

Lingwars

Los mundos de Siri, Google y Alexa

¡Hola!



@sailormercury

Nerea Luis Mingueza

AI Engineer en Sngular

Doctora en Inteligencia Artificial (UC3M)

T3chFest cofounder

Inteligencia Artificial

¿Qué es? ¿Cuáles son las áreas de trabajo?

Áreas de conocimiento

Representación del
conocimiento

Planificación y
razonamiento

Aprendizaje
Automático

Visión por
Computador

Procesamiento del
lenguaje natural

Robótica

Modelo
+
Lenguaje de representación
+
Algoritmo

Inteligencia Artificial Simbólica:

Conceptos (símbolos) + Operadores = solución

Inteligencia Artificial

Aprendizaje Automático

Aprendizaje Profundo

Redes de neuronas

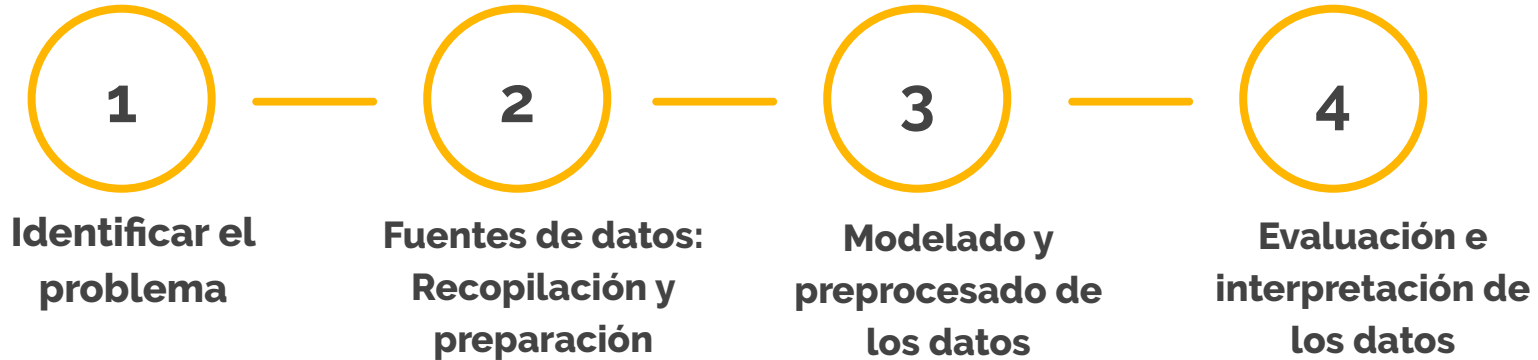


Machine Learning

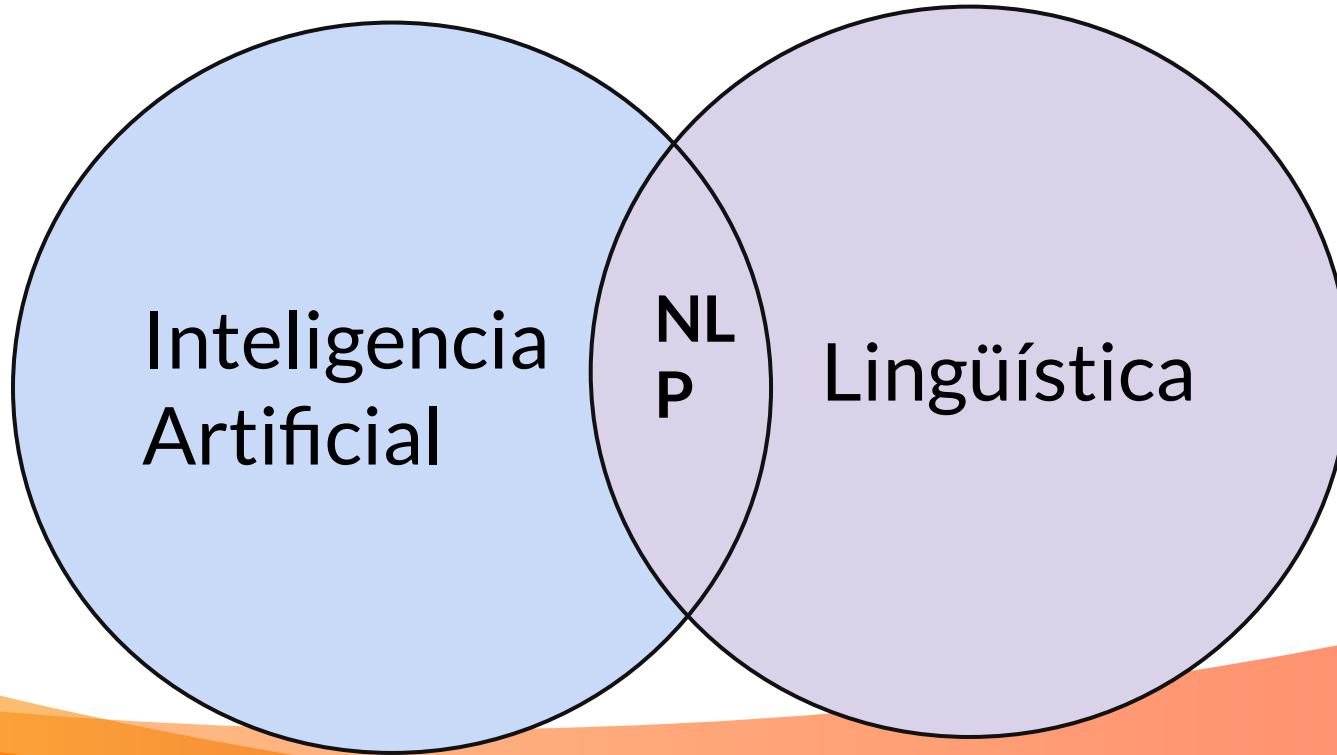
- Conjunto de datos de entrada
- Conjunto de datos de salida (opcional)
- Parámetros
- Toma de decisiones:
 - Árboles
 - Reglas
 - Redes de Neuronas
 - Support Vector Machines...

