



ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

# Fundamentos Machine & Deep Learning

Diplomado Big Data  
y Analítica de Datos 2022-2023

Profesor:

**Rodrigo Sandoval U.**



← Responder   ← Responder a todos   → Reenviar   🗑 Eliminar   📌 Establecer marca   ⋮

### Confirmación hora médica, martes 24 de abril, VidaIntegra



Camila Eniax <camilaeniax@vidaintegra.cl>

21-04-2018 18:23

Para: rodrigo@naturastock.com

Estimado Rodrigo,

Le escribo para confirmar su hora con Paula Leiva Araya, Medico General, para el martes 24 de abril a las 10:45 en VidaIntegra.

¿Asistirá a la hora?

Quedo a la espera de la respuesta.

Saludos,

Camila,  
Ejecutiva Vidaintegra



### RE: Confirmación hora médica, martes 24 de abril, VidaIntegra



Rodrigo Sandoval <rodrigo@naturastock.com>

21-04-2018 18:49

Para: Camila Eniax

Estimada Camila,

Confirmando asistencia este martes 24 de abril a las 10:45.

Atte,

Rodrigo Sandoval

### RE: Confirmación hora médica, martes 24 de abril, VidaIntegra



Camila Eniax <camilaeniax@vidaintegra.cl>

21-04-2018 18:51

Para: rodrigo@naturastock.com



Buenas tardes, Gracias por la respuesta. La hora queda confirmada para el martes 24 de abril con el Paula Leiva Araya a las 10:45. La atención será en VidaIntegra, ubicado en Avenida El Bosque Norte 0110, Las Condes. El bono lo puede comprar en el centro médico a través de sistema imed. Recuerde llevar su cédula de identidad y llegar con 15 minutos de anticipación. Si por algún motivo no puede asistir, le pido notificarme por esta vía para poder reasignar la hora a otra persona. Hasta luego, Saludos, Camila Eniax, Ejecutiva Vidaintegra





**De:** Isabel ...

**Fecha:** 30 de julio de 2018, 17:33:46 CLT

**Para:** Alicia Eniax <[alicia.eniax@...](mailto:alicia.eniax@...)>

**Asunto: Re: Confirmación cita médica, miércoles 1 de agosto, Integramedica**

Si ando con ganas, voy

Isabel ...

El 30-07-2018, a la(s) 08:38, Alicia Eniax  
<[alicia.eniax@...](mailto:alicia.eniax@...)> escribió:

Buenos días Isabel,

Soy Alicia, Ejecutiva de Salud. Le escribo para  
confirmar su hora con el Dr Veronica \_\_\_\_\_ ,  
para el miercoles 1 de agosto a las 13:15, en  
sucursal Las Condes.  
¿Asistirá a la cita?

Muchas gracias y quedo a la espera de su  
respuesta.

Saludos cordiales,  
Alicia Eniax,  
Ejecutiva IntegraMédica

**De:** Alicia Eniax <[alicia.eniax@...](mailto:alicia.eniax@...)>

**Fecha:** 30 de julio de 2018, 17:39:03 CLT

**Para:** <[email](mailto:email)>

**Asunto: Re: Confirmación cita médica, miércoles 1 de agosto, Integramedica**

**Responder a:** Alicia Eniax <[alicia.eniax@...](mailto:alicia.eniax@...)>

Hola,

Gracias por la respuesta. La cita queda confirmada para el  
miercoles 1 de agosto con el Especialidad Veronica Apellido  
Apellido a las 13:15.

La atención será en Sucursal Las Condes. Recuerde llevar su  
cédula de identidad y llegar con 20 minutos de anticipación.  
Puede comprar el bono en nuestros totems de atención  
digital, pagando con Tarjeta de Crédito o Débito.  
Para cambiar los datos de reserva debe llamar al 2 111 2 11 2  
o en recepción antes de la cita.

Saludos,  
Alicia Eniax,  
Ejecutiva Salud



# De las técnicas de NLP ...

La clasificación es útil para tomar decisiones sobre el contenido de un bloque de texto.

La extracción es útil para reconocer elementos clave dentro de un bloque de texto.

Pero al enfocarse en comparaciones más directas (entre palabras o conjuntos de palabras) no se logra el objetivo correcto **de interpretar un texto en lenguaje natural con toda su semántica.**



# Ejemplos y desafíos

*El producto cumple con mis expectativas y no tengo problemas.*

vs.

*Tengo problemas con el producto y no cumple mis expectativas.*

*Contexto 1: Análisis del nivel de empleabilidad mensual.*

*Contexto 2: Reporte médico de un paciente de nombre Julio.*

*Los niveles de Julio se ven mejores que antes.*

*El informe explica con claridad el éxito de las iniciativas comerciales durante el año.*

*El reporte confirma la correcta decisión de estrategias de venta de los últimos 12 meses.*



# Ejemplo de completar palabras ...

Los alumnos van a abrir su \_\_\_\_\_

Alternativas: libro, cuaderno, laptop, mente

Los alumnos no van a abrir su \_\_\_\_\_

Alternativas: pasado, voto, crédito

¿De qué depende?

Del contexto.





# Para lograr entendimiento de lenguaje natural ...

Se necesita conocer el **contexto** y/o la relación lingüística de las palabras.

Se necesita considerar las palabras en una **secuencia**, para poder predecir la próxima palabra.

Se necesita poder **armar una respuesta** que depende de los elementos de la pregunta y del contexto, para que sea útil al receptor.



## NLP

*Natural Language  
Processing  
(Qué se dijo)*

## NLU

*Natural Language  
Understanding  
(Qué significa)*

## NLG

*Natural Language  
Generation  
(Qué se responde)*





## 8.2. NLU y NLG



# Objetivos en NLU y NLG

## Interpretación

- Relación entre porciones de texto y la semántica equivalente.
- Incorporación de contexto en la interpretación.

## Interacción

- Una combinación de los anteriores, con una capacidad de generar respuestas y predecir las palabras adecuadas.



# Técnicas más utilizadas hoy

Word Embedding

Secuencia en redes neuronales

Transformers



# Word Embedding

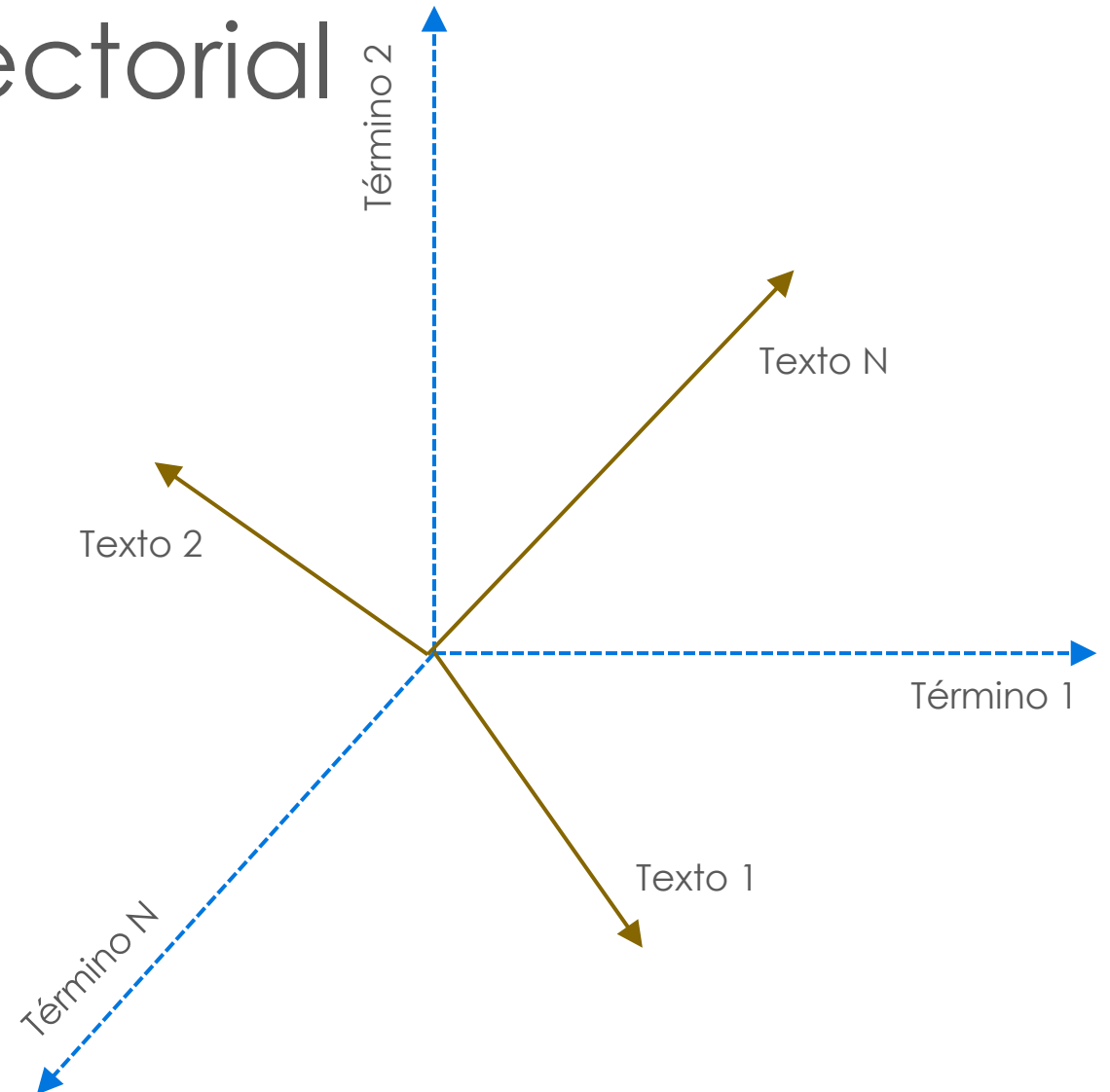
manejo de texto como vectores de ejemplos contextuales



# Modelo de Espacio Vectorial

**VSM** (Vector Space Model) es un espacio en el que el texto es representado por un vector de números, en lugar de su representación original.

