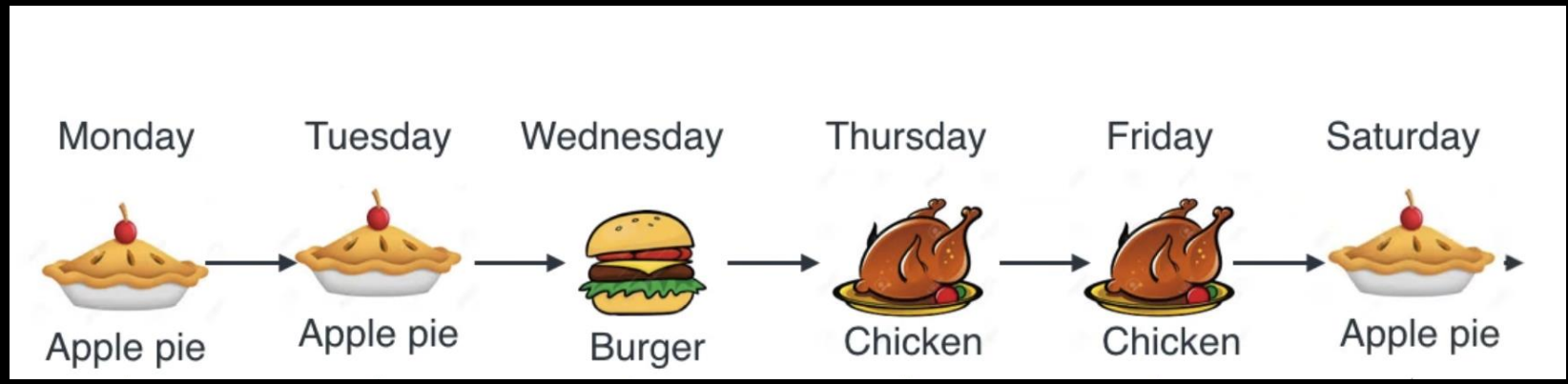
Redes Recurrentes

Las redes neuronales recurrentes (RNN) son un tipo de red neuronal diseñada para procesar datos secuenciales.

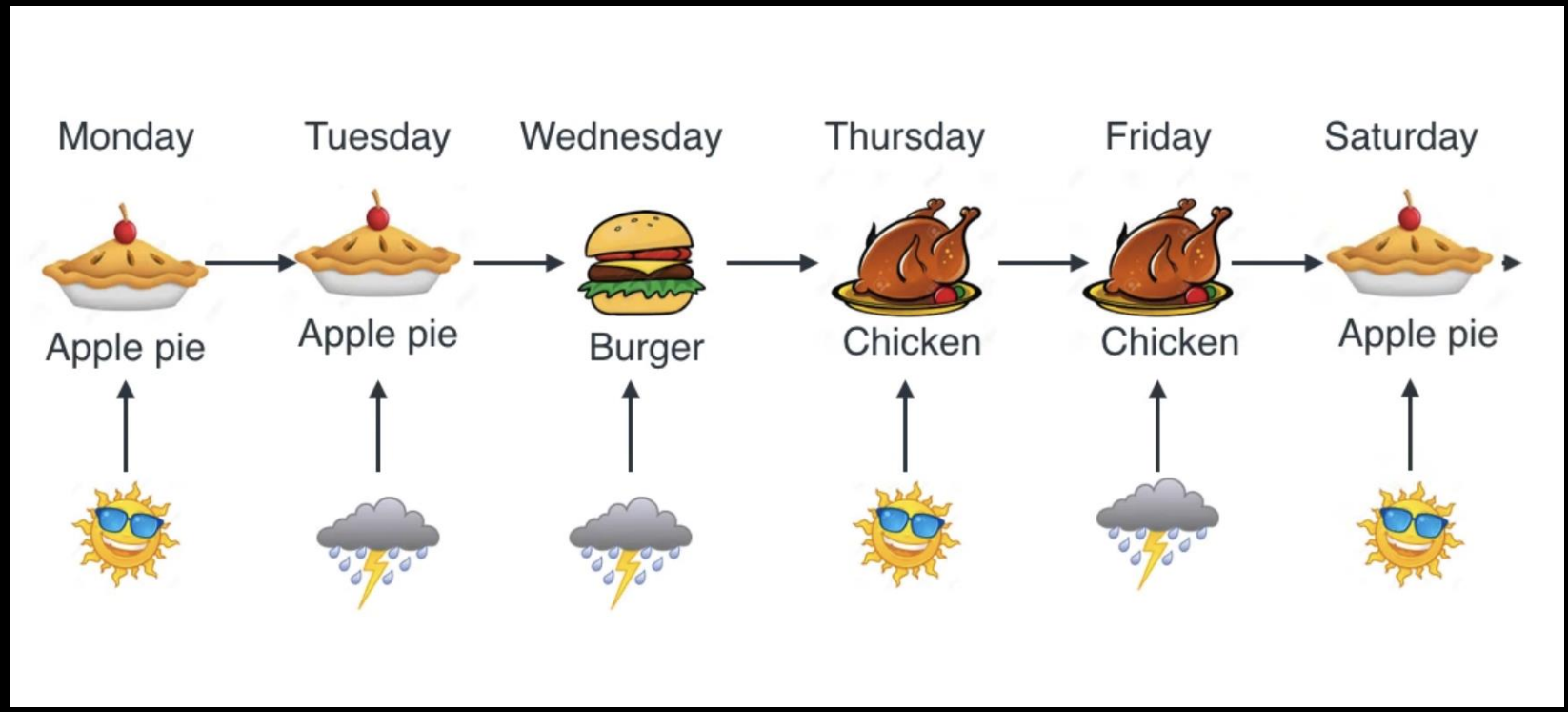
Ideal para problemas en los que la información de estados previos influye en la salida actual.

Permiten modelar dependencias temporales en series de tiempo, lenguaje y otras secuencias.

Como se decide el menu del dia?