Data Mining Pipeline

El proceso para llevar a cabo lo conocido como 'minería de datos'



¿QUÉ ES NATURAL LANGUAGE PROCESSING?



¿QUÉ ES NATURAL LANGUAGE PROCESSING?

Tareas con aplicación directa:

- Traducción de textos
- Transcripción de textos
- Análisis sintáctico-semántico
- Recuperación de información
- Análisis del sentimiento
- Elaboración de pregunta-respuesta
- Expresividad y entonación

LIBERÍAS Y CORPUS DE NATURAL LANGUAGE PROCESSING

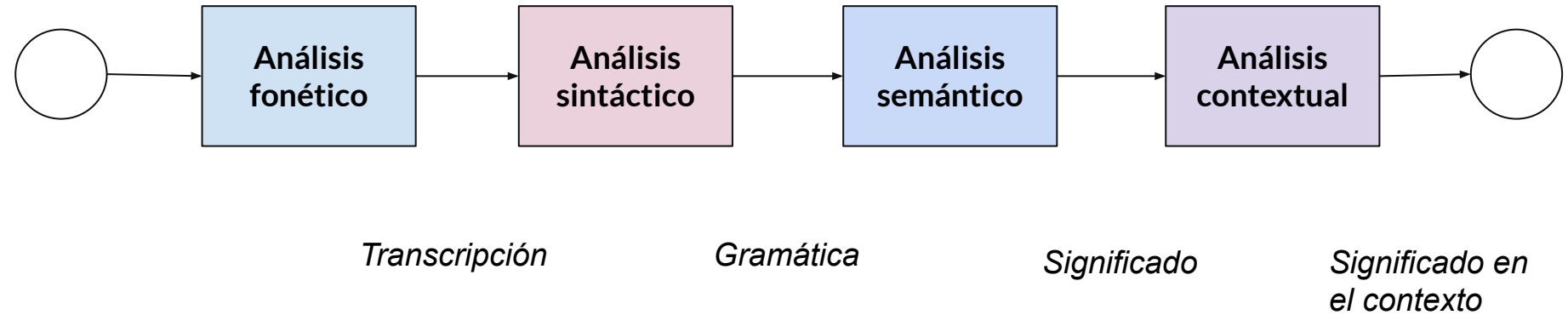
Corpus

- Reseñas
- Sentimiento
- Idiomas
- Noticias
-

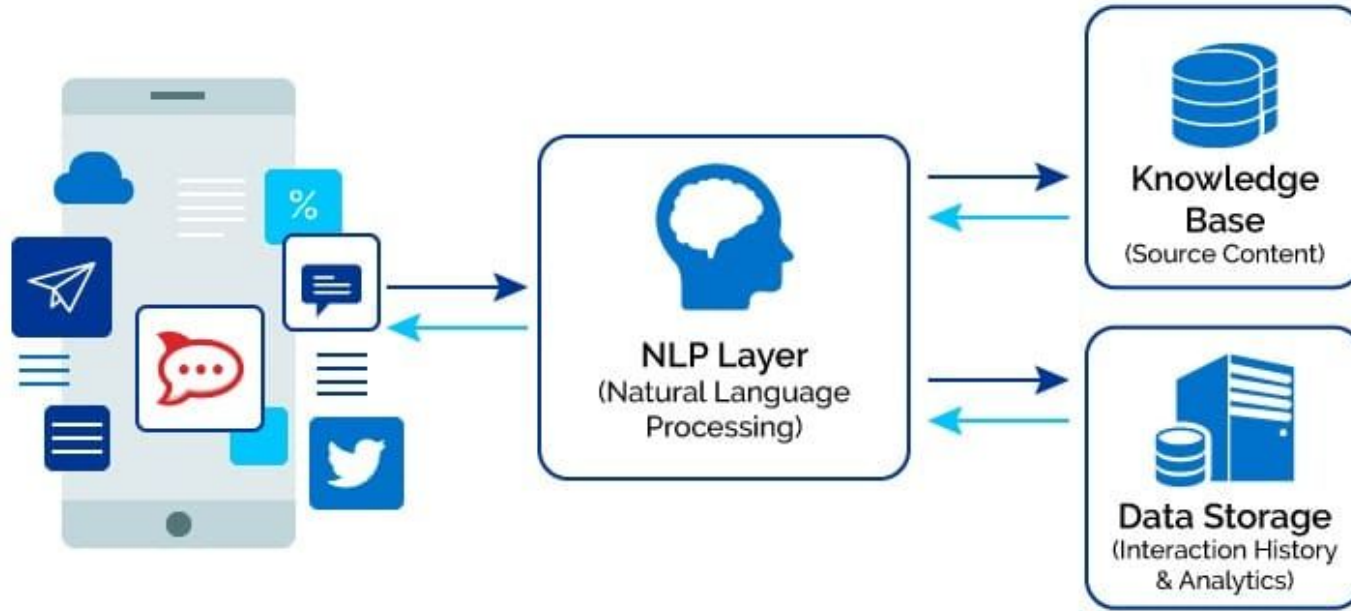
+



EL PROCESO DE RECONOCIMIENTO DEL LENGUAJE



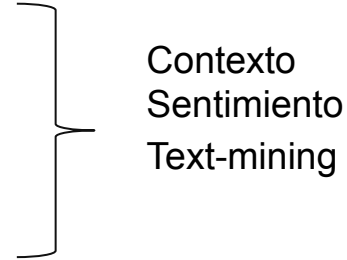
EL PROCESO DE RECONOCIMIENTO DEL LENGUAJE



