



Redes Neuronales para Lenguaje Natural

2024

Grupo de Procesamiento de Lenguaje Natural
Instituto de Computación

Docentes

- Luis Chiruzzo
- Ignacio Sastre

Curso del Grupo de Procesamiento de Lenguaje Natural (PLN) del InCo.

Horarios

Salón 725 - Beige

- martes 16:30 a 18:00
- jueves 16:30 a 18:00

Modalidad del curso

Clases teórico-prácticas

- Teórico expositivo
- Ejercicios de programación y discusión

Entregas obligatorias a lo largo del curso

- Tienen que ir formando grupos de 3 integrantes

Una prueba final escrita obligatoria

- A coordinar con los segundos parciales

Programa

- Introducción a Redes Neuronales
- Representaciones distribuidas de palabras
- Arquitecturas clásicas de aprendizaje profundo
- Modelo atencional
- Modelo transformer
- Grandes modelos de lenguaje y modelos generativos

Bibliografía

- Speech and Language Processing (Third ed.). Jurafsky, D. and J. H. Martin (2024). <https://web.stanford.edu/~jurafsky/slp3/>
- Deep Learning Fundamentals - Sebastian Raschka.
<https://lightning.ai/courses/deep-learning-fundamentals/>
- Deep Learning - Ian Goodfellow, Yoshua Bengio, Aaron Courville - MIT Press (2016)

Software



Tensorflow - <https://www.tensorflow.org/>

Keras - <http://keras.io/>



PyTorch - <https://pytorch.org/>

Huggingface - <https://huggingface.co/>





Ejemplo

Cuéntame la historia de las redes neuronales desde su inicio hasta hoy.



La historia de las redes neuronales es fascinante y abarca varias décadas de investigación y avances tecnológicos. Aquí tienes un resumen de los momentos clave en el desarrollo de las redes neuronales desde su inicio hasta el presente:

1. **Década de 1940-1950: Neurona Artificial**

- El concepto de una "neurona artificial" fue introducido por Warren McCulloch y Walter Pitts en 1943. Propusieron una simplificación matemática de las neuronas biológicas y cómo podrían interactuar para realizar cálculos lógicos.

2. **Década de 1950-1960: Perceptrón**

Ahora escríbelo como si fuera una payada.



En la historia de las redes neuronales me voy a aventurar,
Una payada les traigo, para todos narrar.
Comenzó en los cuarenta, con McCulloch y Pitts,
Una neurona artificial, ¡vaya ingenio sin fin!

Llegó el cincuenta, y el perceptrón apareció,
Rosenblatt lo creó, en el campo se impuso.
Clasificando patrones, brillaba con pasión,
Pero estancado quedó, retos en la misión.

Los sesenta y setenta trajeron su pesar,
Minsky y Papert, cuestionaron sin cesar.

¿Quién fue el primer premio Nobel uruguayo?