Hay diferentes enfoques de vectorización, coincidiendo en referenciar la ocurrencia de los términos en diferentes porciones de texto (documentos o frases)





# Vectores de Distribución

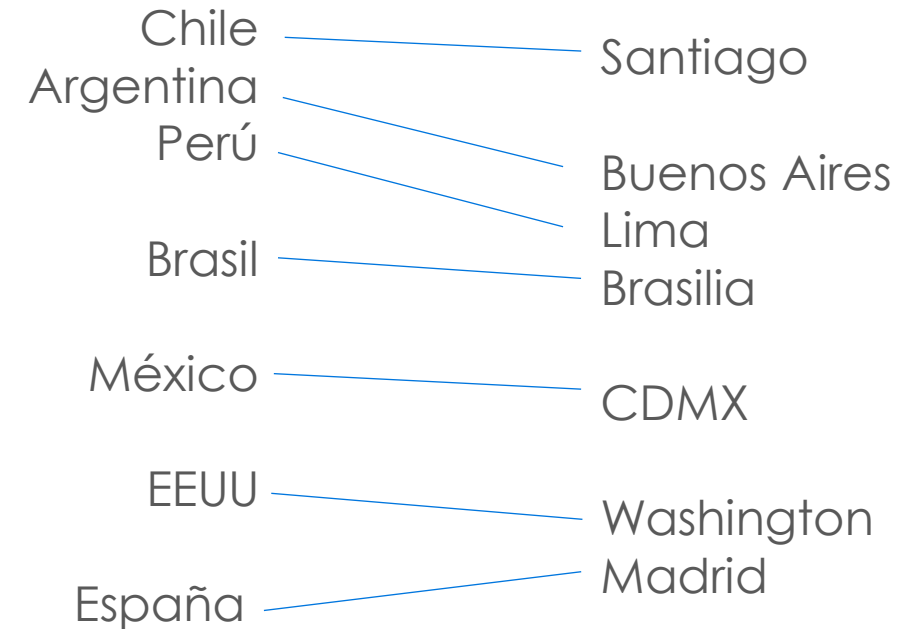
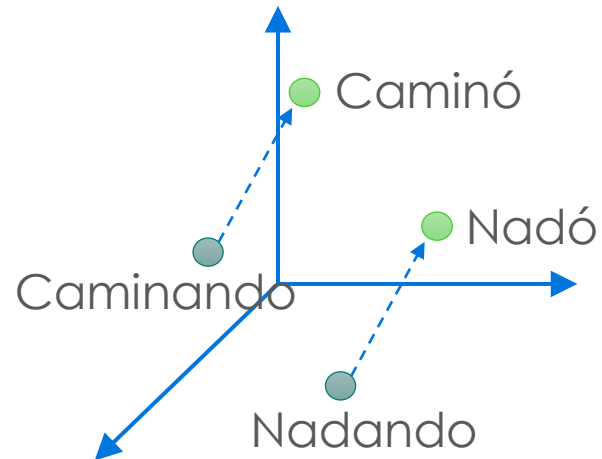
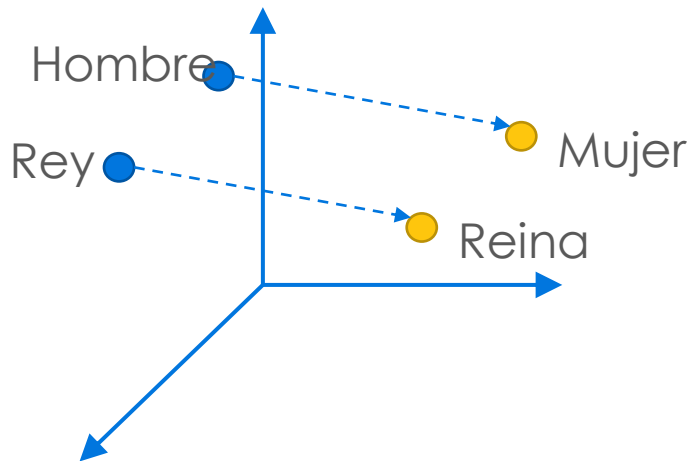
**Hipótesis de Distribución** [Harris, 1954]: las palabras que ocurren en los mismos **contextos**, tienden a tener significados similares.

*“Una palabra es caracterizada por su compañía”*

**Representaciones de Distribución:** las palabras son representadas por vectores de alta dimensionalidad basado en el contexto en que éstas ocurren.

# Word Embedding

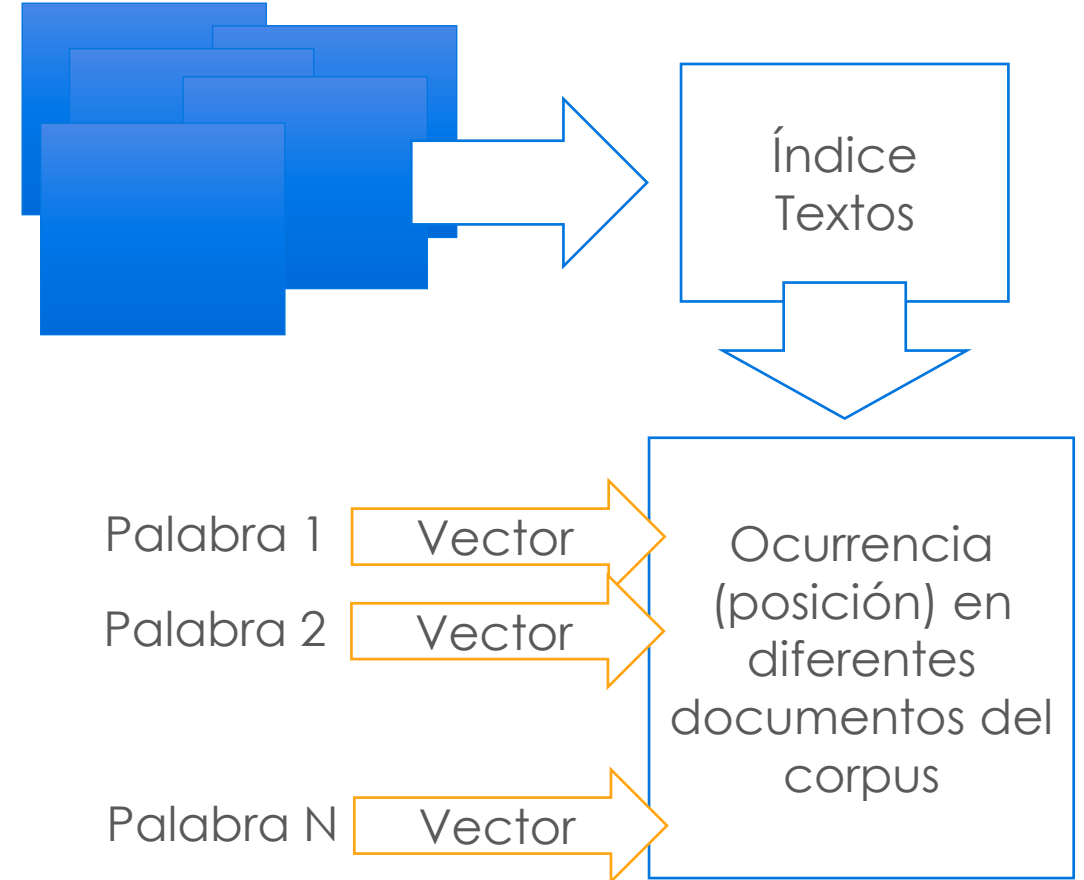
Incorpora la equivalencia o relación entre diferentes palabras en un lenguaje y contexto.





# Vectorización de Palabras – Word Vectors

- Se identifican y vectorizaron todas las palabras encontradas en un corpus (por ej, Wikipedia en español).
- Cada palabra, en un vocabulario, queda asociada a la posición donde – en general – aparece en los textos → Vector
- Entonces, se pueden comparar vectorialmente.





# Diferencia formas de vectores

Vectorización a un Vocabulario:

Simplificar la cantidad de términos a un vocabulario.

Vectorizar el texto  $[p1 \ p2 \ p3 \ ... \ pn] = [0 \ 1 \ 1 \ 0 \ ... \ 0]$

Vectorización Word Embedding

Simplifica el texto en los términos relevantes (sin reducir a la raíz / sin stemming ni lemmatization)

Vectorizar la palabra

$[D1 \ D2 \ D3 \ ... \ Dn] = [24 \ 0 \ 0 \ ... \ 745]$



# Word Embedding desde fuente masiva

## Wikipedia (ESP)

La enciclopedia libre

Portada

Portal de la comunidad

Actualidad

Cambios recientes

Páginas nuevas

Página aleatoria

Ayuda

Donaciones

Notificar un error

En otros proyectos

Wikimedia Commons

Wikiquote

Imprimir/exportar

Crear un libro

Descargar como PDF

Versión para imprimir

Herramientas

Lo que enlaza aquí

Cambios en enlazadas

Subir archivo

Páginas especiales

Enlace permanente

Información de la página

Elemento de Wikidata

Citar esta página

En otros idiomas

Unete al concurso y recibe una postal enviada desde Asia. [Ayúdanos con las traducciones!]

### Actor

Este artículo tiene [referencias](#), pero no tiene [verificabilidad](#). Puedes colaborar agregando referencias a material sin fuentes fiables podría ser [cuestionado](#). Este aviso fue puesto el 27 de diciembre de 2017.

Para otros usos de este término, véase [Actor \(desambiguación\)](#).

Se denomina **actor** (en femenino, **actriz**) a la persona que interpreta una acción, ya sea mediante improvisación o basándose en textos (obra de un [autor](#) o creados a través de improvisaciones individuales y colectivas). El actor construye su trabajo usando la voz, la mímica y otros recursos corporales y emocionales con el fin de transmitir al espectador el conjunto de ideas y acciones propuestas. Puede asimismo usar recursos técnicos u objetos como manipular títeres, o interpretar sobre la imagen o la voz de los demás. Combinando su ocupación con el director y el resto de los profesionales participantes en la obra, y apoyado en estímulos visuales o sonoros, actúa en lugares donde se realizan espectáculos públicos a través de las representaciones.<sup>1</sup>

Los actores de la obra *Los actores de la obra* pintados por Antonio...

Imprimir/exportar

Crear un libro

Descargar como PDF

Versión para imprimir

WIKIPEDIA  
La enciclopedia libre

Portada

Este noviembre es el Mes de Asia en Wikipedia. Únete al concurso y recibe una postal enviada desde Asia. [Ayúdanos con las traducciones!]

Artículo **Discusión**

Leer Ver código Ver historial

Buscar en Wikipedia

Coordenadas:  33°27′00″S 70°40′00″O (mapa)

...mino, véase [Chile \(desambiguación\)](#).

extremo

Su nombre oficial

capital es la

so Nacional está

geográficas.<sup>4</sup>

inental,

sta occidental del

Cono Sur que se extiende entre los paralelos 17°29′57″ S y 56°32′12″ S, mayormente desde la ribera sudoriental del océano Pacífico hasta las cumbres más altas divisorias de aguas de la cordillera de los Andes. Alcanza un largo de 4270 km <sup>26</sup> un ancho máximo de 445 km en

Imprimir/exportar

Crear un libro

Descargar como PDF

Versión para imprimir

República de Chile

Bandera

Escudo

Lema: «Por la razón o la fuerza»

Himno: Himno nacional de Chile

0:00

¿Problemas al reproducir este archivo?

Referencia esencial:

Jorge Pérez

@perez

Universidad de Chile

Instituto Milenio

Fundamentos de los Datos



## DATA

1 tensor found

Palabras comunes

Edit by

label

Tag selection as

Load

Publish

Download

Label

☒ Sphereize data ?