EJEMPLO DE ARQUITECTURA PARA UN BOT

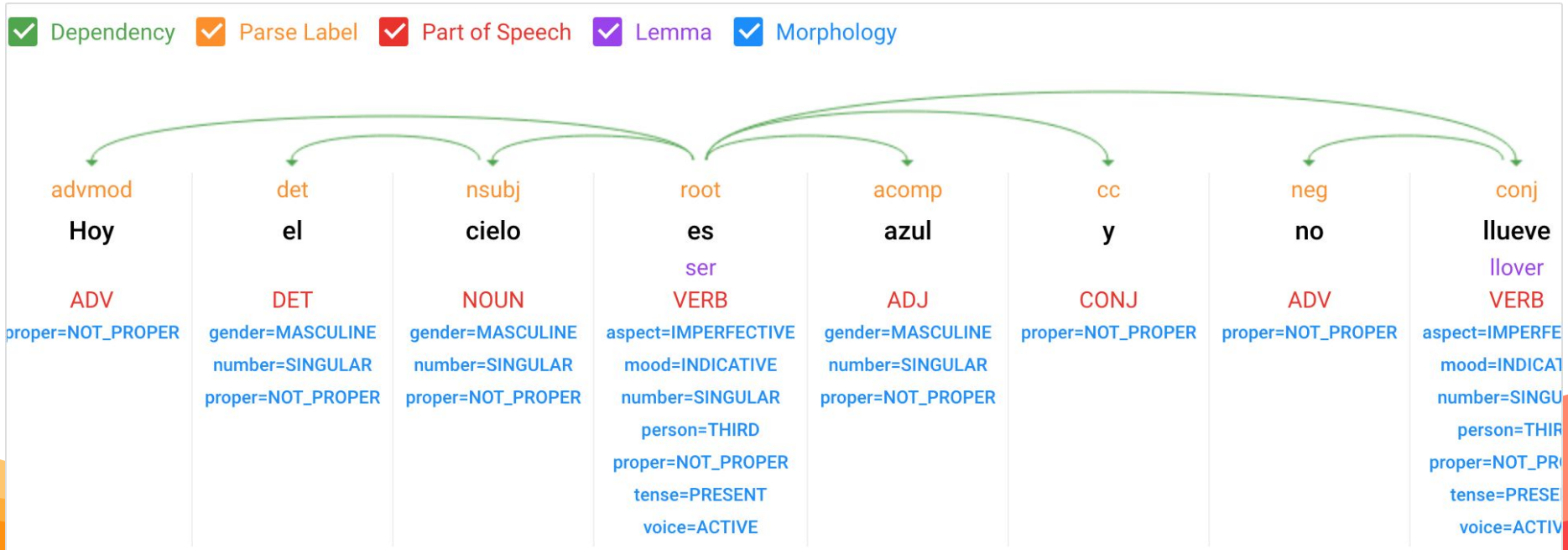
¿QUÉ TENER EN CUENTA EN LA TRANSCRIPCIÓN?

- Frecuencia del habla
- Pausas
- Lista de *stop-words* o palabras vacías
- Lista de palabras más usadas

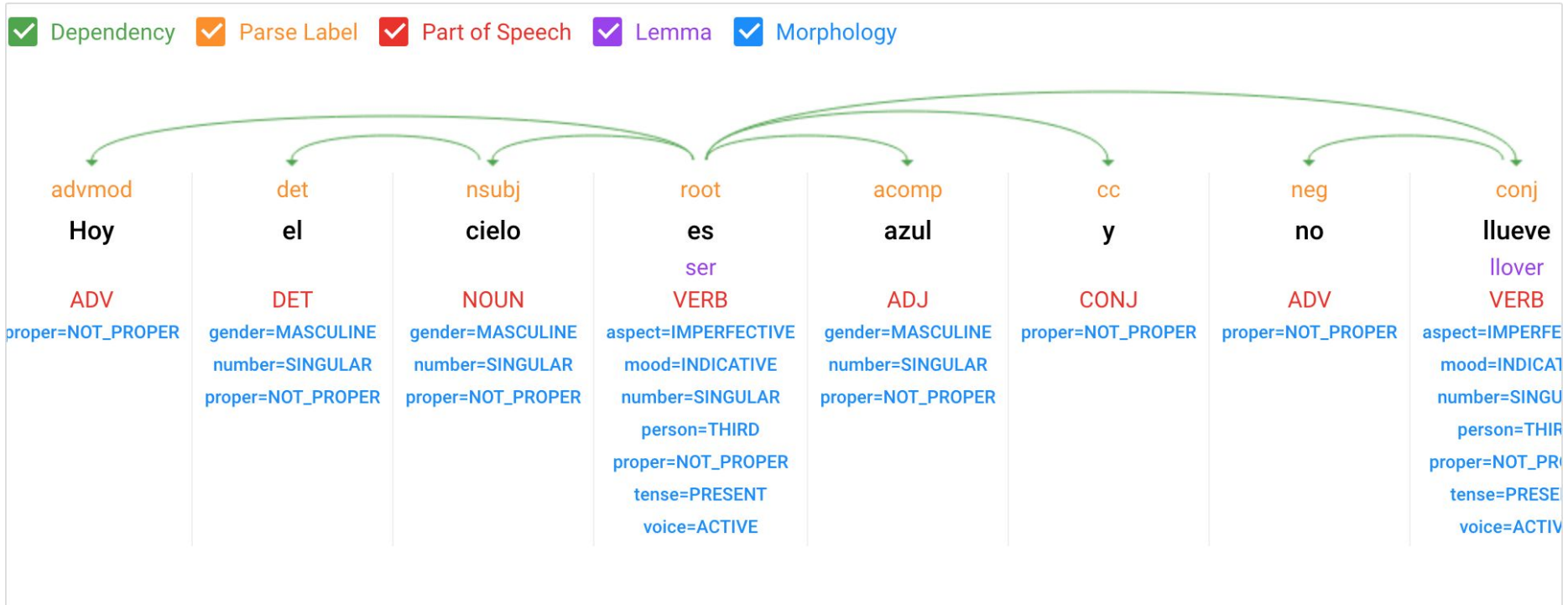


Un *corpus* es un conjunto de textos que comparten idioma y temática (si son más específicos). Se almacenan electrónicamente y se procesan a modo de análisis, validación y verificación de reglas lingüísticas.