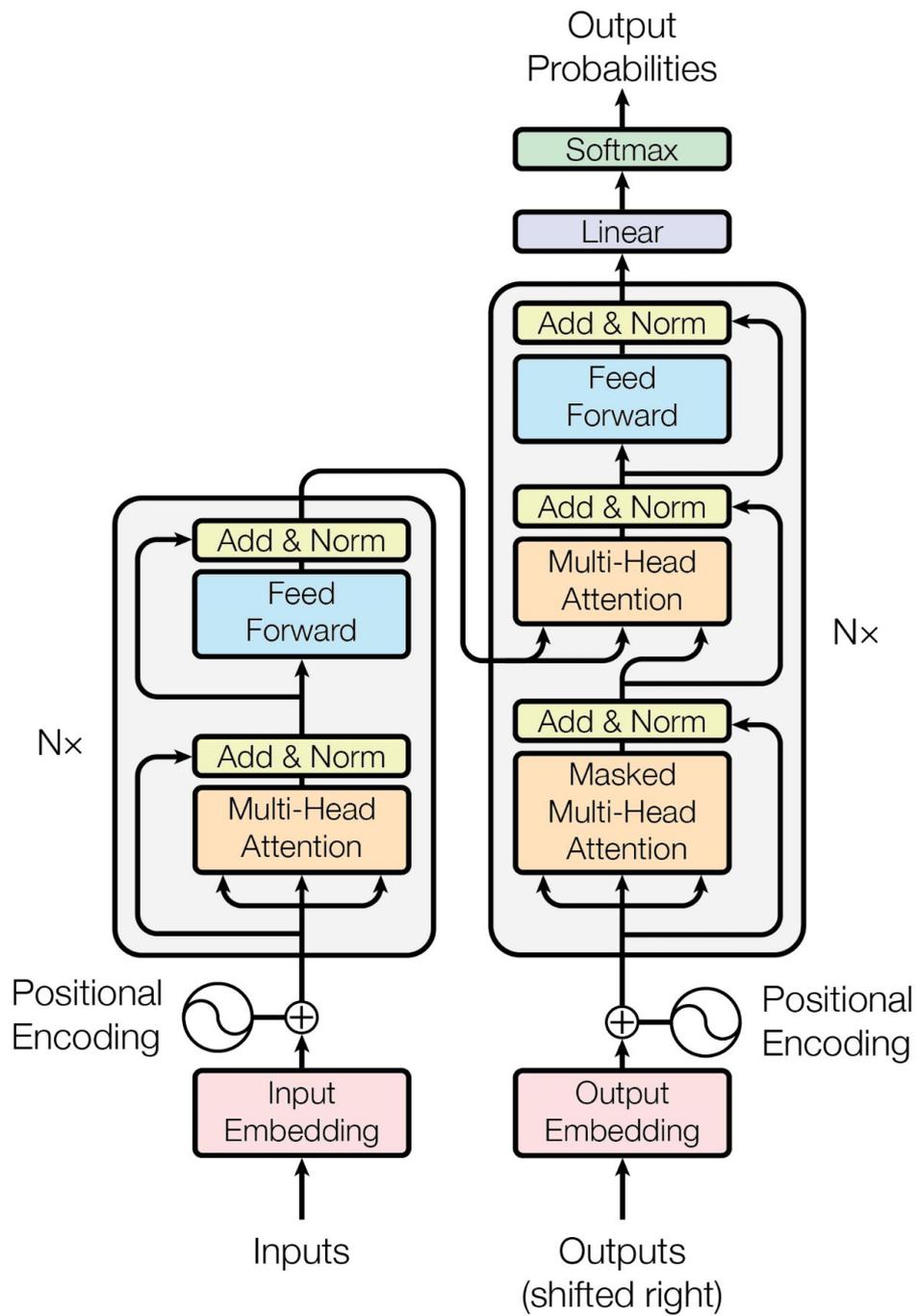


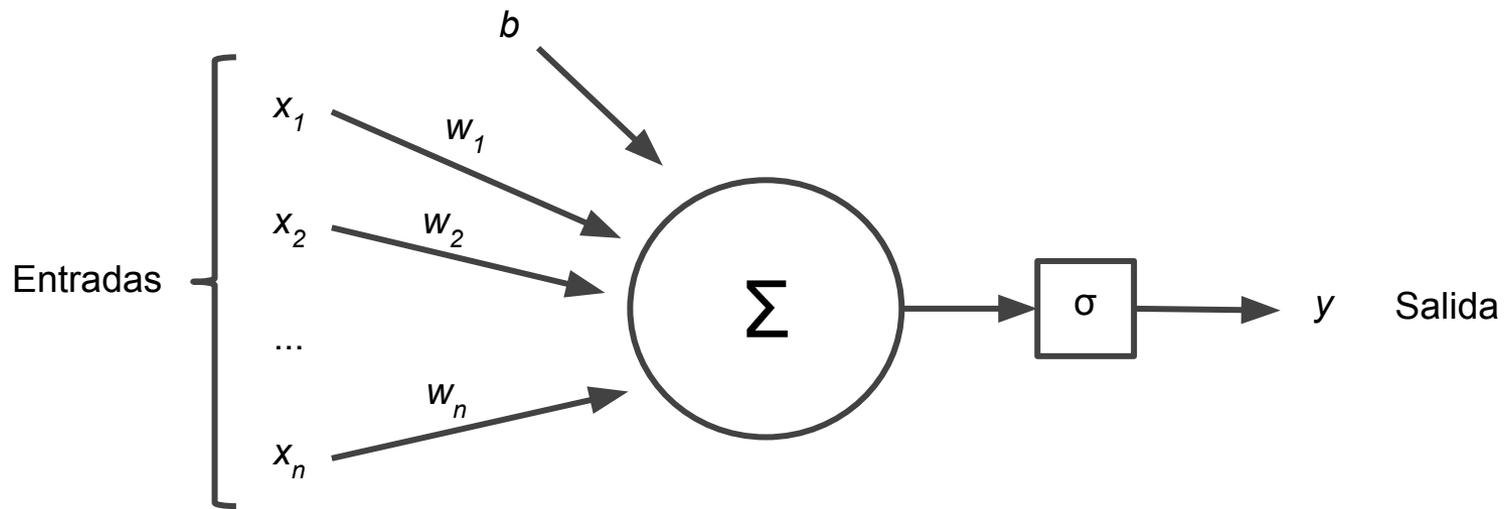
El primer uruguayo en ganar un Premio Nobel fue José Enrique Rodó, un escritor y ensayista uruguayo que recibió el Premio Nobel de Literatura en 1917.

