establish that state-of-the-art convolutional networks for image classification trained with stochastic gradient methods easily fit a random labeling of the training data. This phenomenon is qualitatively unaffected by explicit regularization, and occurs even if we replace the true images by completely unstructured random noise. We corroborate these experimental findings with a theoretical construction showing that simple depth two neural networks already have perfect finite sample expressivity as soon as the number of parameters exceeds the number of data points as it usually does in practice.

We interpret our experimental findings by comparison with traditional models.

Malentendido



Support Vector Machines

Teoremas de Generalización

Error de práctica - error de entrenamiento $\leq \sqrt{m/n}$

m= dimensión del espacio de funciones

n=cantidad de ejemplos

Stronger generalization bounds for deep nets via a compression approach

Sanjeev Arora* Rong Ge† Behnam Neyshabur‡ Yi Zhang§

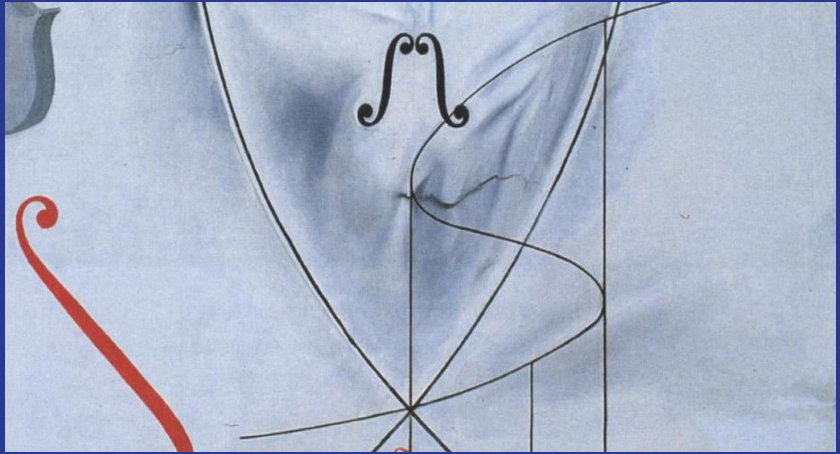
Abstract

Deep nets generalize well despite having more parameters than the number of training samples. Recent works try to give an explanation using PAC-Bayes and Margin-based analyses, but do not as yet result in sample complexity bounds better than naive parameter counting. The current paper shows generalization bounds that're orders of magnitude better in practice. These rely upon new succinct reparametrizations of the trained net — a compression that is explicit and efficient. These yield generalization bounds via a simple compression-based framework introduced here. Our results also provide some theoretical justification for widespread empirical success in compressing deep nets.

Analysis of correctness of our compression relies upon some newly identified "noise stability"properties of trained deep nets, which are also experimentally verified. The study of these properties and resulting generalization bounds are also extended to convolutional nets, which had eluded earlier attempts on proving generalization.

Escuela de Matemáticas Bourbaki.

1. Cursos a la medida.
2. Cursos individuales y en grupos pequeños.
3. Diversos niveles.



Sitio Web

alfonso@escuela-bourbaki.com

LinkedIn

alfonso@escuela-bourbaki.com

Meetup

alfonso@escuela-bourbaki.com