ANÁLISIS SINTÁTICO-SEMÁNTICO



ANÁLISIS SINTÁTICO-SEMÁNTICO





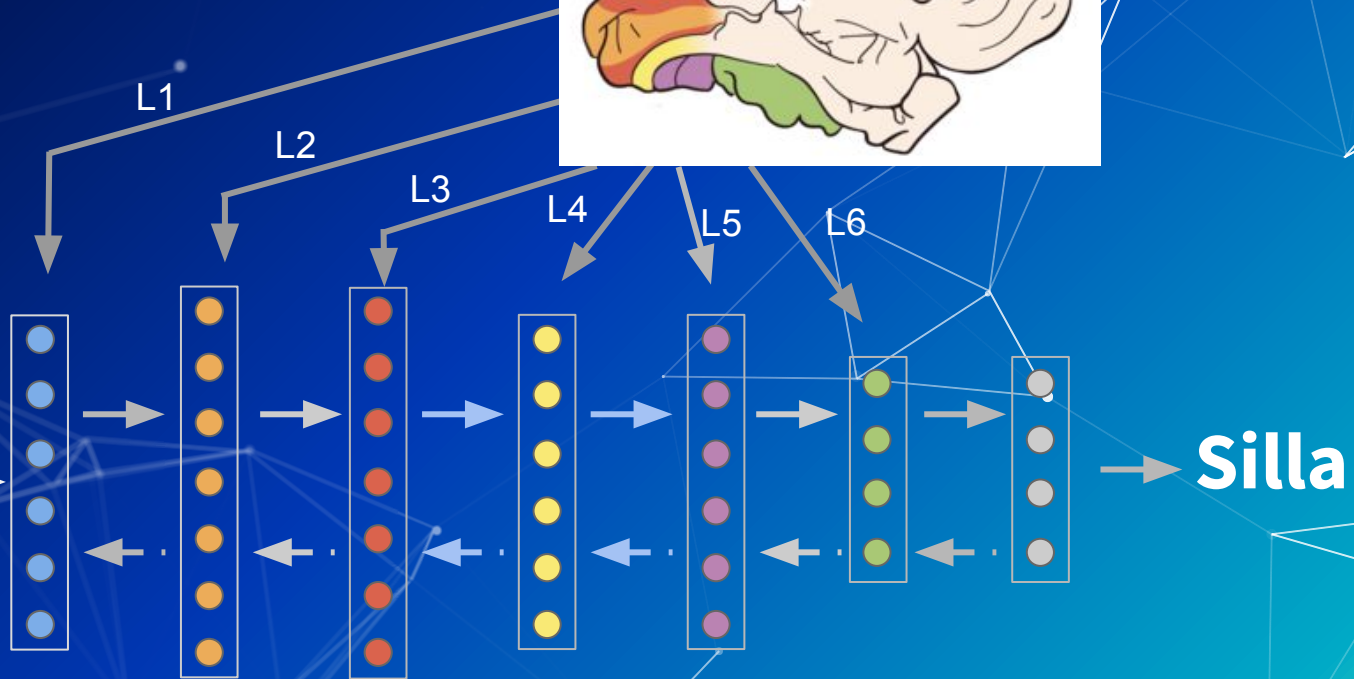
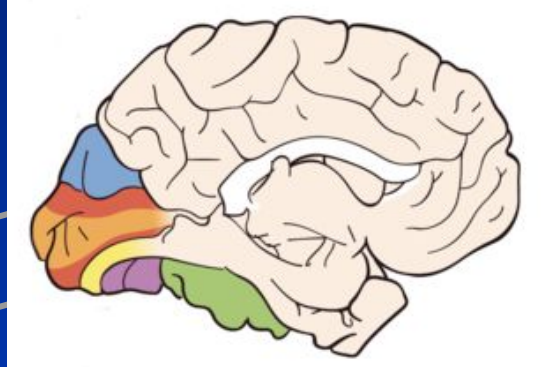
Redes de neuronas

¿Neuronas? ¿No son las mías? Si yo tengo muchas

Principales componentes

- > Utilizamos elementos llamados '**neuronas**'
- > Conectadas de forma múltiple entre sí formando una '**red**'
- > Damos importancia a los '**datos de entrada**'
- > Escogemos **pesos, función de activación**
- > Interpretamos la **salida**:
 - Semántica de los datos
 - Características Comunes

» Copiar proceso cognitivo humano



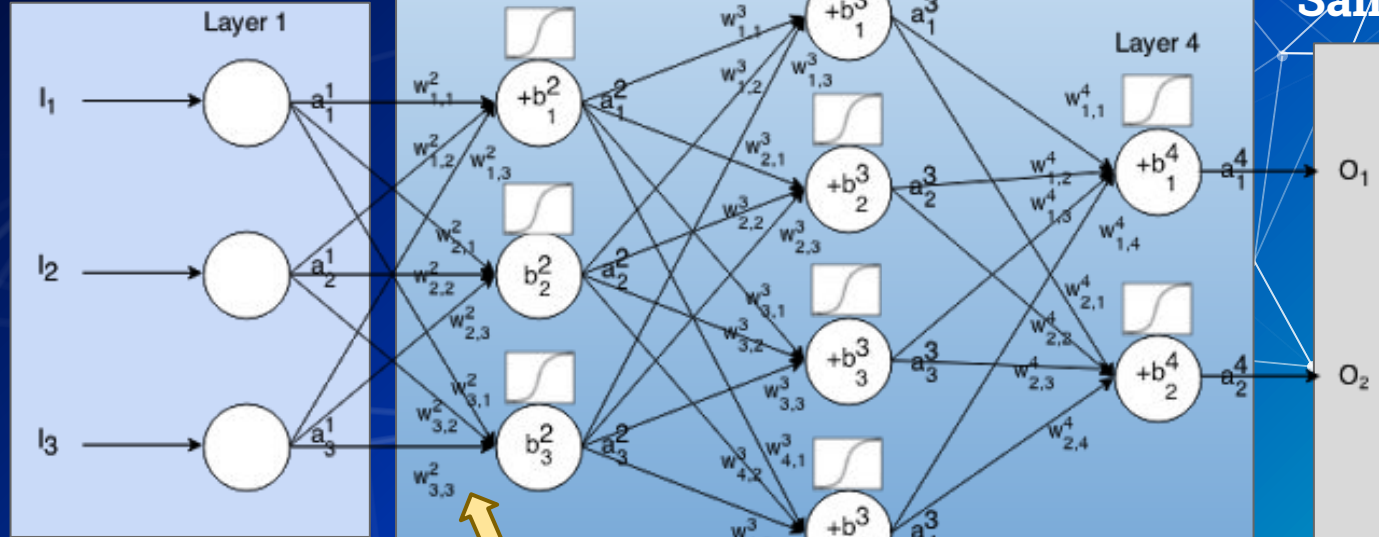
Colección de funciones matemáticas **sencillas**

Función de activación

Salida

Capas ocultas

Datos de entrada



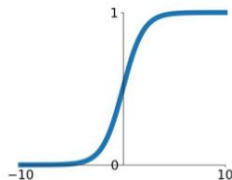
Pesos

Gradiente, función de coste, umbral...