¿Cuántos premios Nobel ganó José Enrique Rodó?



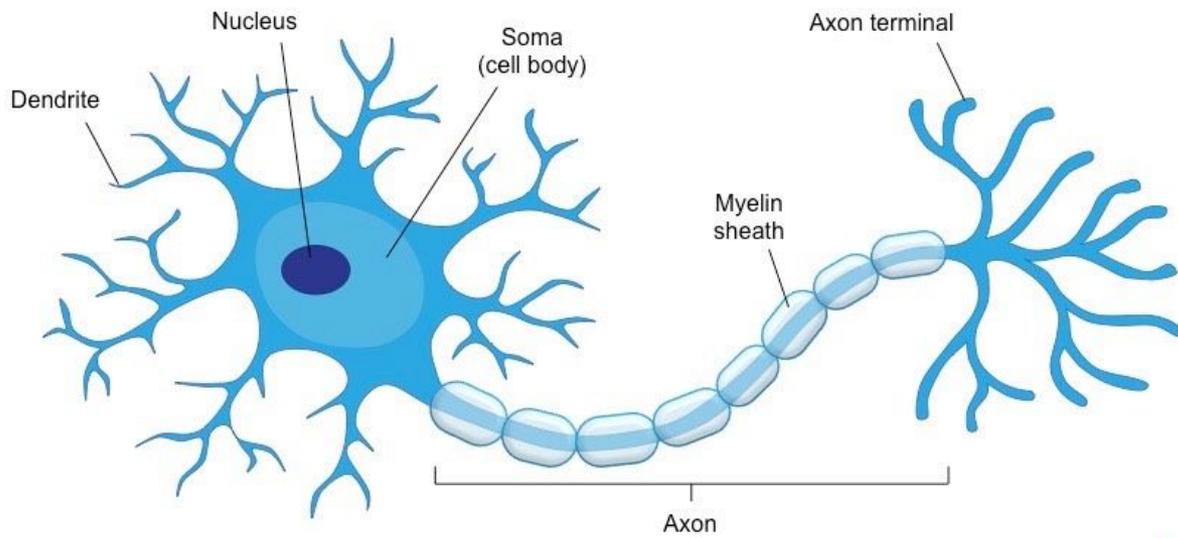
José Enrique Rodó no ganó el Premio Nobel de Literatura en realidad. Mis disculpas por la información incorrecta anterior. El primer uruguayo en ganar un Premio Nobel fue José Serebrier, un director de orquesta y compositor nacido en Uruguay, quien ganó el Premio Nobel de Literatura en 2004.