Checkpoint:

Metadata: [https://raw.githubusercontent.com/jorgeperezrojas/jorgeperezrojas.github.io/master/common\\_emb.txt](https://raw.githubusercontent.com/jorgeperezrojas/jorgeperezrojas.github.io/master/common_emb.txt)

UMAP

T-SNE

PCA

CUSTOM

X

Component #1

Y

Component #2

Z

Component #3

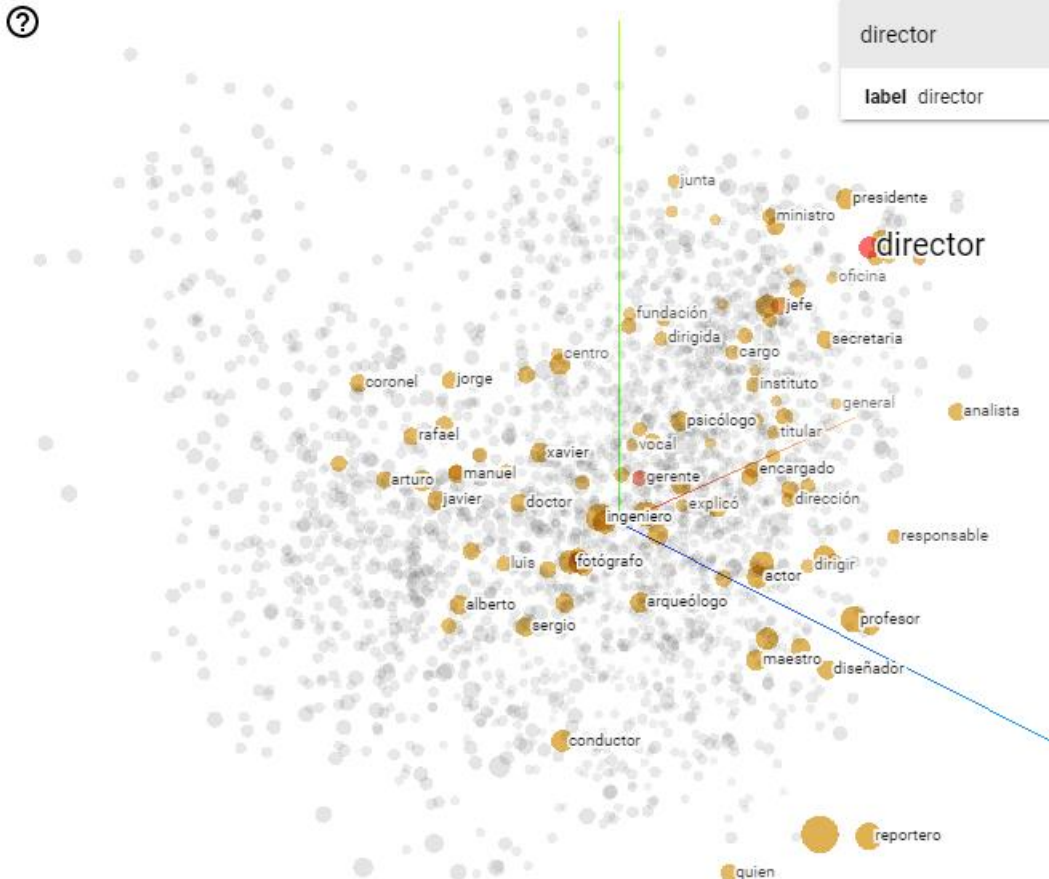
☒

PCA is approximate. ?

Total variance described: 11.2%.



Points: 2810 | Dimension: 300 | Selected 101 points



director

label director

Show All  
DataIsolate 101  
pointsClear  
selection

Search

by

label

neighbors ?  100

distance

COSINE

EUCLIDEAN

Nearest points in the original space:

gerente	0.377
jefe	0.513
ejecutivo	0.529
profesor	0.584
responsable	0.613
presidente	0.624
secretario	0.632
dirección	0.633
dirigir	0.648
analista	0.653
técnico	0.664
instituto	0.677
general	0.681
actor	0.685
titular	0.690
nombrado	0.693
encargado	0.693

BOOKMARKS (0) ?



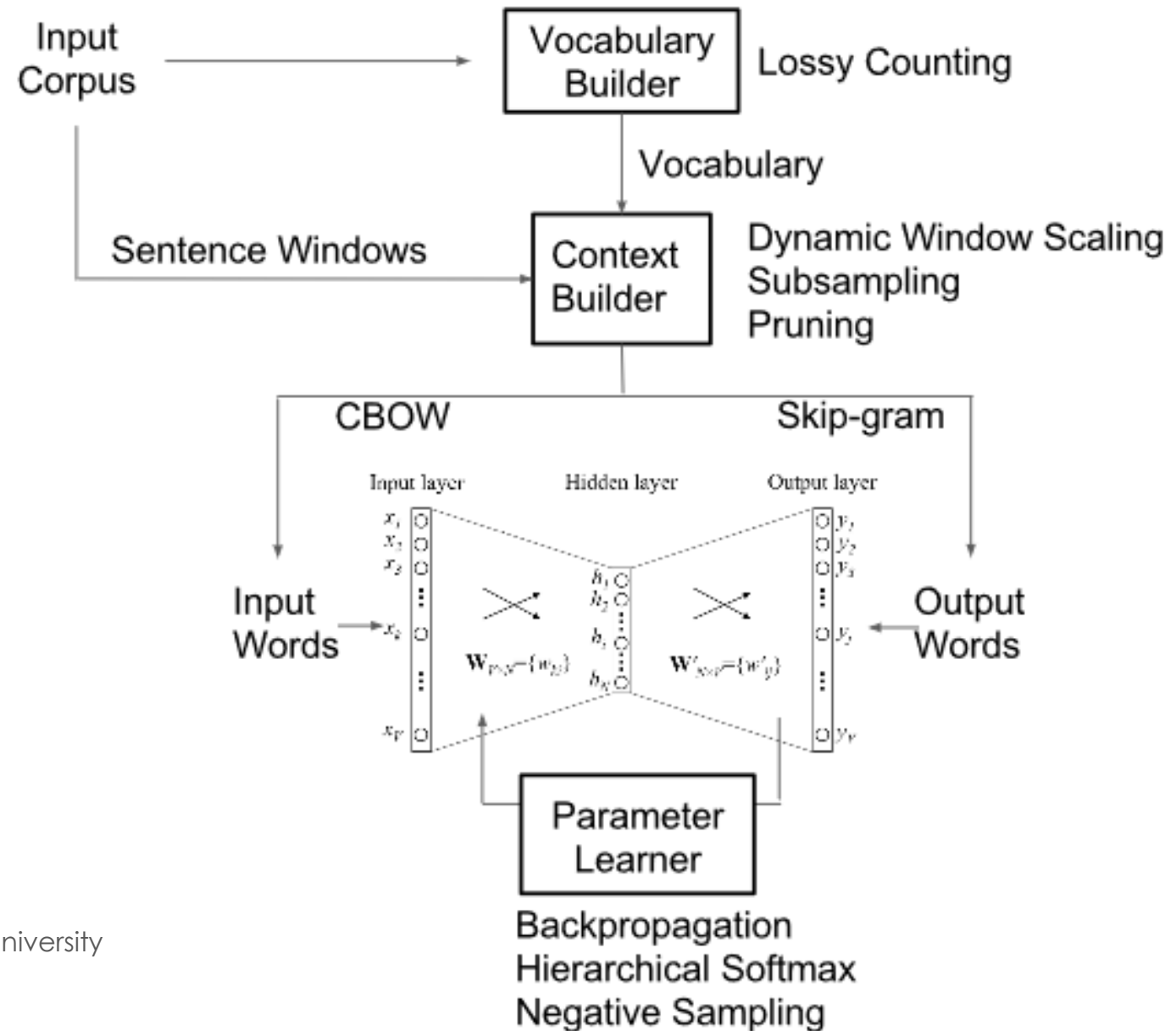


Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

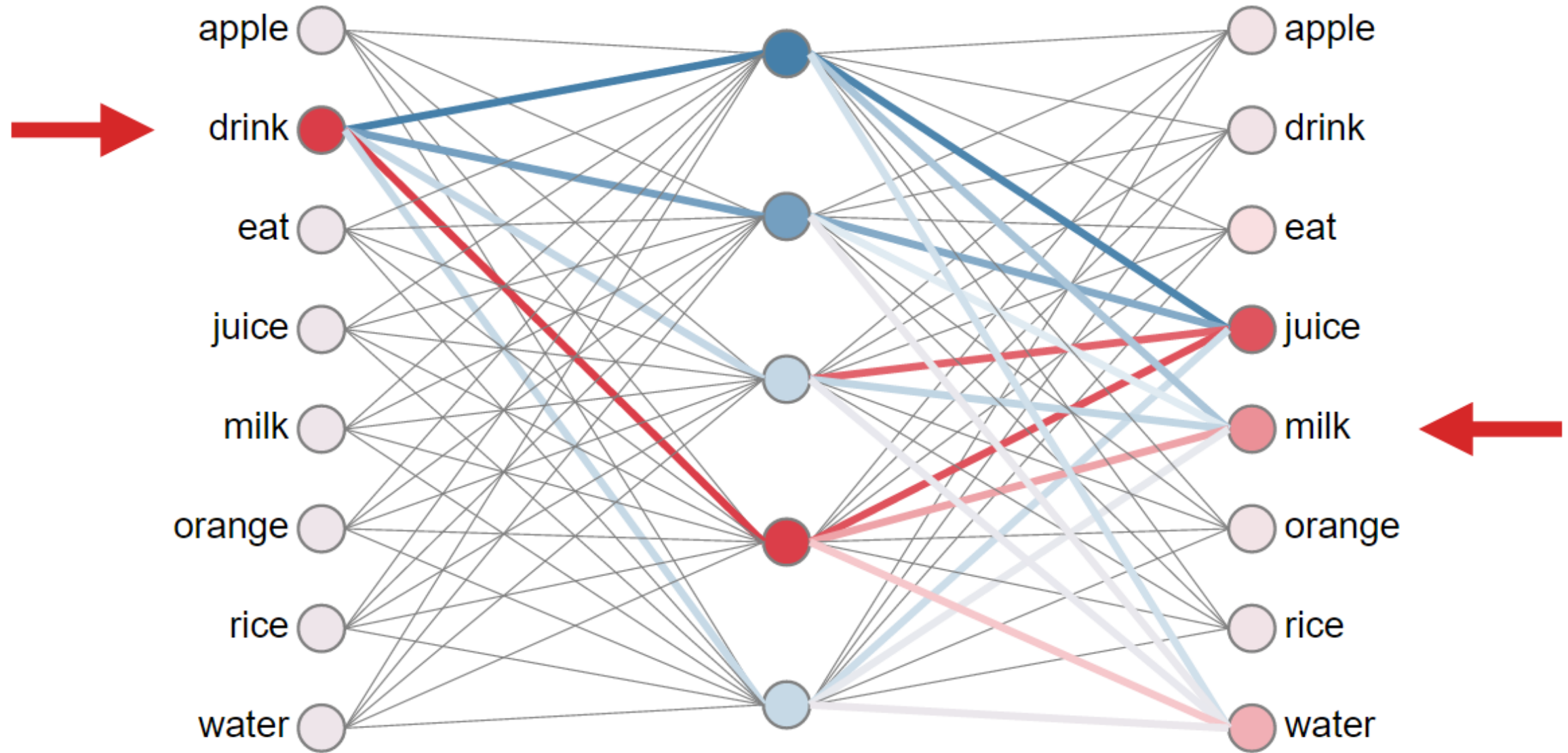


counts	I	like	enjoy	deep	learning	NLP	flying	.
I	0	2	1	0	0	0	0	0
like	2	0	0	1	0	1	0	0
enjoy	1	0	0	0	0	0	1	0
deep	0	1	0	0	1	0	0	0
learning	0	0	0	1	0	0	0	1
NLP	0	1	0	0	0	0	0	1
flying	0	0	1	0	0	0	0	1
.	0	0	0	0	1	1	1	0

# Arquitectura









Word2Vec  
(w2v)  
Google

GloVe  
Stanford  
Univ

ELMo  
AllenNLP

fastText

Gensim

Otros...