Funciones de activación

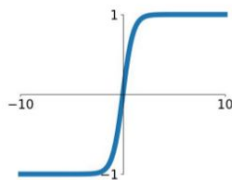
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



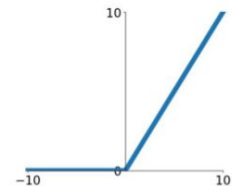
tanh

$$\tanh(x)$$



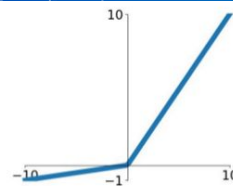
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

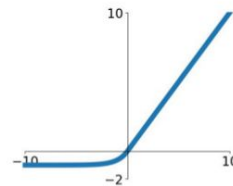


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

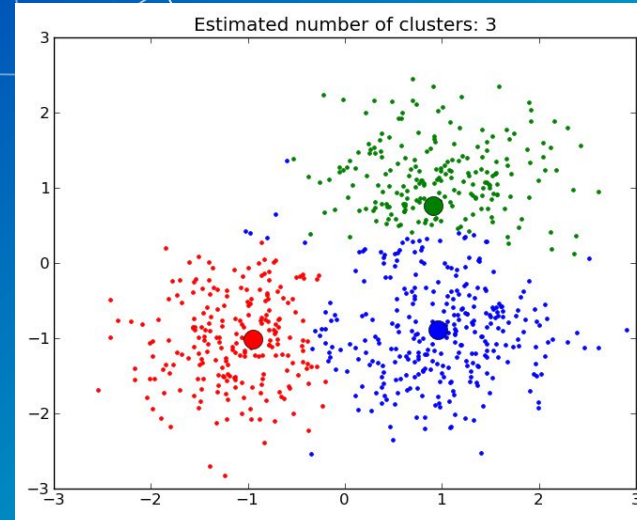
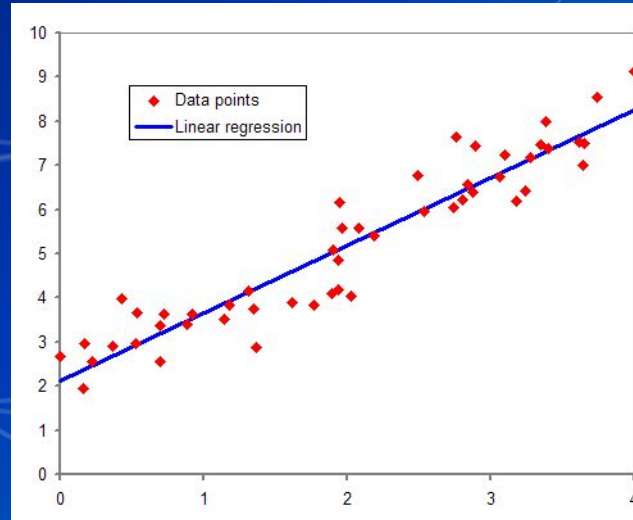
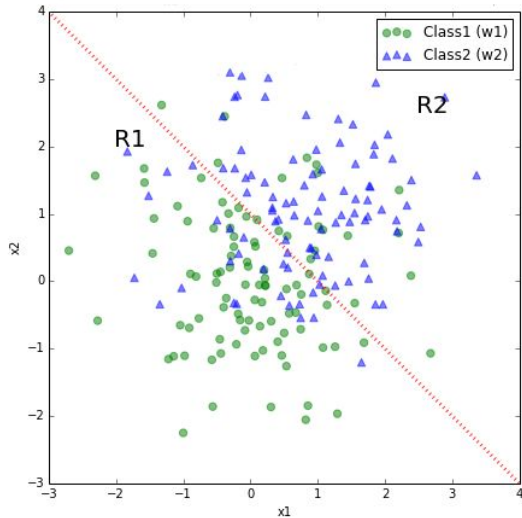
ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



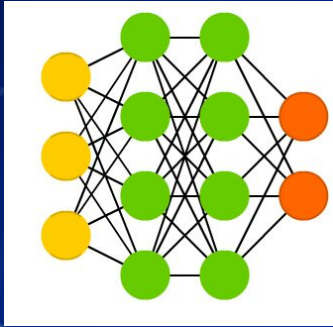
¿Cuál es el objetivo?

- Supervisado: clasificación/regresión
- No supervisado: estimación/agrupamiento

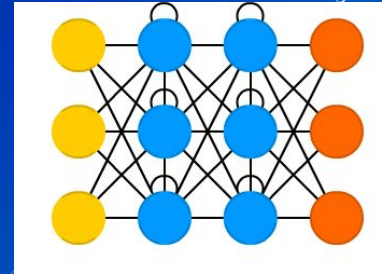


Tipos de redes de neuronas

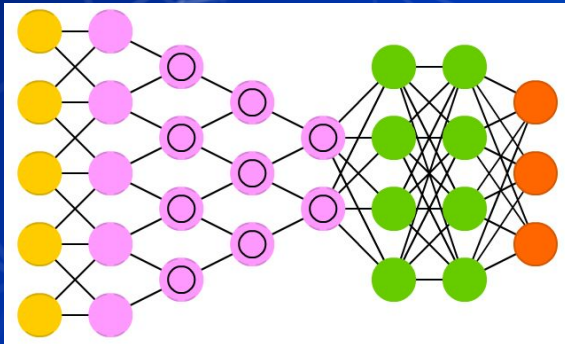
Feed Forward (FFN)