Como se decide el menu del dia?



Por suerte a alguien se le ocurrio registrar esto

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

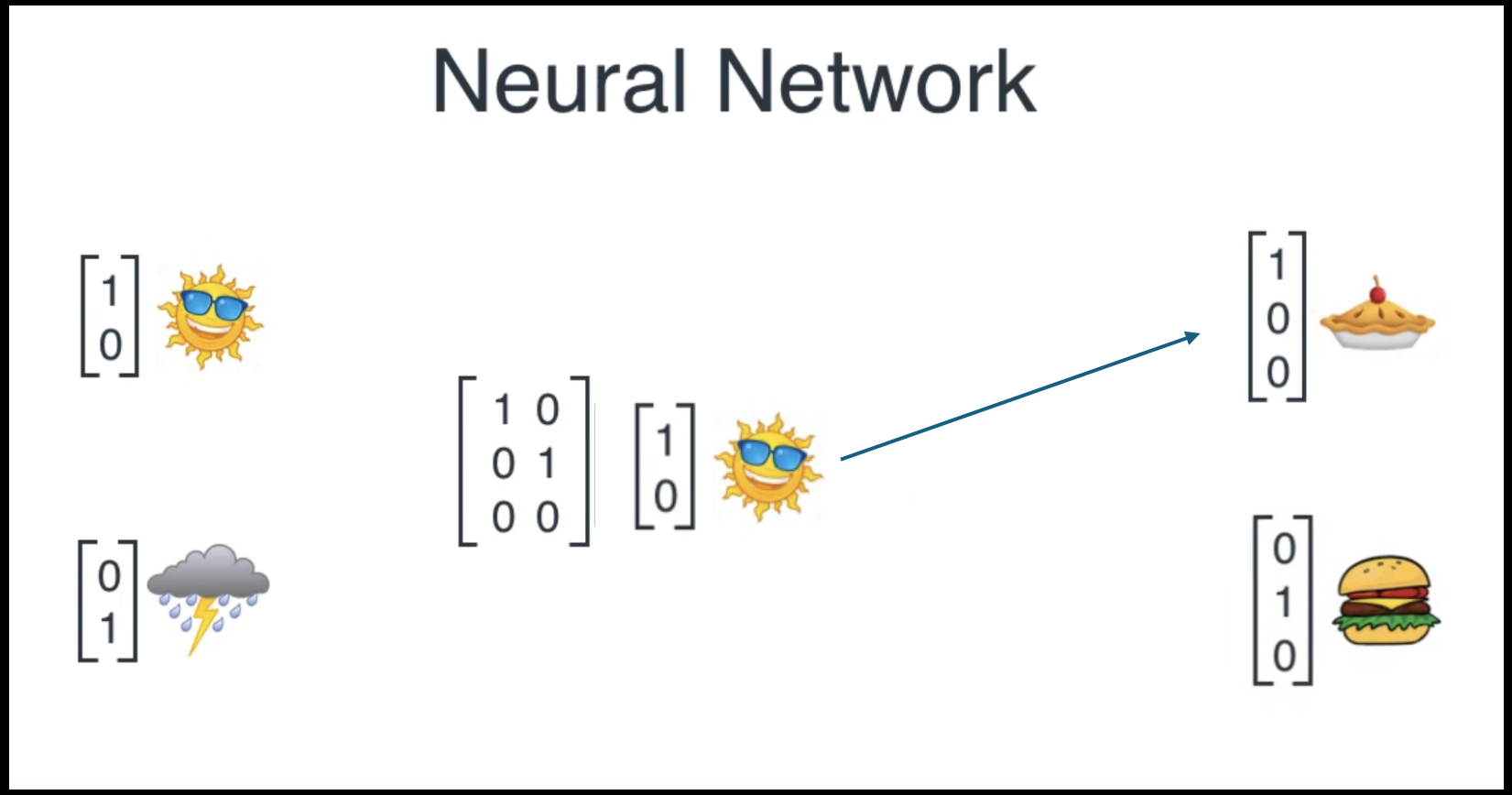
$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Supongamos una preimera version del esquema de las cartas, muy sencilla