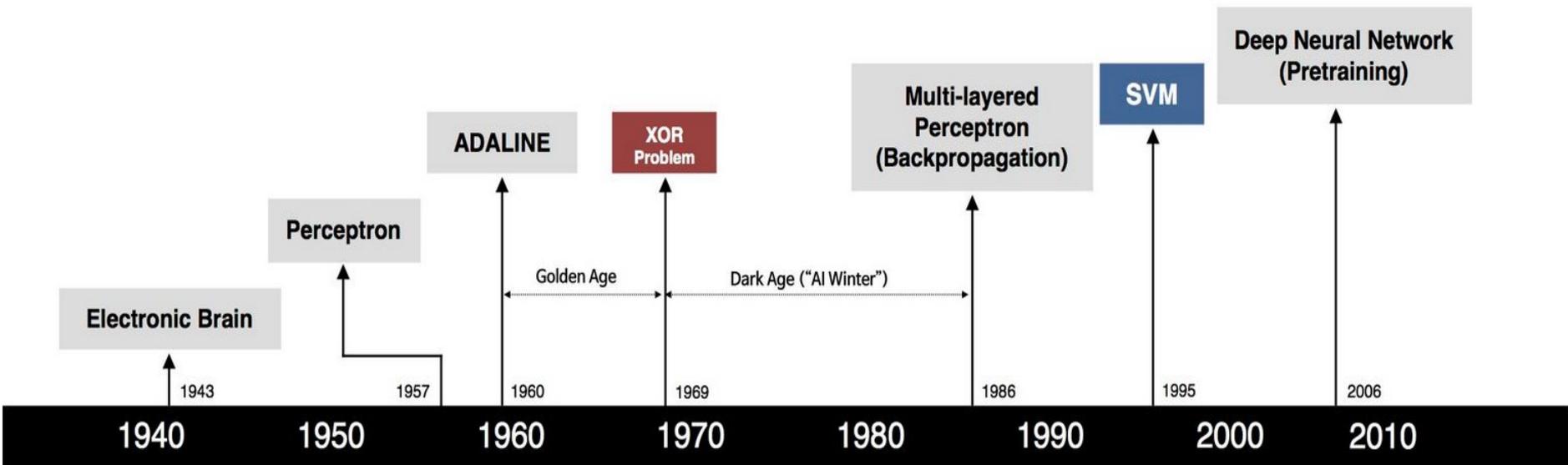




$$y = \sigma\left(\sum_i x_i w_i + b\right)$$

Neurona de
McCulloch-Pitts,
1943





S. McCulloch - W. Pitts



F. Rosenblatt



B. Widrow - M. Hoff



M. Minsky - S. Papert



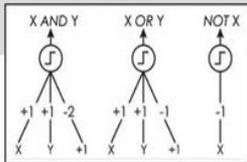
D. Rumelhart - G. Hinton - R. Williams



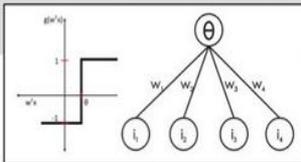
V. Vapnik - C. Cortes



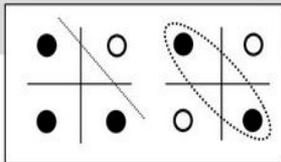
G. Hinton - S. Ruslan



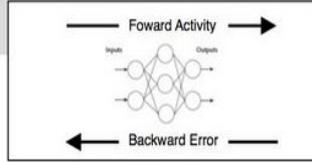
- Adjustable Weights
- Weights are not Learned



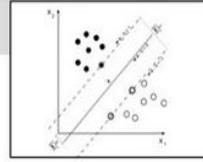
- Learnable Weights and Threshold



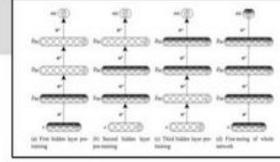
- XOR Problem



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



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

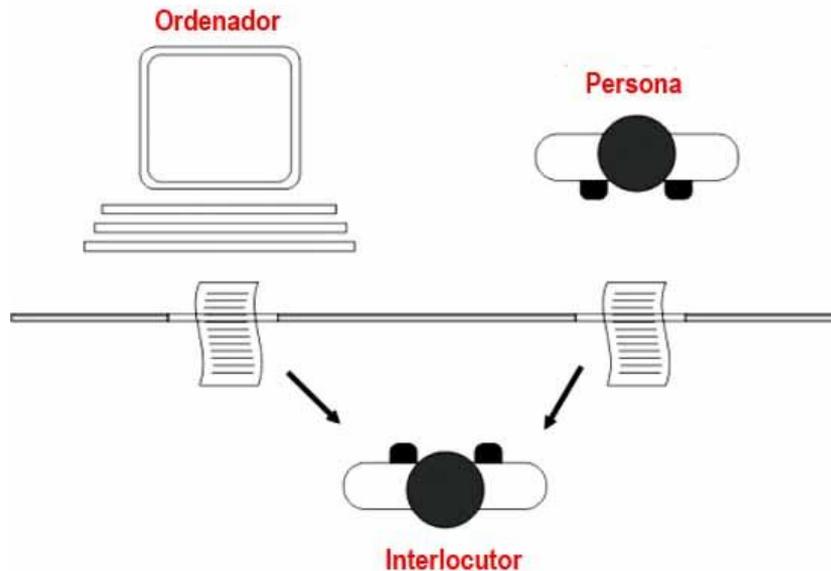


- Hierarchical feature Learning

Alan Turing: "Computing Machinery and Intelligence", 1950

(I propose to consider the question, "Can machines think?")