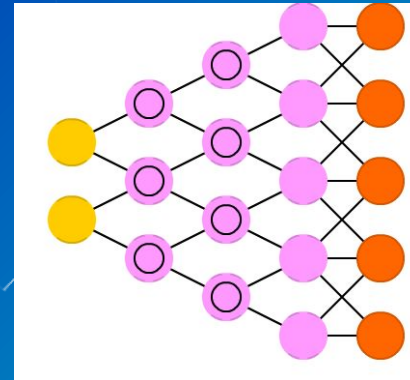
Recurrentes (RNN)



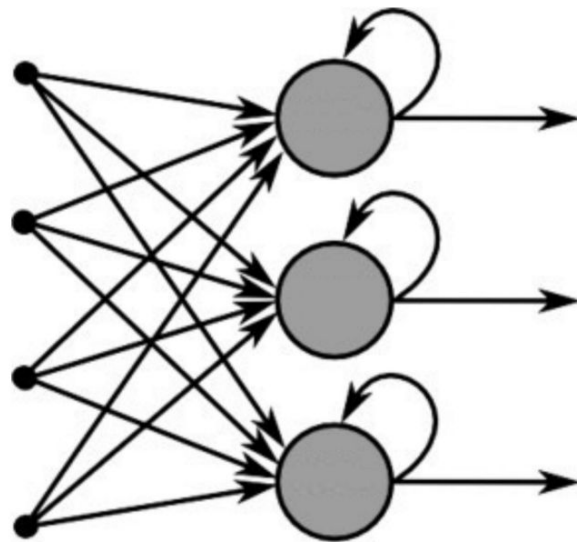
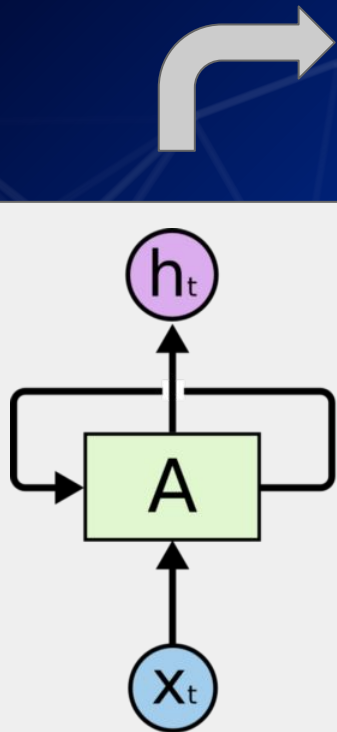
Convolucionales (CNN)



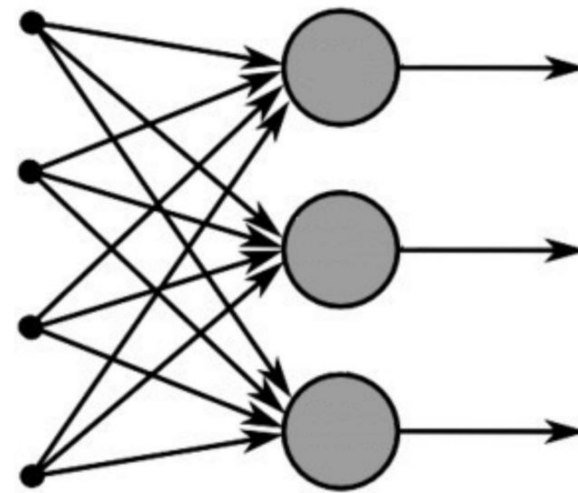
Deconvolucionales (DNN)