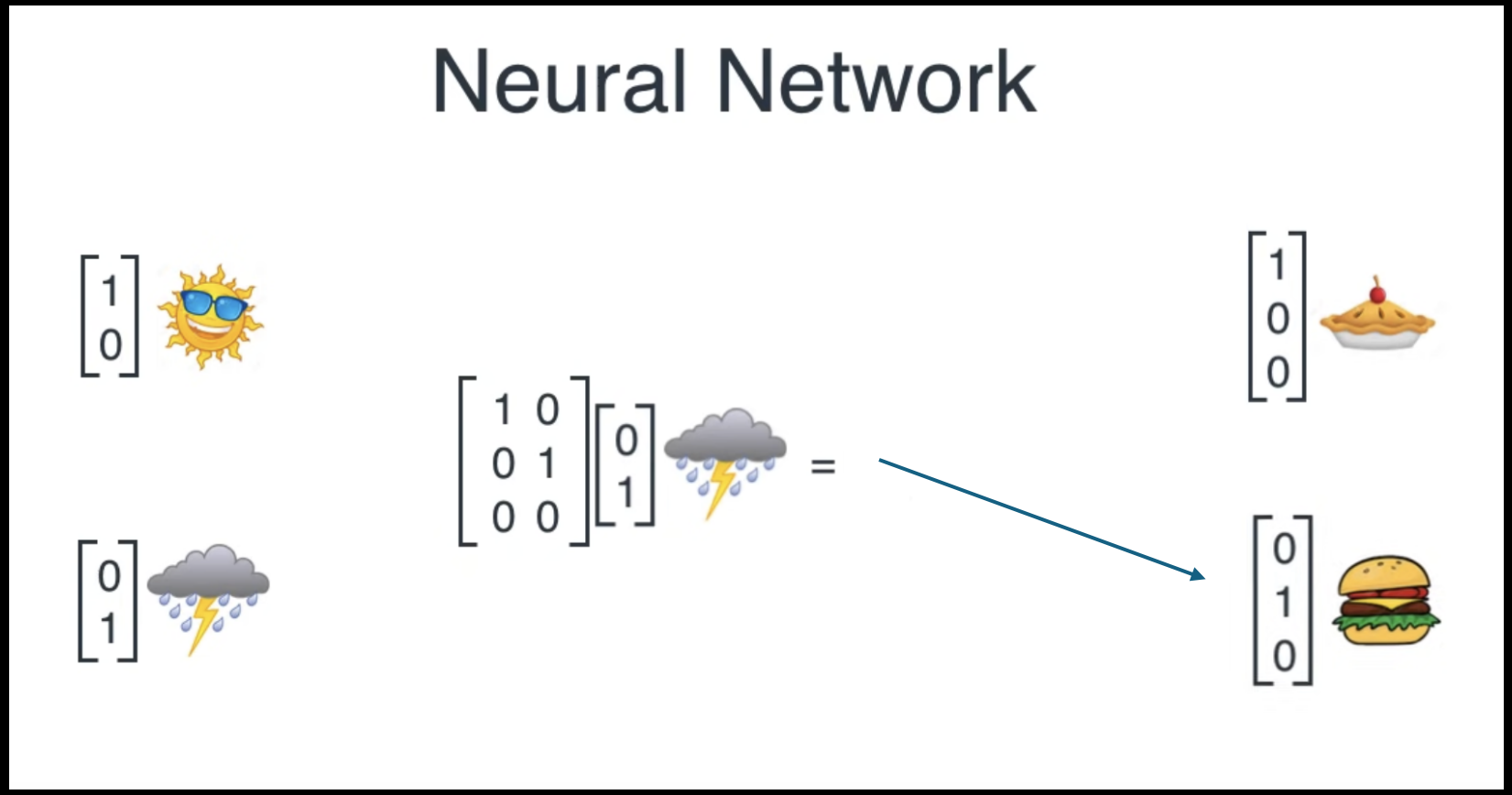
Regla: soleado, contento, torta

nublado, hamburguesa

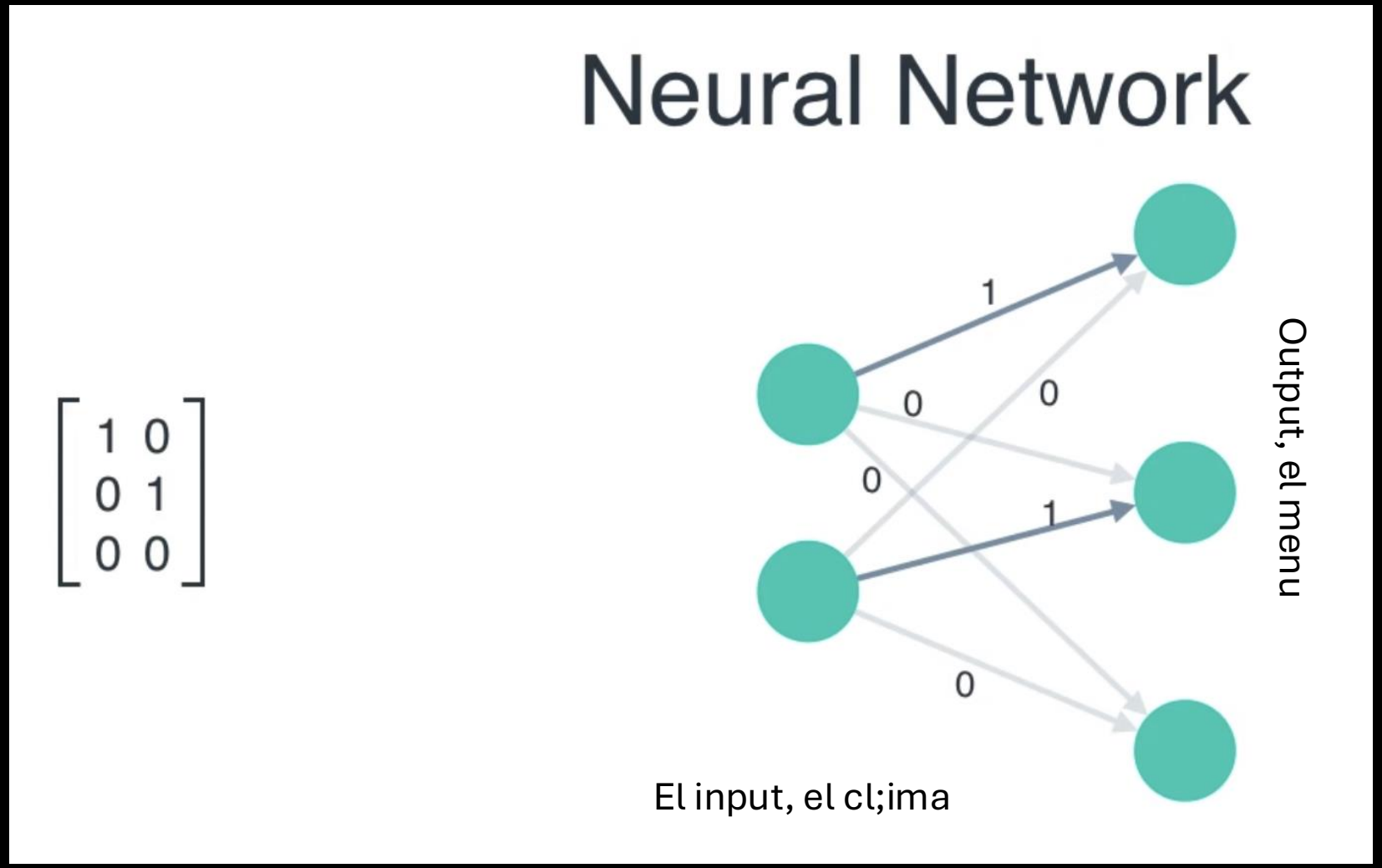
Matematica de la regla



Matematica de la regla



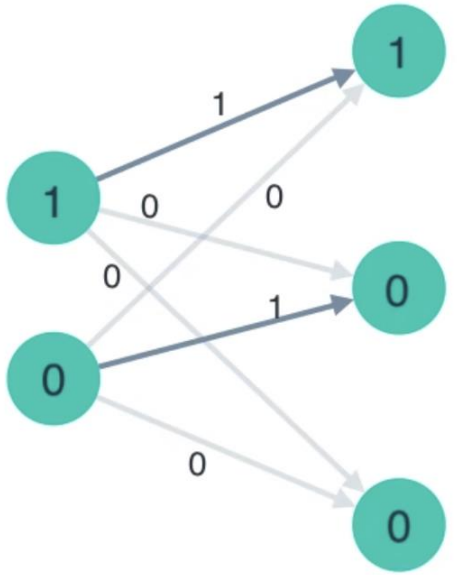
La matematica, escrita como una matriz



Neural Network

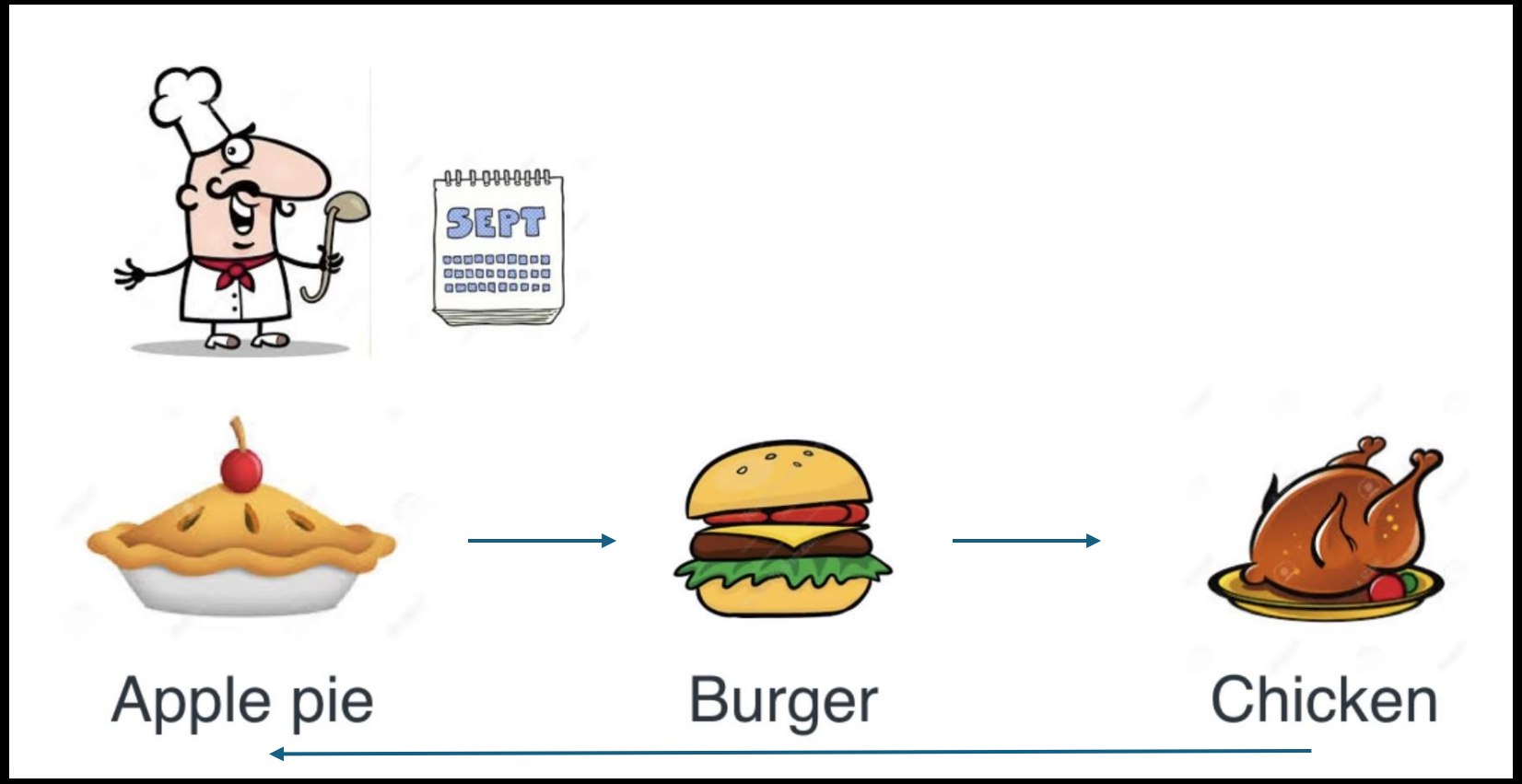
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$



$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

El cocinero TOC



Simple Neural Network

$$\begin{bmatrix}1\\0\\0\end{bmatrix}$$ 

$$\begin{bmatrix}0\\1\\0\end{bmatrix}$$ 

$$\begin{bmatrix}0\\0\\1\end{bmatrix}$$ 

$$\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}\begin{bmatrix}1\\0\\0\end{bmatrix}$$  $= \begin{bmatrix}0\\1\\0\end{bmatrix}$ 

Simple Neural Network

$$\begin{bmatrix}1\\0\\0\end{bmatrix}$$ 

$$\begin{bmatrix}0\\1\\0\end{bmatrix}$$ 

$$\begin{bmatrix}0\\0\\1\end{bmatrix}$$ 

$$\begin{bmatrix}0 & 0 & 1\\1 & 0 & 0\\0 & 1 & 0\end{bmatrix}$$ $$\begin{bmatrix}0\\0\\1\end{bmatrix}$$  $$=$$ $$\begin{bmatrix}1\\0\\0\end{bmatrix}$$ 