Test de Turing





¿Y qué es el PLN?

Procesamiento de Lenguaje Natural

- Recuperación de información
- Traducción automática
- Respuestas a preguntas
- Análisis de sentimiento
- Chatbots



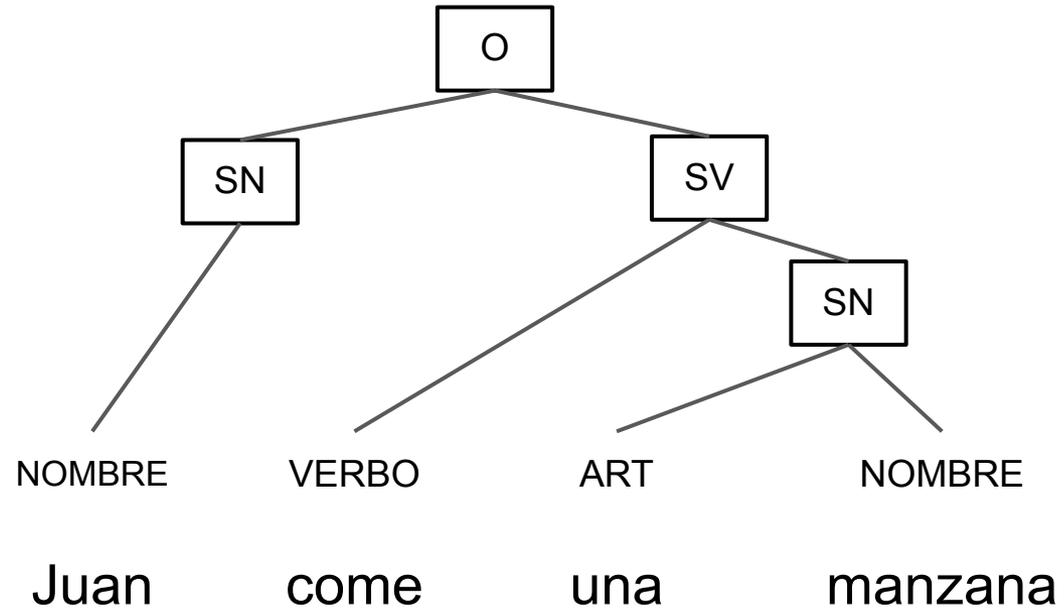
Busca en Google o escribe una URL

receta de merengue italiano → 意大利蛋白酥皮食譜



Procesamiento de Lenguaje Natural

- Parsing
- POS Tagging



Análisis de Sentimiento

Qué bien que
marcaron a Messi



Gracias a la gente de
Brookfield por su excelente
atención. <https://t.co/r38i4q7clV,P>



pues a mi me ha
encantado #sherlock



@user top secret: el #psoe de
andalucía despilfarra otros 90 millones de
euros en subvenciones. <http://t.co/aiwhnxrj>



Traducción Automática

- Mitâ ohupytyva'erâ vy'apavê → Un niño debe ser feliz
- Conadis aty'i omba'apohína oúvo. → Reunión de trabajo de la Subcomisión de Conadis.
- Odenunsia oñemondaha mokõi mymba → Denuncian robo de dos animales vacunos
- Jekupyty niko oĩ ñandepype → La solidaridad está en todos nosotros
- Rogueru mandi'o, kumanda, avati, opáicha ore roñemitÿva. → Trajimos mandioca, poroto, maíz, todo tipo de producción que tenemos.

Autocompletar



Debido a las copiosas

lluvias $P=0.8$

nevadas $P=0.05$

árbol $P=0.000000001$

...

¿Cómo lo hace el celular?
Contando qué tan frecuentes
son las palabras y las frases
(*modelo de lenguaje de N-gramas*)

¿Y cómo lo hace ChatGPT?



The notion “probability of a sentence” is an entirely useless one, under any known interpretation of this term.