Redes Recurrentes

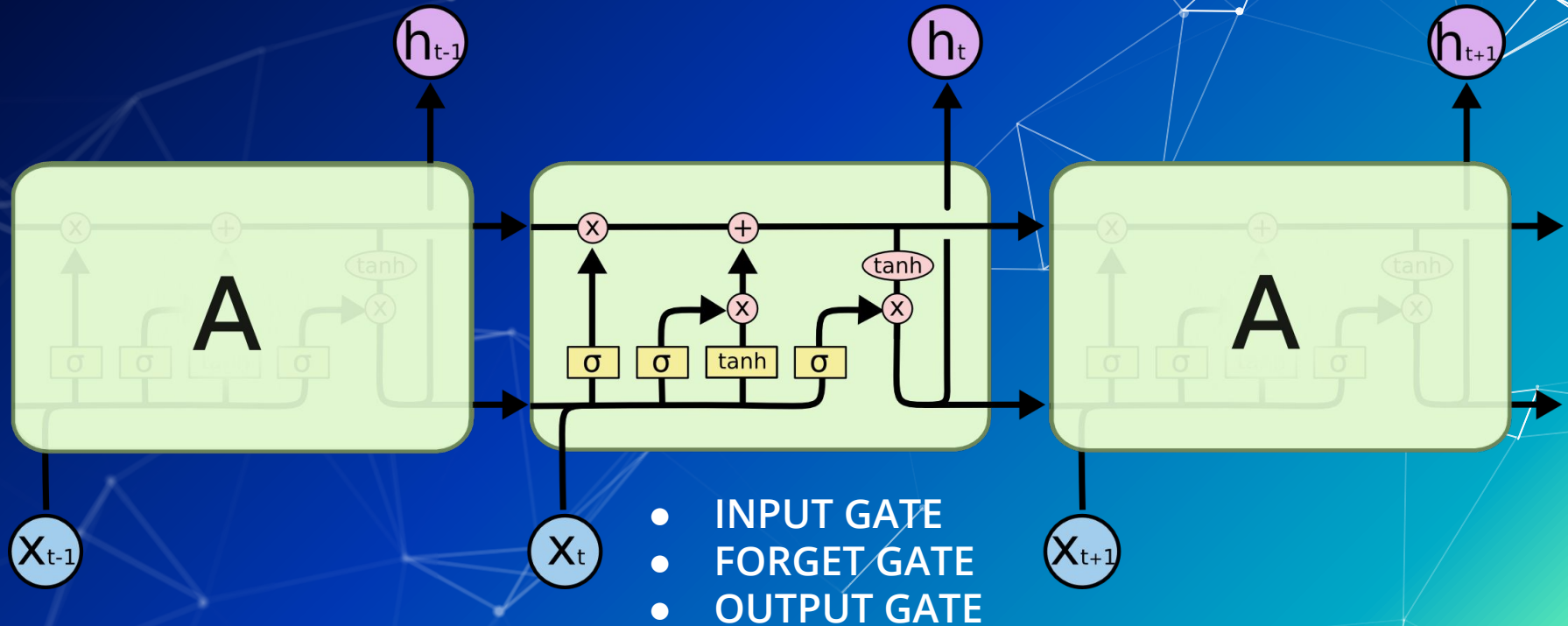


Recurrent Neural Network



Feed-Forward Neural Network

Redes Long Term Short Memory



Recursos de Redes de Neuronas/ML

- **AI for Everyone**
- **Elements of AI**
- **Fast.ai**
- **Codelabs de Google**
- **Jupyter Notebooks**



Epoch
000,396

Learning rate
0.03

Activation
ReLU

Regularization
L1

Regularization rate
0.01

Problem type
Classification

DATA

Which dataset do you want to use?



Ratio of training to test data: 50%



Noise: 0



Batch size: 10



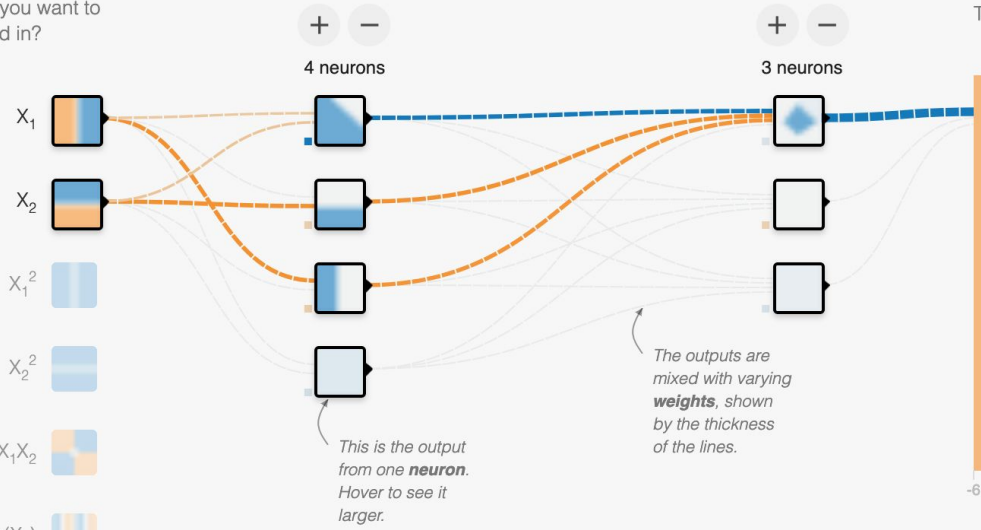
REGENERATE

FEATURES

Which properties do you want to feed in?

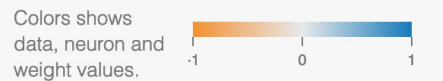
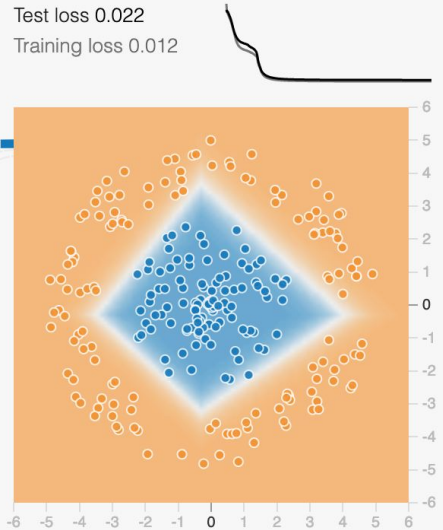
- X_1
- X_2
- X_1^2
- X_2^2
- X_1X_2
- $\sin(X_1)$
- $\sin(X_2)$