HOMBRE → REY

MUJER → ?



REINA

HOMBRE →  
ACTOR

MUJER → ?



ACTRIZ

JUGAR → JUEGA

CANTAR →  
?



CANTA

SANTIAGO →  
CHILE

BOGOTÁ →  
?



COLOMBIA

HOMBRE →  
DOCTOR

MUJER → ?



ENFERMERA

SESGO PROPIO  
DE LOS TEXTOS  
DE REFERENCIA



LUNES, MARTES, MIÉRCOLES, ENERO



ENERO

BLANCO, AZUL, ROJO, CHILE



CHILE

EVERTON, COBRELOA, HUACHIPATO,  
TALCA



TALCA

TALCA, PARIS, LONDRES



TALCA

# EJEMPLO

Pregunta: “¿Cuál es la ley que protege los derechos del consumidor?”

PAUTA

Relación entre proveedores de bienes o servicios y los consumidores está regulada por la Ley N° 19.496 de protección de los derechos de los consumidores. En septiembre del 2018, se publica la Ley N° 21.081, que fortalece la protección de los consumidores, dotando al Servicio con facultad para fiscalizar y mejorando diversos procesos

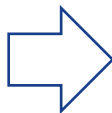
RESPUESTA

Existe una ley que vela por la buena relación entre los clientes y los vendedores de servicios o productos y es la número 19496, llamada “ley de los derechos del consumidor”. Esta se complementa por la ley número 21081, que ahonda en los resguardos, estableciendo formas y capacidades de fiscalización

¿Cuál es la distancia sintáctico-semántica entre ambos textos?

Distancia = 0,35

$0,35 < 0,7$



¡SON EQUIVALENTES! (puntaje completo)



# Interacción con agentes de diálogo

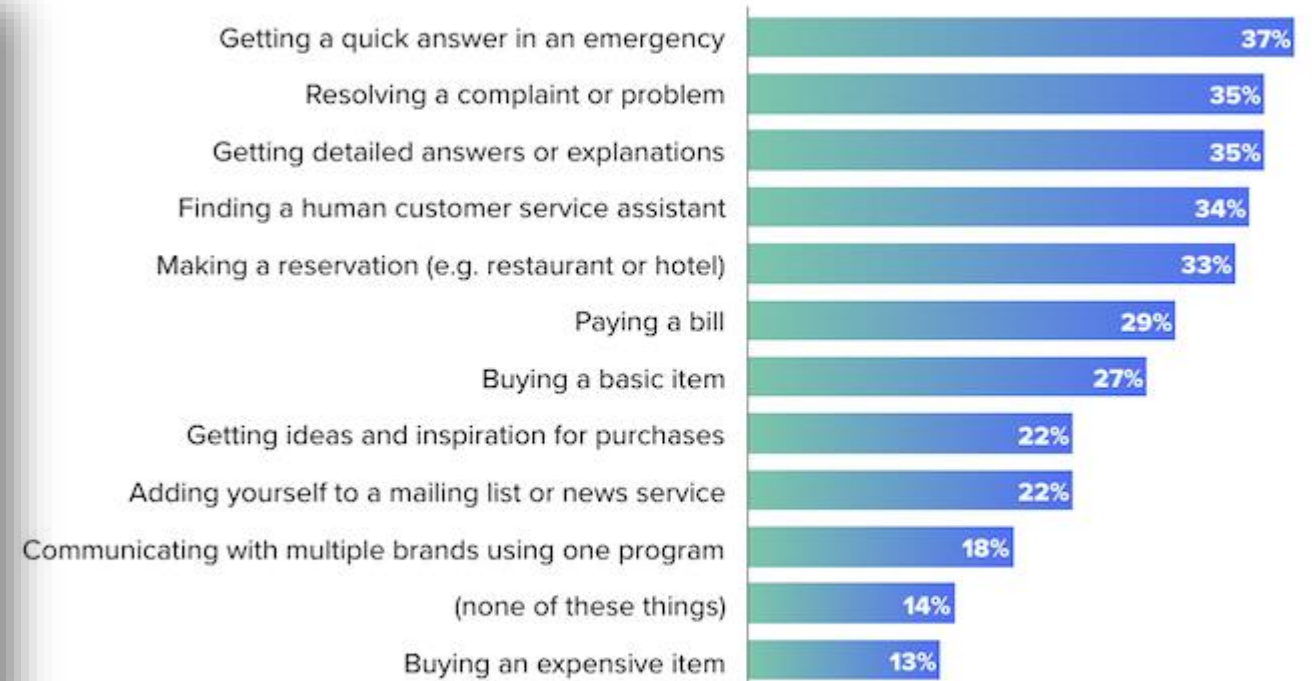
# Chatbots

## Conversational Bots



## Predicted Use Cases for Chatbots

*What do you predict you would use a chatbot for?*



2018 State of Chatbots Report



# ¿Puede un chatbot entender de verdad?

**Un chatbot está entrenado en reconocer palabras y/o términos claves en lo que escribe un usuario y responde con textos pre-estructurados lo que ese usuario – seguramente – espera como respuesta.**

**Un chatbot no entiende las preguntas de verdad.**

**Un chatbot no es capaz de explicar cómo armó una respuesta en particular.**

**Un chatbot no razona sobre las preguntas para responderlas.**



# Entonces, ¿cómo se puede mejorar un bot conversacional?

Lo más importante es lograr interpretar correctamente lo que un usuario está preguntando o diciendo y contestar con texto generado (no necesariamente desde plantillas de respuestas recurrentes), que tenga sentido.

Para ello, la generación de lenguaje natural (NLG) debe ser capaz de armar una secuencia de palabras, que reflejen algún sentido correcto.

Esto se logra con modelos de predicción de palabras en una frase.



# Predicción de la siguiente palabra



el androide



el androide **libre**

Eliminar

el androide **feliz**

el androide **libre xiaomi mi a1**

el androide **libre calibrar bateria**

el androide **verde**

el androide **libre telegram**

el androide **21**

el androide **black**

el androide **libre aplicaciones gratis**

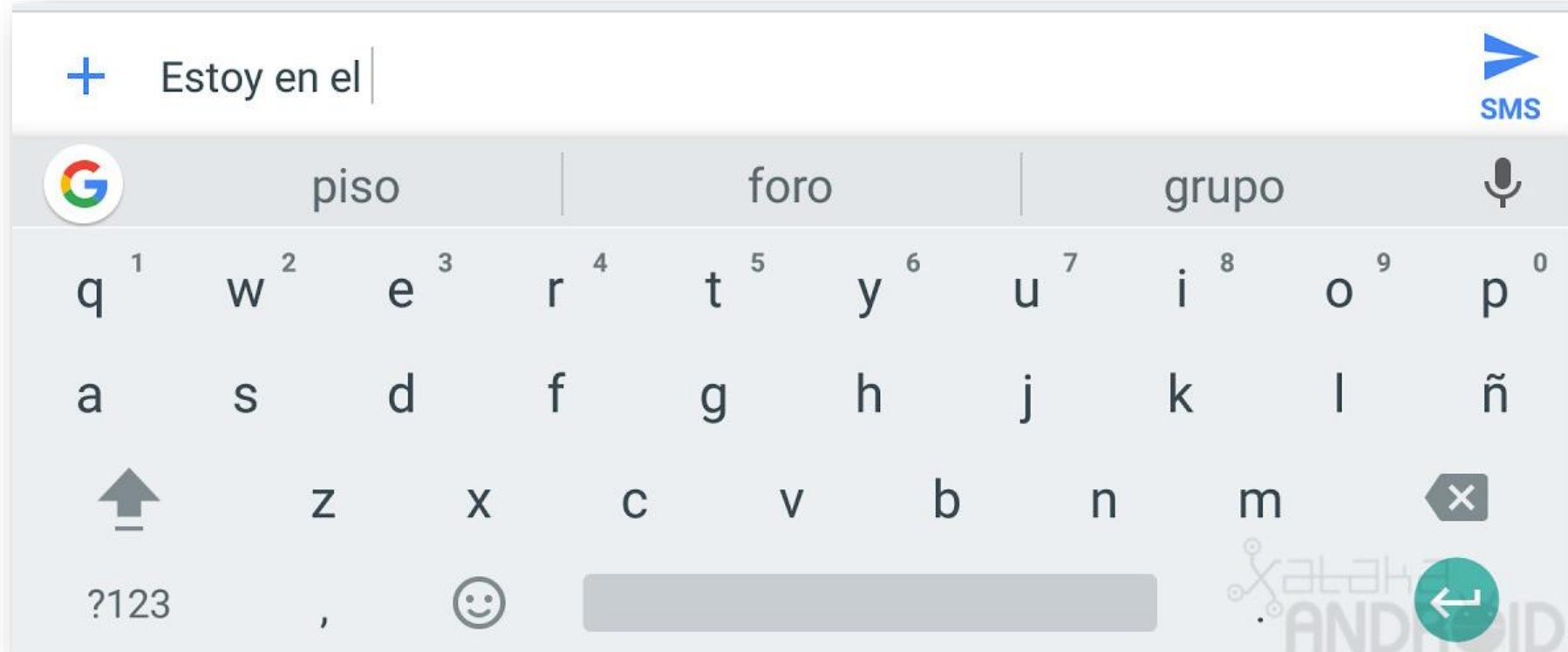
el androide **libre xiaomi**

Buscar con Google

Voy a tener suerte

[Más información](#)

[Denunciar predicciones inadecuadas](#)





# ¿Qué tipo de modelo servirá para predecir la próxima palabra?



Redes Neuronales Recurrentes – RNN  
También conocido como

# MODELAMIENTO SECUENCIAL





Gran potencia hasta la aparición de los Transformers.