*Noam Chomsky
1969*

*Anytime a linguist leaves
the group the recognition
rate goes up.*

*Fred Jelinek
1988*



Historia del PLN

- Inicios en los 1950s, traducción automática
- Hasta los 1980s, sistemas basados en reglas
- 1990s y 2000s, sistemas estadísticos
- 2010s hasta hoy, redes neuronales

Historia del PLN y RN

- 2001 - Primer Modelo de Lenguaje neuronal (Bengio)

- 2008 - Multi-task Learning (Collobert - Weston)

- 2013 - Word Embeddings (Mikolov)



- 2015 - Modelo atencional (Bahdanau)

Las RN se pueden usar ampliamente para PLN

- 2017 - Transformer (Vaswani)



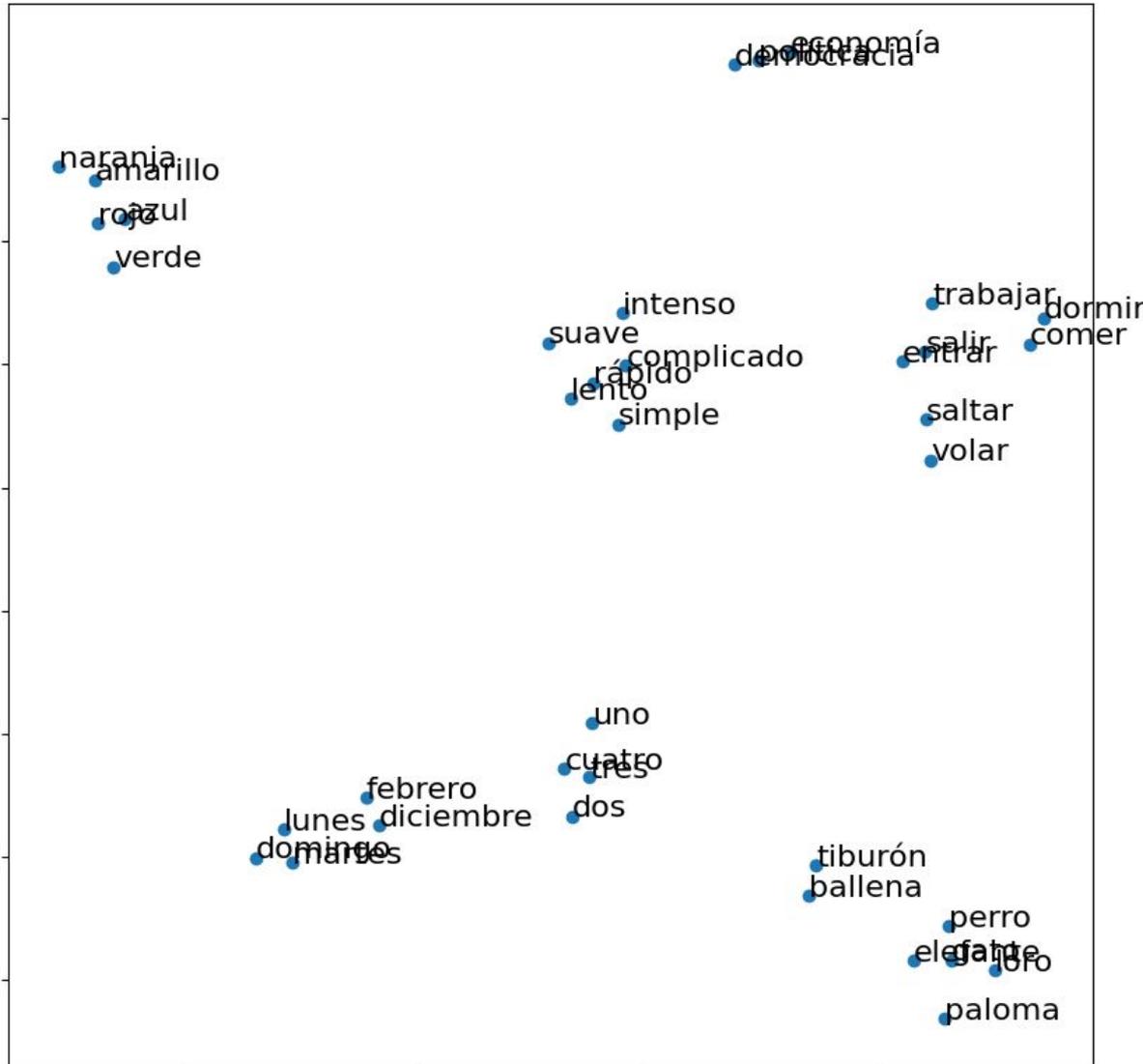
- 2020s - Modelos generativos (Brown)

Las RN son el estado del arte en PLN



Las RN permiten hacer cosas antes imposibles para PLN

Word Embeddings



Redes Recurrentes

The Unreasonable Effectiveness of Recurrent Neural Networks

May 21, 2015

We'll train RNNs to generate text character by character and ponder the question "how is that even possible?"