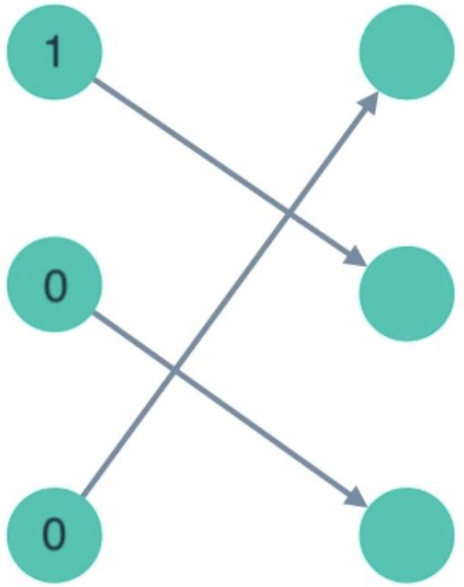
$$\begin{bmatrix}0\\1\\0\end{bmatrix}$$ 

$$\begin{bmatrix}0\\0\\1\end{bmatrix}$$ 

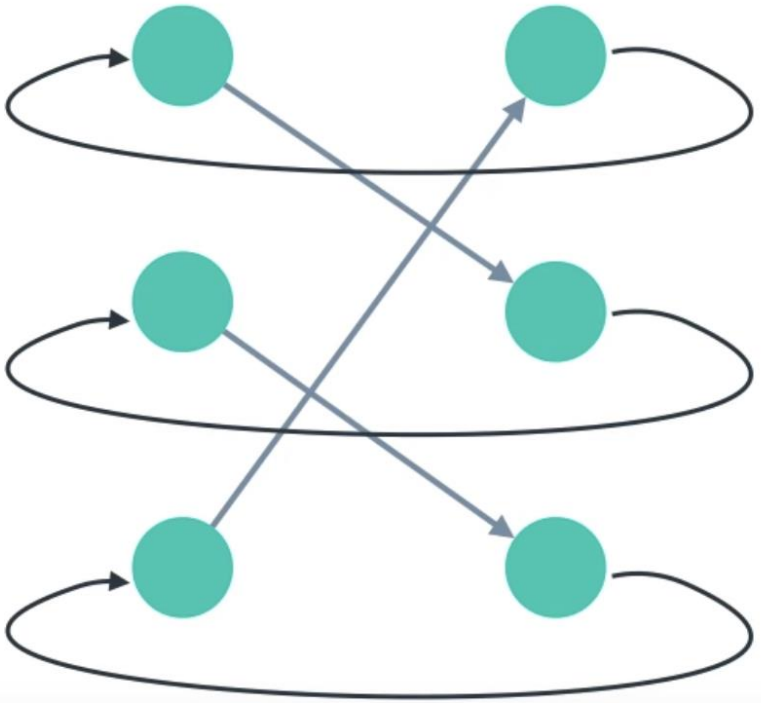
Simple (Recurrent) Neural Network

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

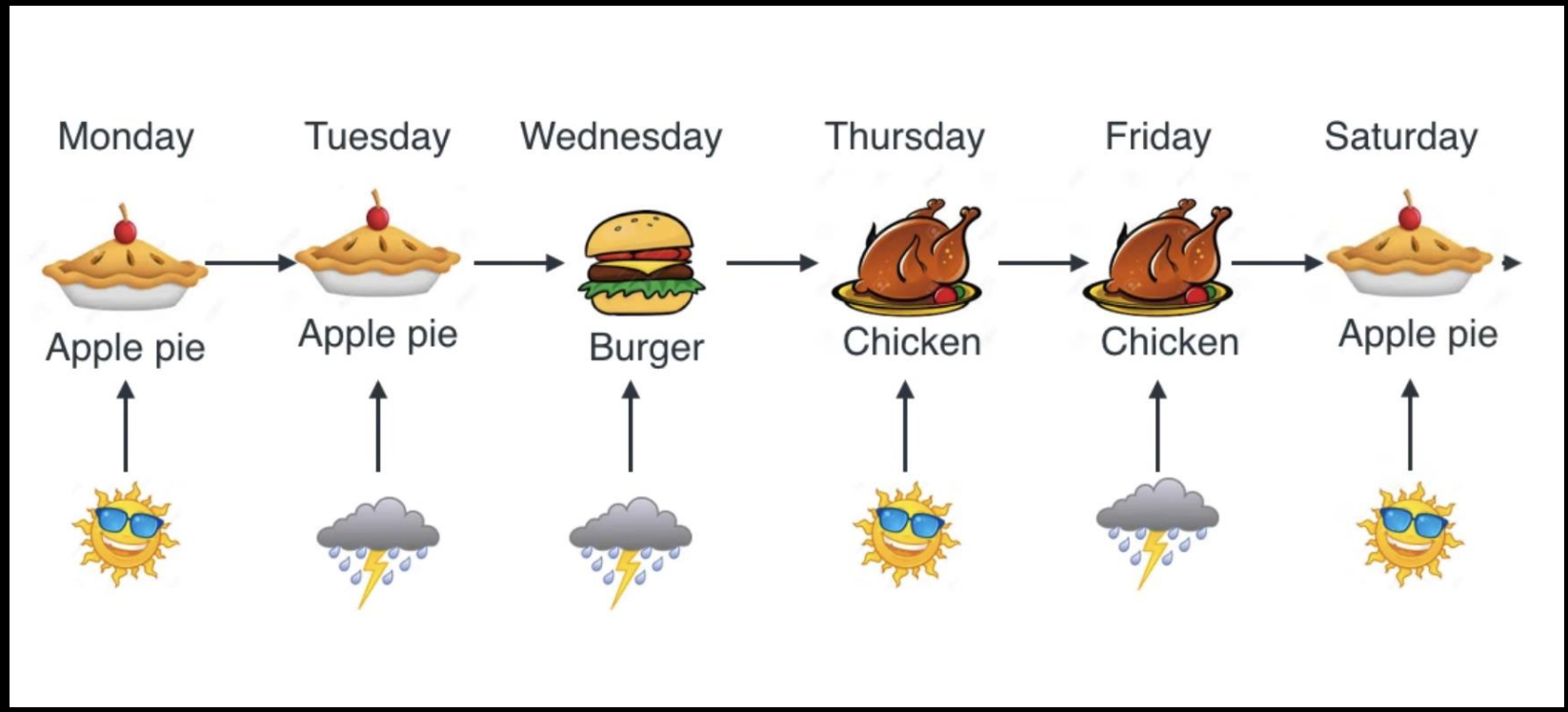
$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$