El punto central: Modelamiento de secuencia de datos

Dimensión de TIEMPO o de ORDEN está implícitamente presente en el contexto modelado

Está enfocado en el procesamiento de una secuencia de valores  $x^{(1)}, \dots, x^{(t)}$

Manejan secuencias notoriamente largas, pero también, de largo variable.

Para lograrlo, se debe aprovechar el concepto de compartir parámetros entre partes del modelo (modelamiento estadístico y Machine Learning, años 80)

# RNN – En contexto de NLP

Predicción del siguiente elemento:

*Los alumnos abrieron su \_\_\_\_\_ (libro, cuaderno, laptop, mente)*

Más formalmente, dada una secuencia de palabras  $x_1, x_2, \dots, x_n$ , computar la distribución de probabilidad de la siguiente  $x_{t+1}$ :

$$P(x_{t+1} \mid x_t, x_{t-1}, \dots, x_1)$$

Donde  $x_{t+1}$  puede ser cualquier palabra de un vocabulario

$$V = \{w_1, w_2, \dots, w_{|V|}\}$$

Un sistema que hace esto se conoce como un **Modelo de Lenguaje**.

# Modelamiento de Lenguaje

Más simple: un **Modelo de Lenguaje** es un sistema que le asigna la probabilidad a una porción de texto.

Por ej si se tiene un texto  $x_1, x_2, \dots, x_T$  entonces la probabilidad de este texto (de acuerdo al **Modelo de Lenguaje**) es:

$$P(x_1, x_2, \dots, x_T) = P(x_1) * P(x_2 | x_1) * \dots * P(x_T | x_{T-1}, \dots, x_1)$$

$$= \prod_{t=1}^T P(x_t | x_{t-1}, \dots, x_1) \quad \rightarrow \text{Lo entrega el Modelo de Lenguaje}$$

# Modelos de Lenguaje de n-gramas

- Los estudiantes abrieron su cuaderno
- **Pregunta:** ¿cómo aprender un Modelo de Lenguaje?
- **Respuesta** (antes de las redes profundas): aprender un Modelo de Lenguaje de n-gramas.
- **Definición:** un n-grama es un grupo de palabras consecutivas:
  - Unigramas:  
“Los”, “estudiantes”, “abrieron”, “su” “cuaderno”
  - Bigramas:  
“Los estudiantes”, “estudiantes abrieron”, “abrieron su”, “su cuaderno”
  - Trigramas:  
“Los estudiantes abrieron”, “estudiantes abrieron sus” , “abrieron su cuaderno”
  - 4-gramas:  
“Los estudiantes abrieron su”, “estudiantes abrieron su cuaderno”



# Modelos de Lenguaje de n-gramas

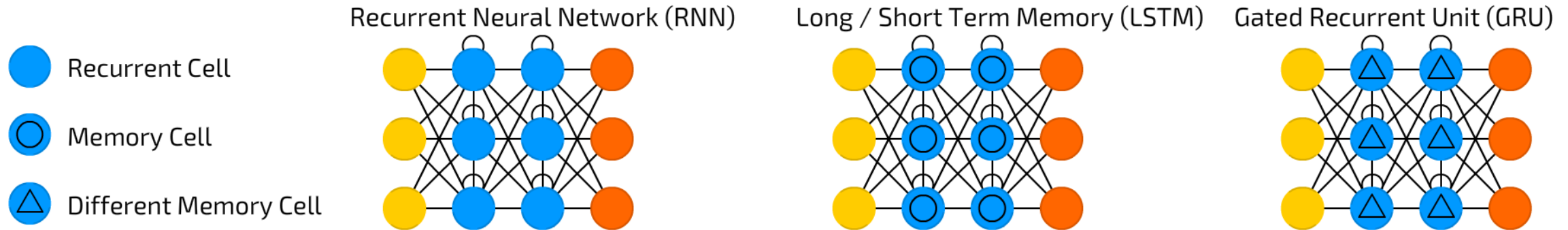
- Entonces, recolectar estadísticas de la frecuencia de los diferentes n-gramas y usarlos para predecir la siguiente palabra.
- Esto sirve también para generar texto: predecir las siguientes palabras/frases.
- Estos Modelos de Lenguaje se pueden construir con **Redes Profundas**

**RNN: Redes Neuronales Recurrentes**

**LSTM: Long-Short Term Memory Networks**

**GRU: Gated Recurrent Units**

# Modelos de Redes con Memoria/Secuencia



Estos modelos trabajan con celdas (o células) de memoria y de recurrencia. En la práctica, hay capas de entrada, de salida, y escondidas, siendo éstas últimas las que funcionan diferente en cada uno de estos 3 modelos. En particular, las neuronas recurrentes tienen conexiones entre la misma capa y con eso, son capaces de procesar secuencias. Esto ocurre al pasar estados intermedios o escondidos, que se asocian a *timesteps*, los que se combinan los valores de los nodos de entrada.

(Más detalles, en anexos)



# Gracias



**rsandova@ing.puc.cl**  
rodrigo@RSolver.com



**@RSandovalSolver**



**/in/RodrigoSandoval**

**www.RodrigoSandoval.net**  
**www.RSolver.com**





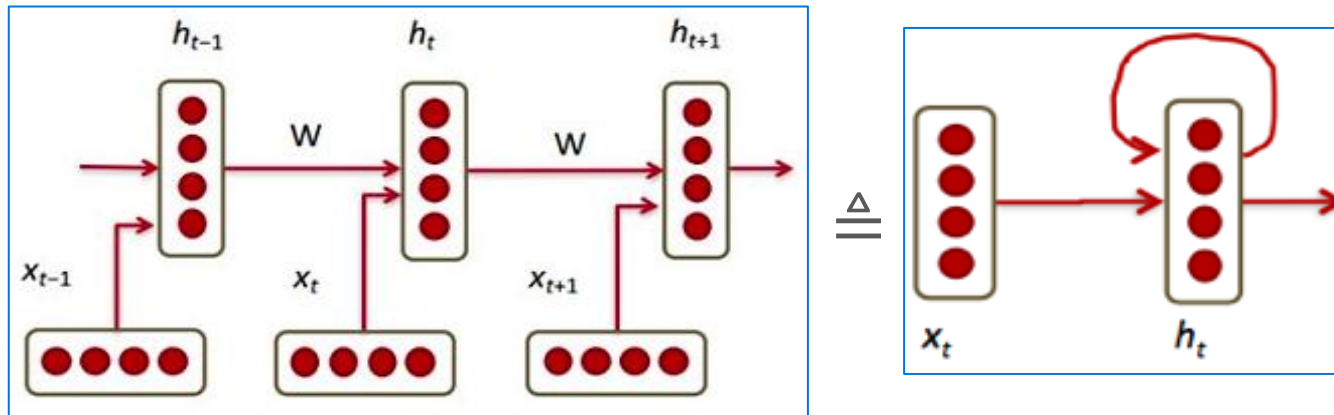
# Anexo: RNNs, GRU, LSTM

# Recurrent Neural Networks (RNNs)

La principal idea para RNN en texto es:

Condición en **todas las palabras previas**

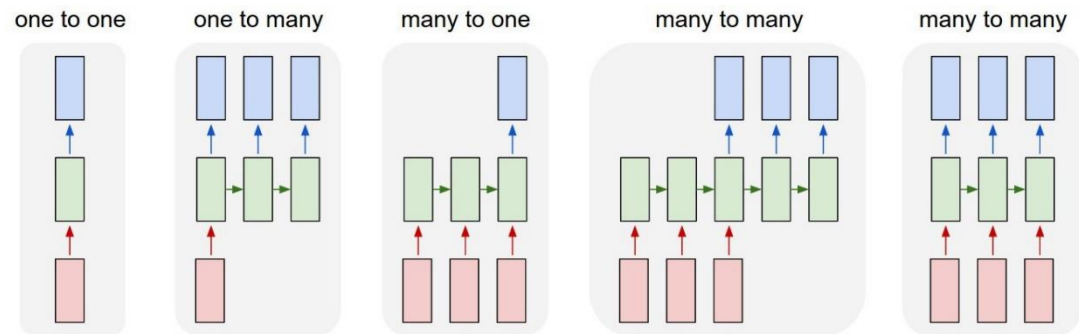
Usar el mismo set de pesos en todos los pasos de tiempo  $h_t = \sigma(W^{(hh)}h_{t-1} + W^{(hx)}x_t)$



<https://pbs.twimg.com/media/C2j-8j5UsAACgEK.jpg>

😊 Apilarlos, bien!

😞 Vanishing gradient problem



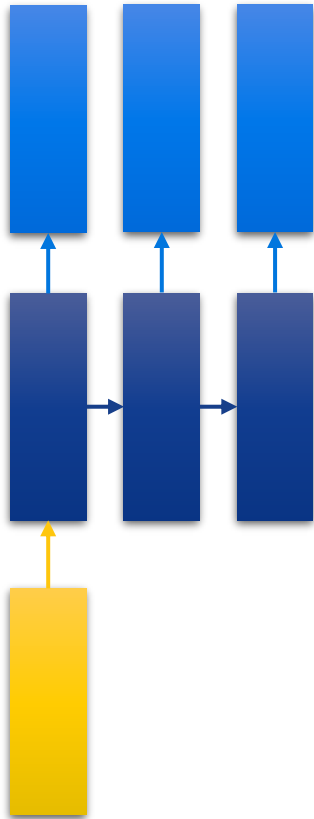
<https://discuss.pytorch.org/uploads/default/original/1X/6415da0424dd66f2f5b134709b92baa59e604c55.jpg>