PANDARUS:

Alas, I think he shall be come approached and the day
When little strain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:

They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:

Well, your wit is in the care of side and that.

Second Lord:

They would be ruled after this chamber, and
my fair nudes begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:

Come, sir, I will make did behold your worship.

VIOLA:

I'll drink it.

```
/*  
 * Increment the size file of the new incorrect UI_FILTER group information  
 * of the size generatively.  
 */  
static int indicate_policy(void)  
{  
    int error;  
    if (fd == MARN_EPT) {  
        /*  
         * The kernel blank will coeld it to userspace.  
         */  
        if (ss->segment < mem_total)  
            unblock_graph_and_set_blocked();  
        else  
            ret = 1;  
        goto bail;  
    }  
    segaddr = in_SB(in.addr);  
    selector = seg / 16;  
    setup_works = true;  
    for (i = 0; i < blocks; i++) {  
        seq = buf[i++];  
        bpf = bd->bd.next + i * search;  
        if (fd) {  
            current = blocked;  
        }  
    }  
    rw->name = "Getjbbregs";  
    bprm_self_clearl(&iv->version);  
    regs->new = blocks[(BPF_STATS << info->historidac)] | PFMR_CLOBATHINC_SECONDS << 12;  
    return segtable;  
}
```

Grandes Modelos de Lenguaje

Fine-tuning

The model is trained via repeated gradient updates using a large corpus of example tasks.



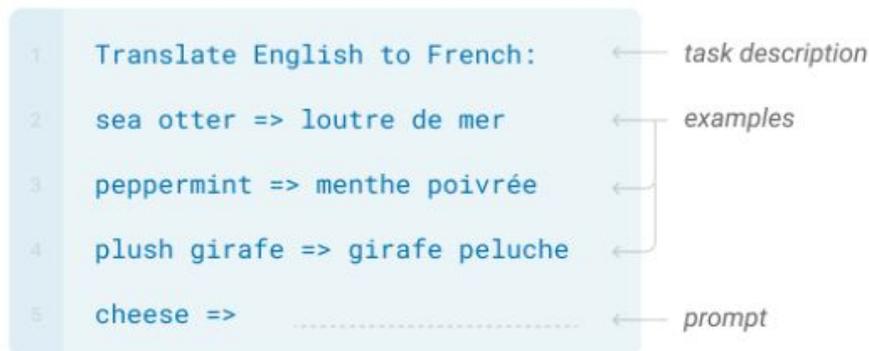
Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.



Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



PLN y RN

- Grandes cantidades de datos
- Gran demanda de cómputo
- Entrenamiento de modelos
- Evaluación

Tipos de problemas

1 palabra (o un par) \rightarrow 1 categoría

frío y caliente son sinónimos?
antónimos?

n palabras \rightarrow 1 categoría

este tweet tiene sentimiento positivo?
este tweet es un chiste?
este mail es spam?

n palabras \rightarrow 0..n categorías

de qué temas habla este texto?
qué emociones presenta este tweet?

n palabras \rightarrow n categorías

POS-tagging, NER, chunking,
parsing, SRL

n palabras \rightarrow m palabras

traducción automática
respuestas a preguntas
resúmenes automáticos
chatbots...

*Todos los problemas los podemos
reducir a esto?*



Metodología de las RN

- Aprendizaje automático supervisado

Training, desarrollo, test

- Familia muy flexible de funciones

Podemos modelar casi cualquier función (Cybenko, 1989)

Pero sobreajusta en seguida!

- Cómo encontramos la mejor función posible?

Metodología de las RN

- Cómo representamos la entrada?
One-hot, BOW, embeddings, tokenización
- Arquitecturas:
Fija: MLP
Secuenciales: CNN / LSTM
Atencional: Transformer
- Tamaño de los modelos
En general son miles de millones de parámetros!