Simple (Recurrent) Neural Network

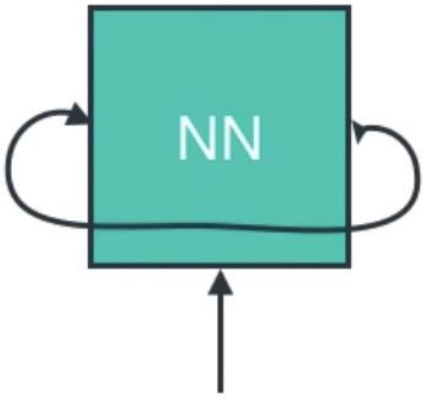


El cocinero que se consiguio una vida



Hay un orden, pero esta afectado por el clima

Recurrent Neural Network



Vectors

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Vamos a necesitar unas prescripciones matematicas mas ricas...

More Complicated RNN

$$\left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right]$$

Food

$$\left[\begin{array}{cc} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \hline 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{array}\right]$$

Weather

Food

$$\begin{bmatrix}1\\0\\0\end{bmatrix}$$ 

$$\begin{bmatrix}0\\1\\0\end{bmatrix}$$ 

$$\begin{bmatrix}0\\0\\1\end{bmatrix}$$ 

$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\\hline 0&0&1\\1&0&0\\0&1&0\end{bmatrix}\begin{bmatrix}1\\0\\0\end{bmatrix}$$  $= \begin{bmatrix}1\\0\\0\\\hline 0\\1\\0\end{bmatrix}$

Food

Food

$$\begin{bmatrix}1\\0\\0\end{bmatrix}$$ 

$$\begin{bmatrix}0\\1\\0\end{bmatrix}$$ 

$$\begin{bmatrix}0\\0\\1\end{bmatrix}$$ 

$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\\hline 0&0&1\\1&0&0\\0&1&0\end{bmatrix}\begin{bmatrix}1\\0\\0\end{bmatrix}$$  $=$ $$\begin{bmatrix}1\\0\\0\\\hline 0\\1\\0\end{bmatrix}$$  Same

 Next day

Food

Weather

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\left[\begin{array}{cc} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \hline 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{array}\right]$$

Weather

Weather

$\begin{bmatrix}1\\0\end{bmatrix}$ 

$\begin{bmatrix}0\\1\end{bmatrix}$ 

$\left[\begin{array}{cc}1&0\\1&0\\1&0\\\hline 0&1\\0&1\\0&1\end{array}\right]$ $\begin{bmatrix}1\\0\end{bmatrix}$  = $\left[\begin{array}{c}1\\1\\1\\\hline 0\\0\\0\end{array}\right]$  Same

 Next day

Weather

Weather

$$\begin{bmatrix}1\\0\end{bmatrix}$$ 

$$\begin{bmatrix}0\\1\end{bmatrix}$$ 

$$\begin{bmatrix}1 & 0\\1 & 0\\1 & 0\\\hline 0 & 1\\0 & 1\\0 & 1\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}$$ 