# RNN: Secuencias de Procesos

Uno  
a uno

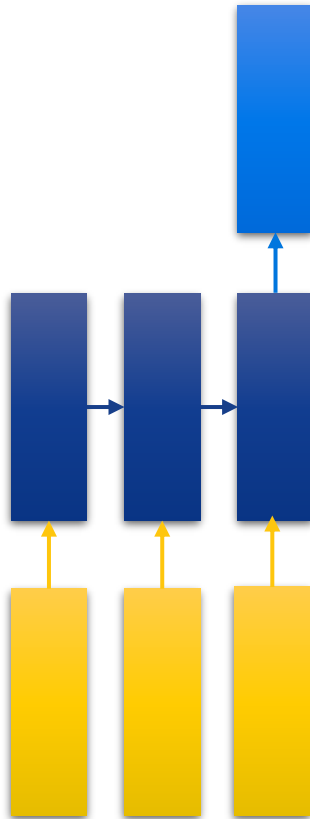


Uno a varios



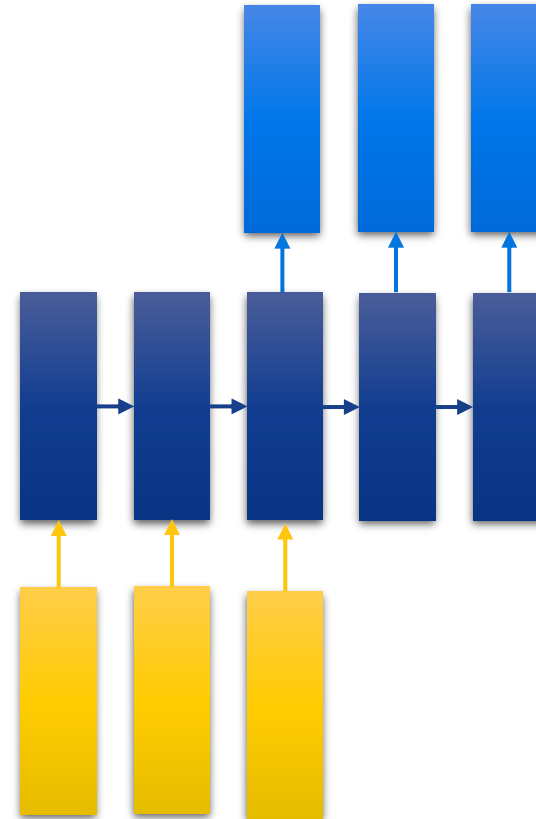
Ejemplo: Etiquetado  
de imágenes  
Imagen →  
**secuencia** de  
palabras

Varios a uno



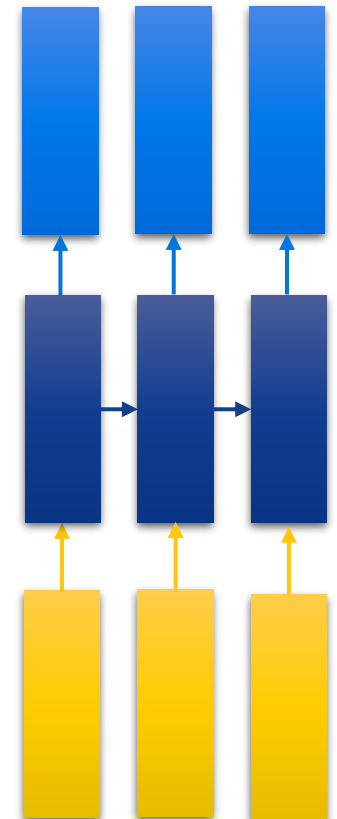
Ejemplo: Análisis de  
Sentimiento  
Secuencia de  
palabras → una clase

Varios a varios



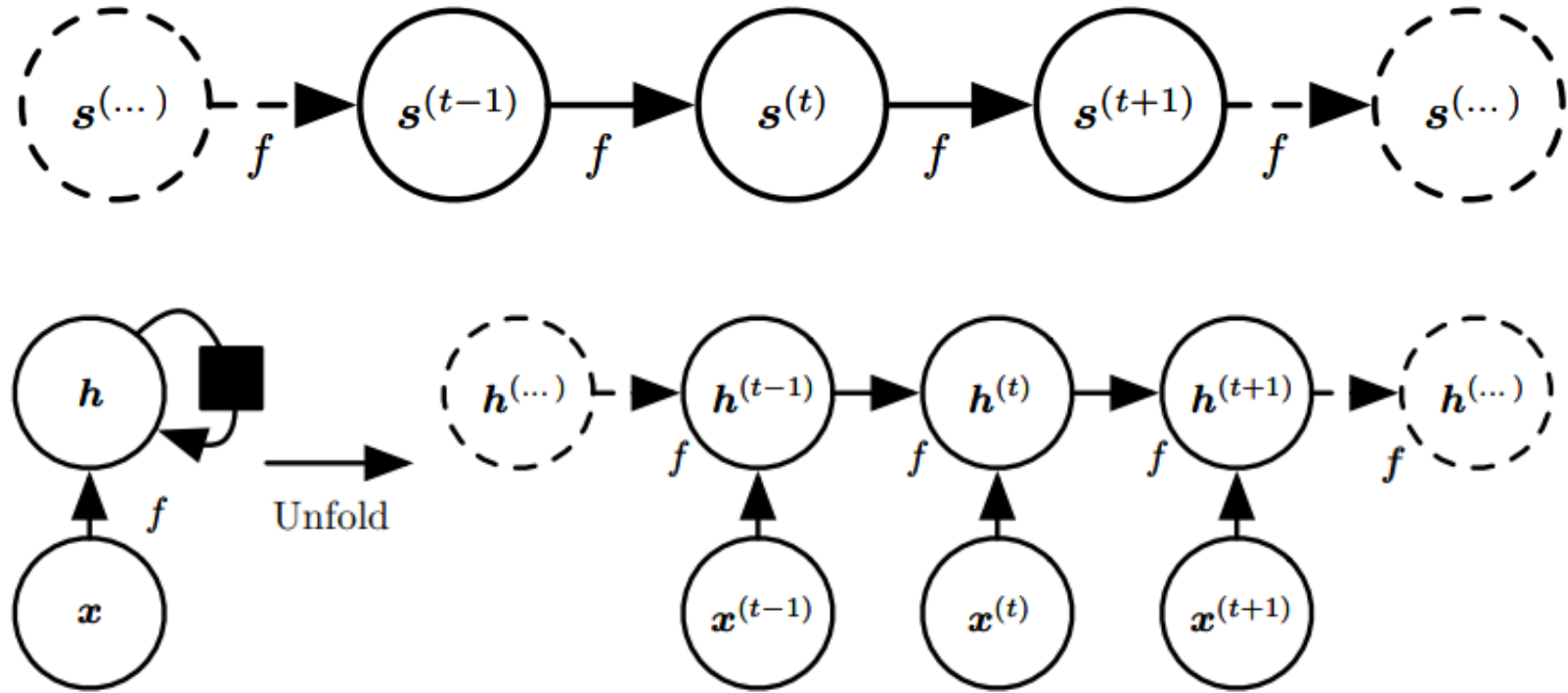
Ejemplo: Traducción  
de Máquina  
Secuencia palabras  
→ secuencia  
palabras

Varios a varios



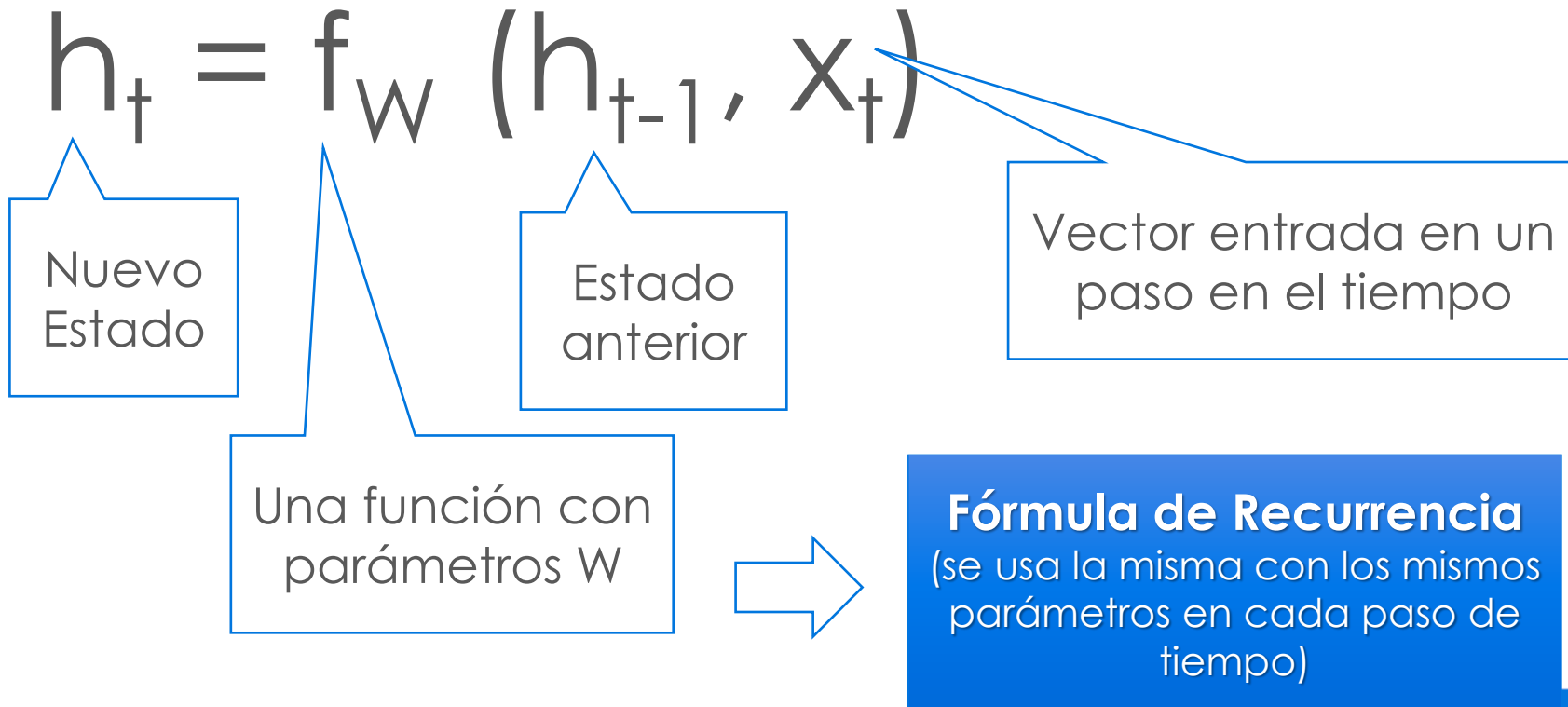
Ejemplo: Clasificación  
de video a nivel de  
frame

- Sistemas dinámicos clásicos
  - Cada nodo es un estado en instante  $t$ .
- Unfolding comp graphs. RNN sin salida



Se procesa la entrada  $x$  incorporándola al estado  $h$  pasado hacia adelante en el tiempo

Se puede procesar una secuencia de vectores  $x$  al aplicar una fórmula de recurrencia en cada paso en el tiempo:

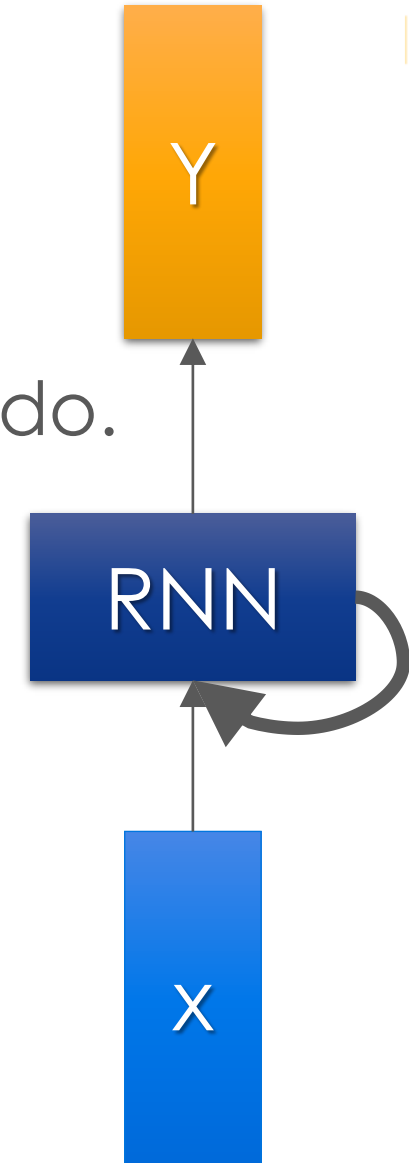


El estado consiste en un vector  $\mathbf{h}$  único, escondido.

$$\mathbf{h}_t = f_w (\mathbf{h}_{t-1}, \mathbf{x}_t)$$

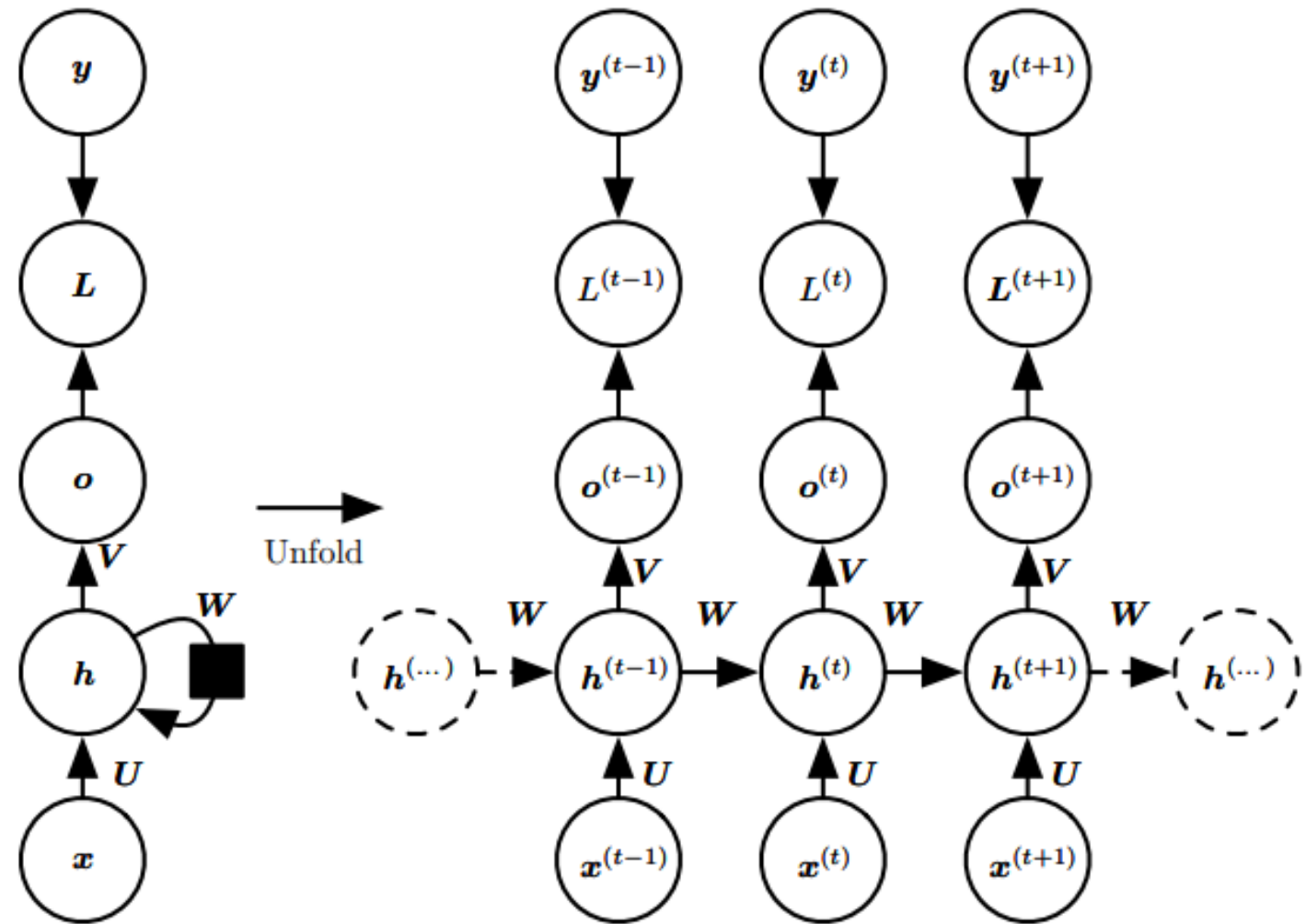
$$\mathbf{h}_t = \tanh (W_{hh} \mathbf{h}_{t-1} + W_{xh} \mathbf{x}_t)$$

$$\mathbf{y}_t = W_{hy} \mathbf{h}_t$$



# Recurrent Hidden Units

- Grafo computacional para calcular la pérdida en entrenamiento de una RNN que mapea una secuencia de valores  $\mathbf{x}$ , correspondiente a una secuencia de salida  $\mathbf{o}$  (salida por softmax)
- Pérdida:  $L$
- Conexiones escondidas:  $h$ .
- Pesos:  $W$