2 HIDDEN LAYERS

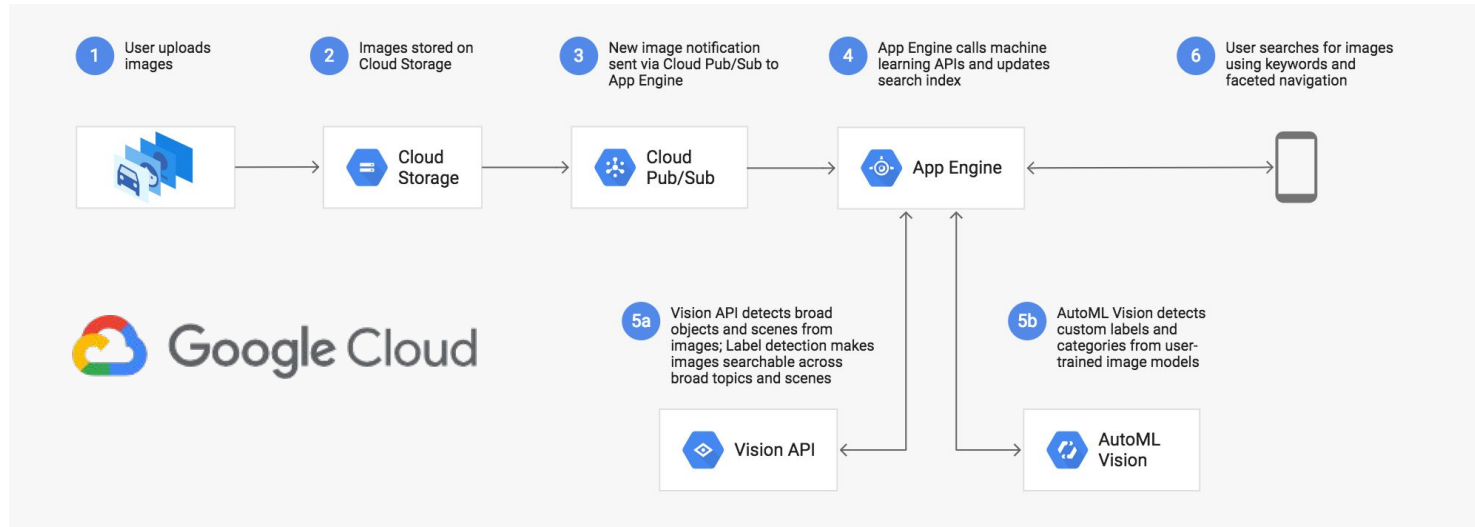


OUTPUT

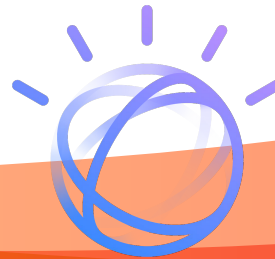
Test loss 0.022
Training loss 0.012



SERVICIOS DE APRENDIZAJE



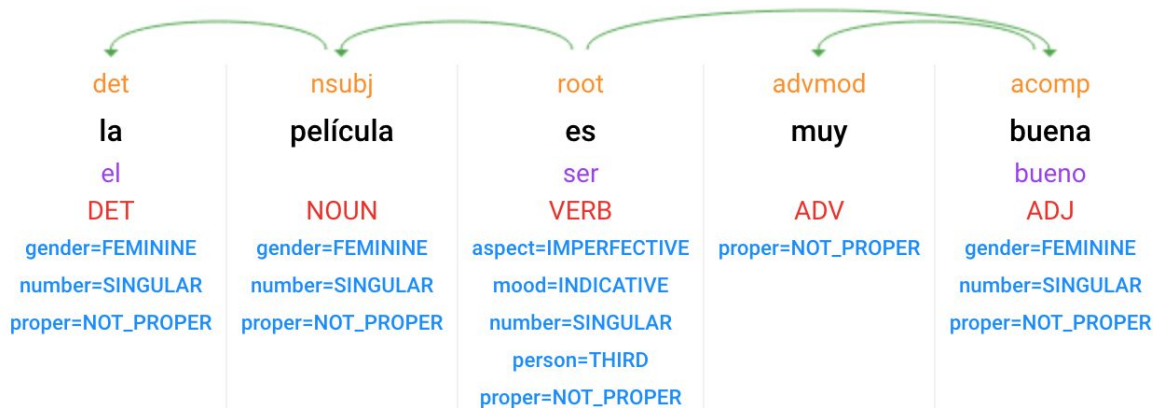
Cognitive Services



EJEMPLO CON GOOGLE CLOUD

AutoML + NLP

<https://cloud.google.com/natural-language/automl/docs/beginners-guide?hl=es-419>



Document & Sentence Level Sentiment

Entire Document

la película es muy buena

Score Magnitude

0.8

0.8

0.8

0.8

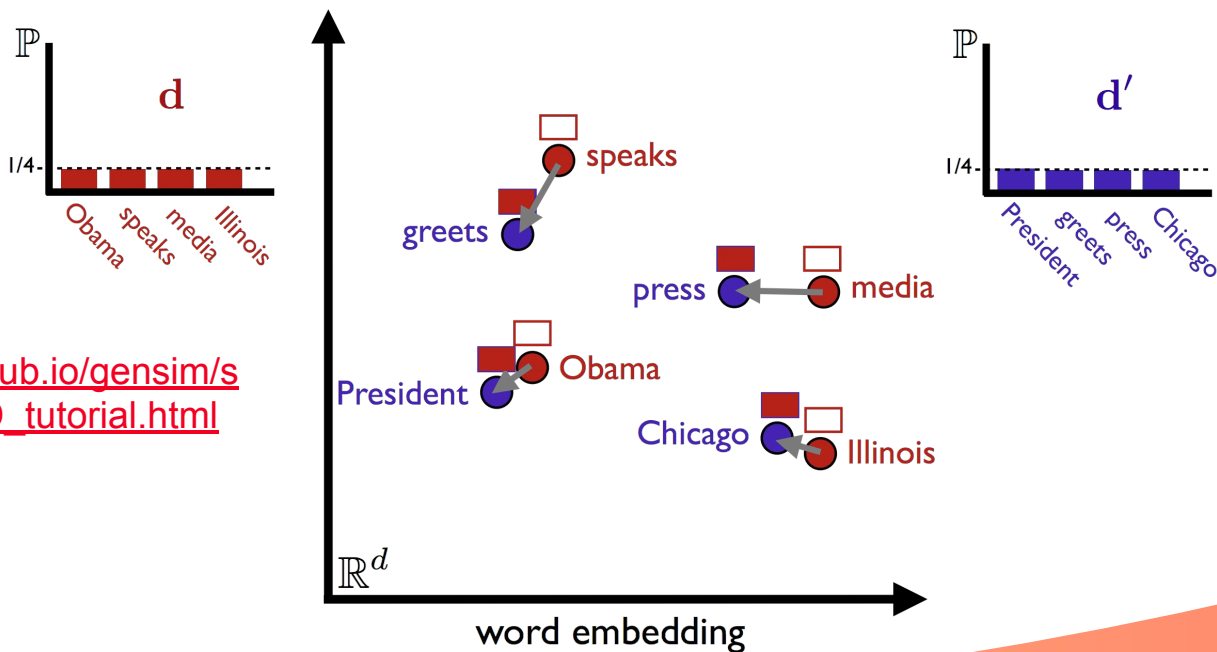
Score Range

-1.0 — -0.25

-0.25 — 0.25

0.25 — 1.0

WORD EMBEDDINGS



https://markroxor.github.io/gensim/s-tatic/notebooks/WMD_tutorial.html

Recursos de Word Embeddigs

<https://github.com/facebookresearch/StarSpace>

<https://towardsdatascience.com/word-embeddings-exploration-explanation-and-exploitation-with-code-in-python-5dac99d5d795>

¡Gracias!

¿Preguntas?

Contáctame en:

- @sailormerqury
- [linkedin.com/in/nerealuis](https://www.linkedin.com/in/nerealuis)