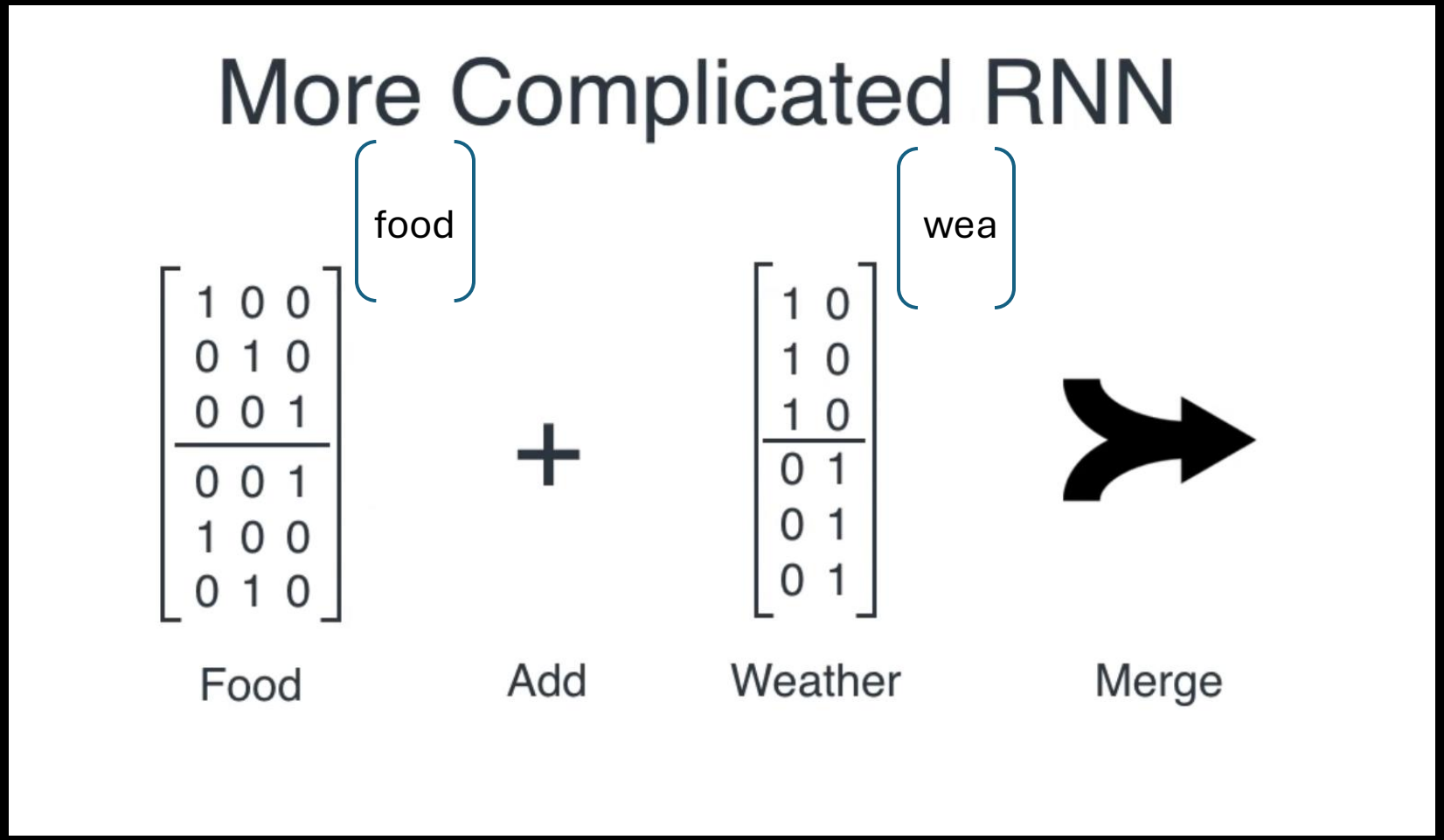
Weather

$$=\begin{bmatrix}0\\0\\0\\\hline 1\\1\\1\end{bmatrix}$$ 

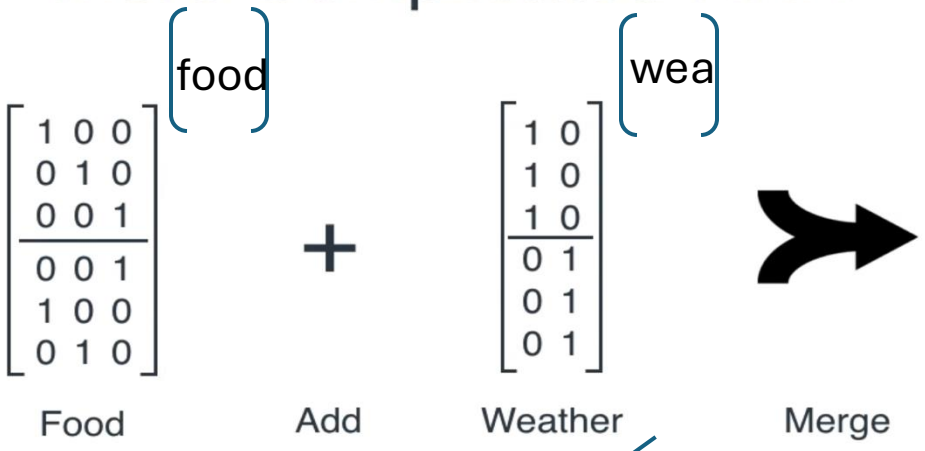
Same

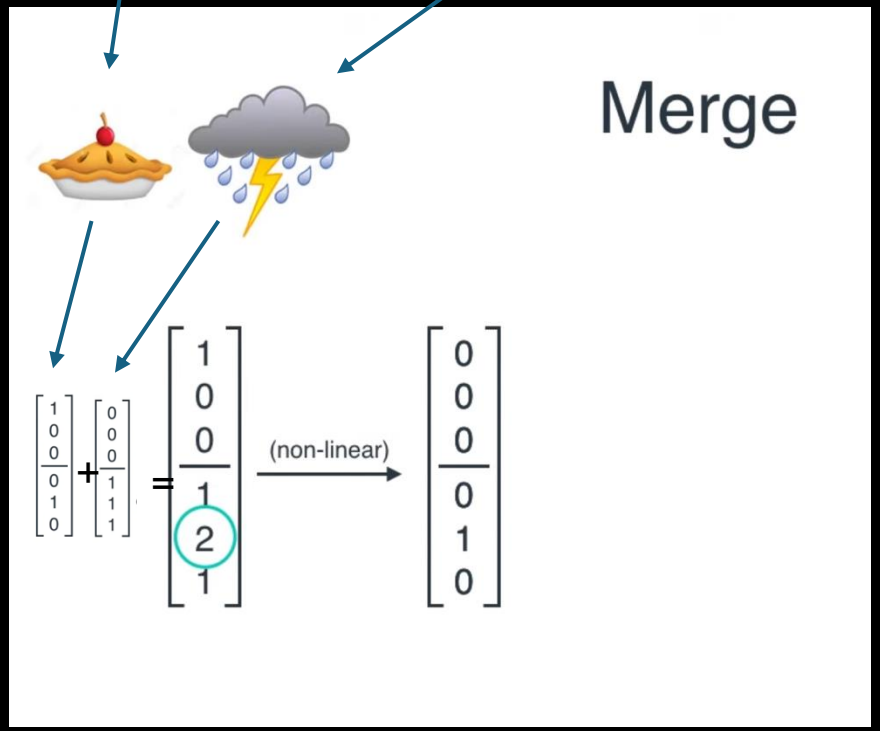
Next day

Vamos a necesitar unas prescripciones matematicas mas ricas...