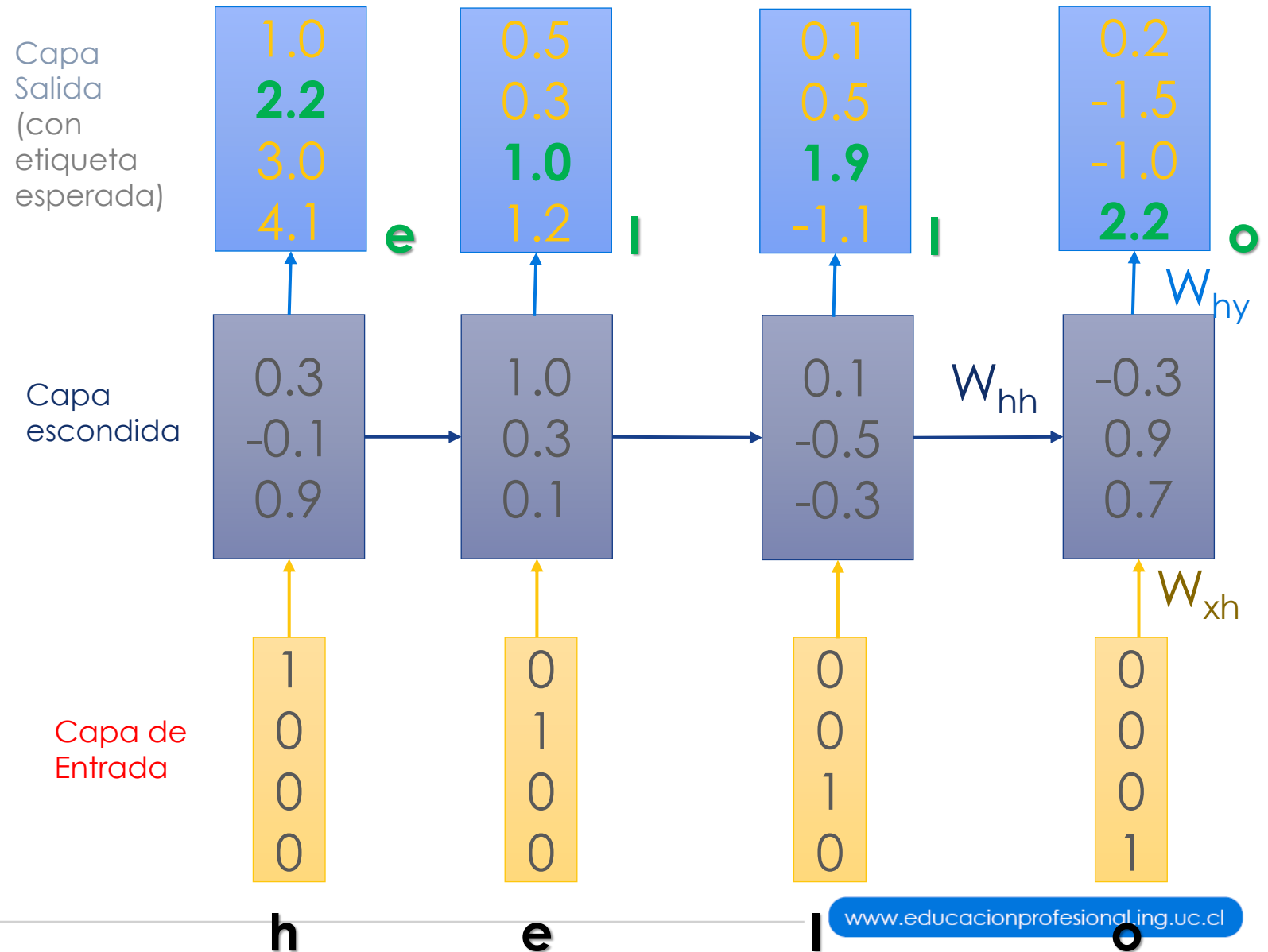






# Ejemplo: Modelo de lenguaje a nivel de caracteres

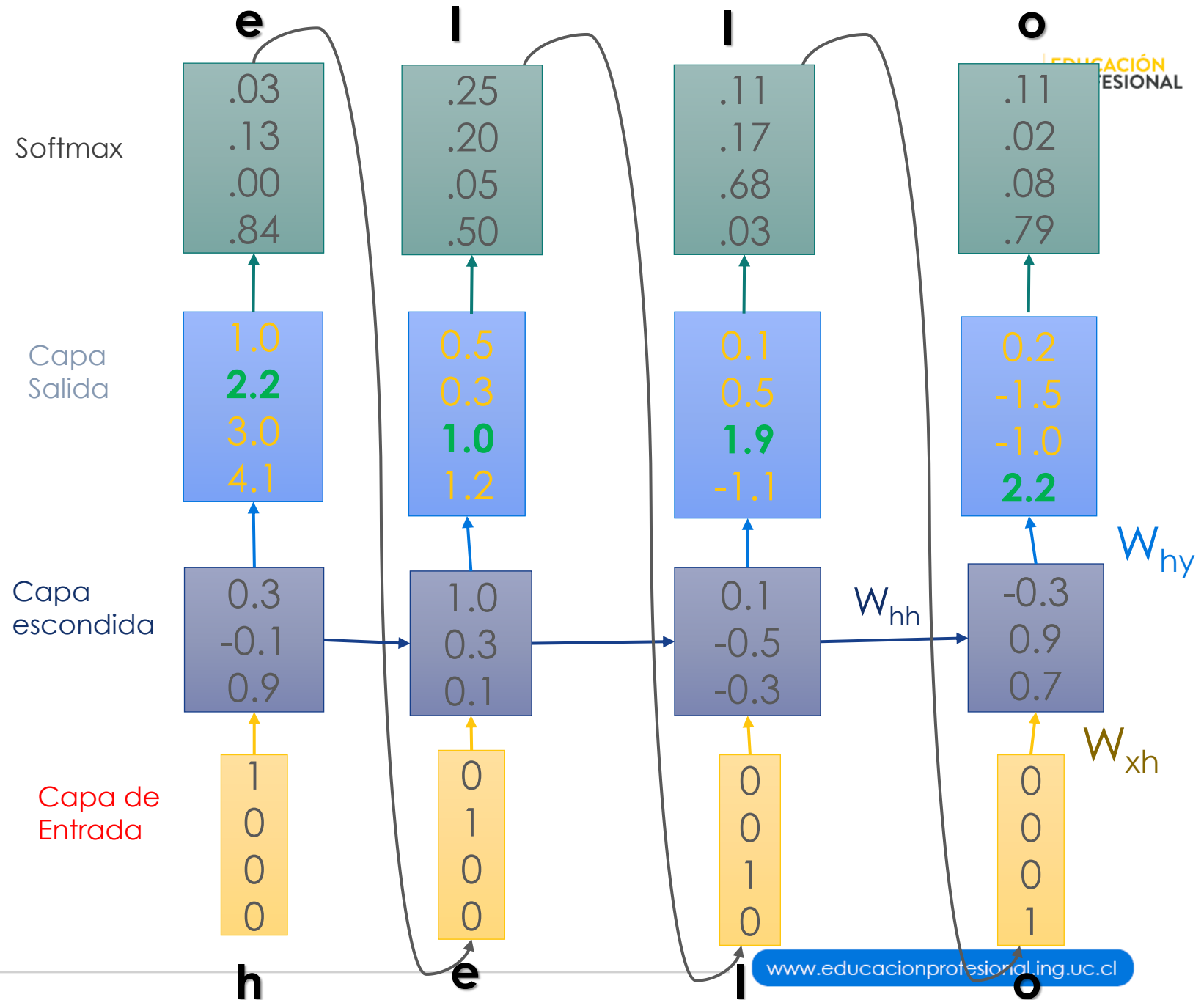
- Vocabulario:  
[h, e, l, o]
- Secuencia de ejemplo para entrenamiento:  
"hello"





## Ejemplo: Modelo de lenguaje a nivel de caracteres

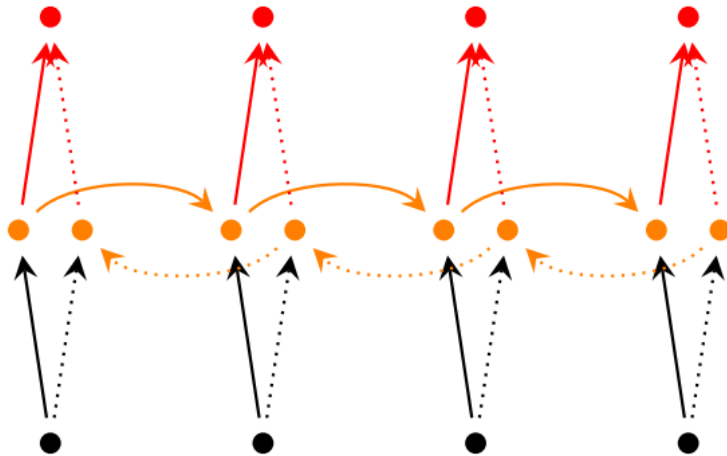
- Vocabulario: [h, e, l, o]
- Secuencia de ejemplo para entrenamiento: "hello"
- Se retroalimentan los caracteres, uno a la vez, de vuelta al modelo



# RNNs bidireccionales

Idea principal: incorporar contexto a la izquierda y a la derecha

La salida puede depender de los elementos anteriores y de los posteriores



$$\vec{h}_t = \sigma(\vec{W}^{(hh)}\vec{h}_{t-1} + \vec{W}^{(hx)}x_t)$$

$$\overleftarrow{h}_t = \sigma(\overleftarrow{W}^{(hh)}\overleftarrow{h}_{t+1} + \overleftarrow{W}^{(hx)}x_t)$$

$$y_t = f([\vec{h}_t; \overleftarrow{h}_t])$$

Pasado y futuro alrededor del token o palabra

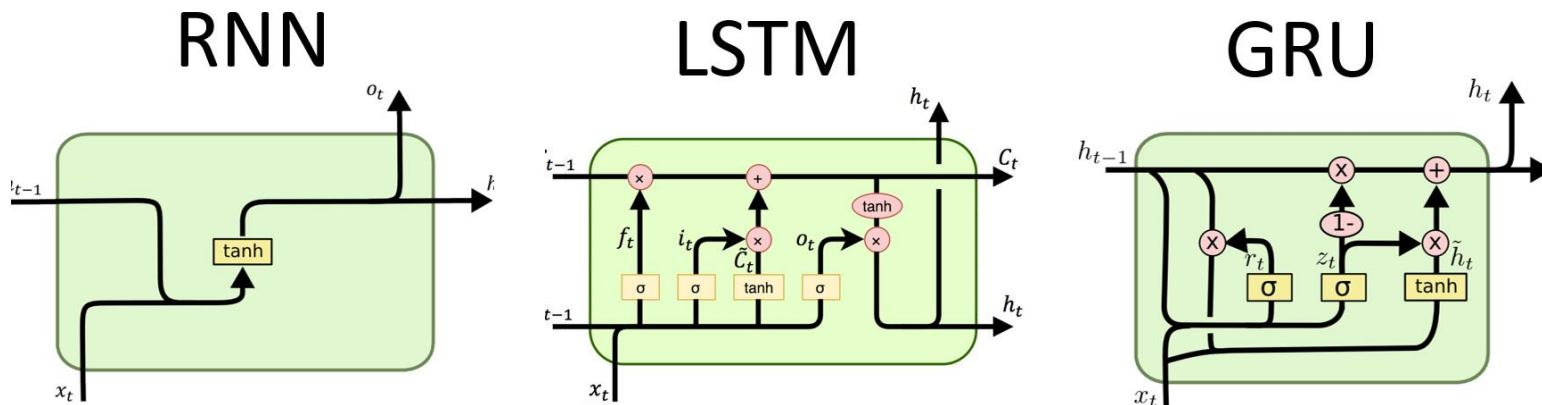
<http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/>

Dos RNNs apiladas una sobre la otra

La salida se computa en base al estado escondido de ambas RNNs  $[\vec{h}_t; \overleftarrow{h}_t]$

# GRU: Predicción de secuencias

- Se mantiene el concepto de las unidades recurrentes, igual que en RNN (y LSTM).
- A diferencia de las RNN, que usan un estado escondido para recordar información, las LSTM y GRU usan puertas (gates) para controlar qué se recuerda y qué se olvida antes de actualizar el estado escondido.
  - La LSTM usa un estado de celda para la memoria de largo plazo.
- GRU y LSTM son similares, pero la primera usa menos puertas.



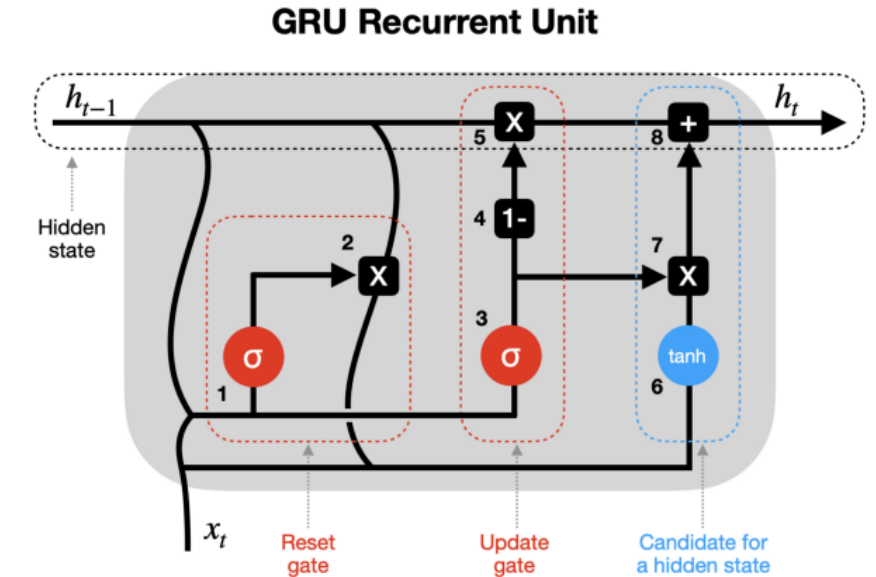
# Proceso en la GRU

**1-2 Reset gate.** Se combina el estado escondido  $h_{t-1}$  y la nueva entrada  $x_t$  y pasan por la *reset gate*. Aquí se aplica una función sigmoide determinando lo que se recuerda (1) y lo que se descarta (0) o se mantiene parcialmente (rango intermedio). El paso 2 restablece el estado escondido multiplicándolo con salidas del paso 1.

**3-4-5 Update gate.** Paso 3 es análogo al 1, pero los pesos y biases usados para escalar son diferentes. Luego de pasar el vector combinado por la sigmoide se resta del vector con 1s (paso 4) y se multiplica por el estado escondido anterior (paso 5).

**6-7-8 Estado escondido candidato.** Luego del paso 2, se combinan las salidas con las nuevas entradas  $x_t$  ponderándolas por sus pesos y agregando bias antes de pasar por una activación  $\tanh$  (paso 6). Entonces el estado escondido candidato se multiplica por los resultados de una update gate (paso 7) y sumado al  $h_{t-1}$  modificado.

Luego se repite para  $t+1$  hasta procesar la secuencia completa.



$h_{t-1}$  - hidden state at previous timestep  $t-1$  (memory)

$x_t$  - input vector at current timestep  $t$

$h_t$  - hidden state at current timestep  $t$

**X** - vector pointwise multiplication    **+** - vector pointwise addition

**tanh** - tanh activation function

**$\sigma$**  - sigmoid activation function

**⌢** - concatenation of vectors

**---** - states

**---** - gates

**---** - updates

Fuente: <https://towardsdatascience.com/gru-recurrent-neural-networks-a-smart-way-to-predict-sequences-in-python-80864e4fe9f6>