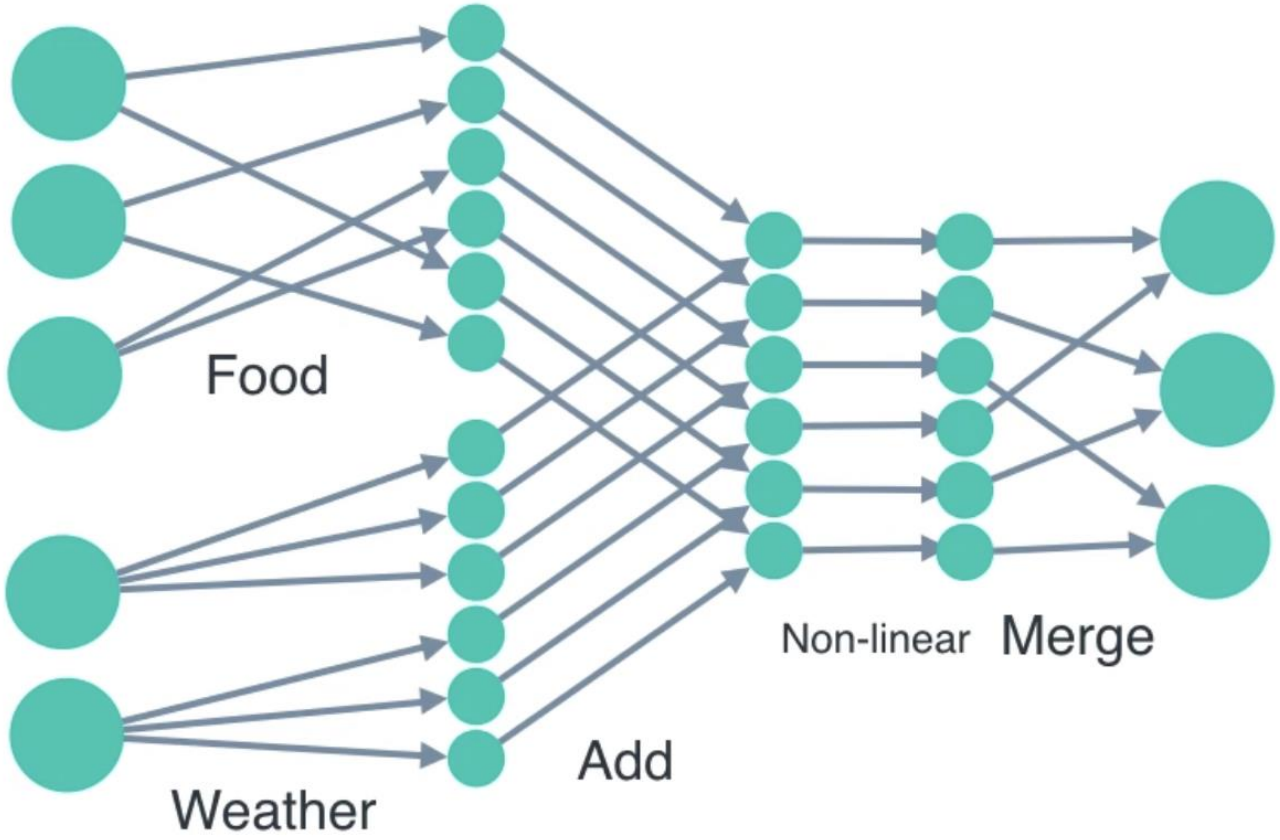


More Complicated RNN

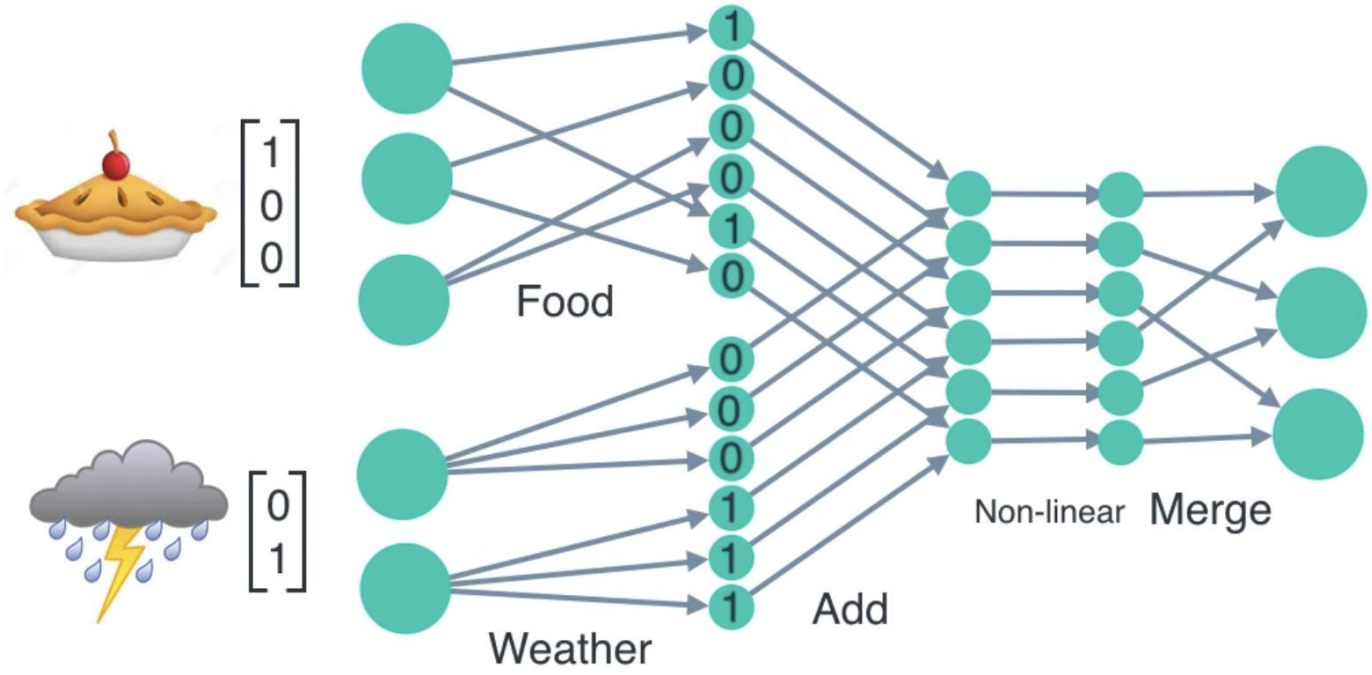




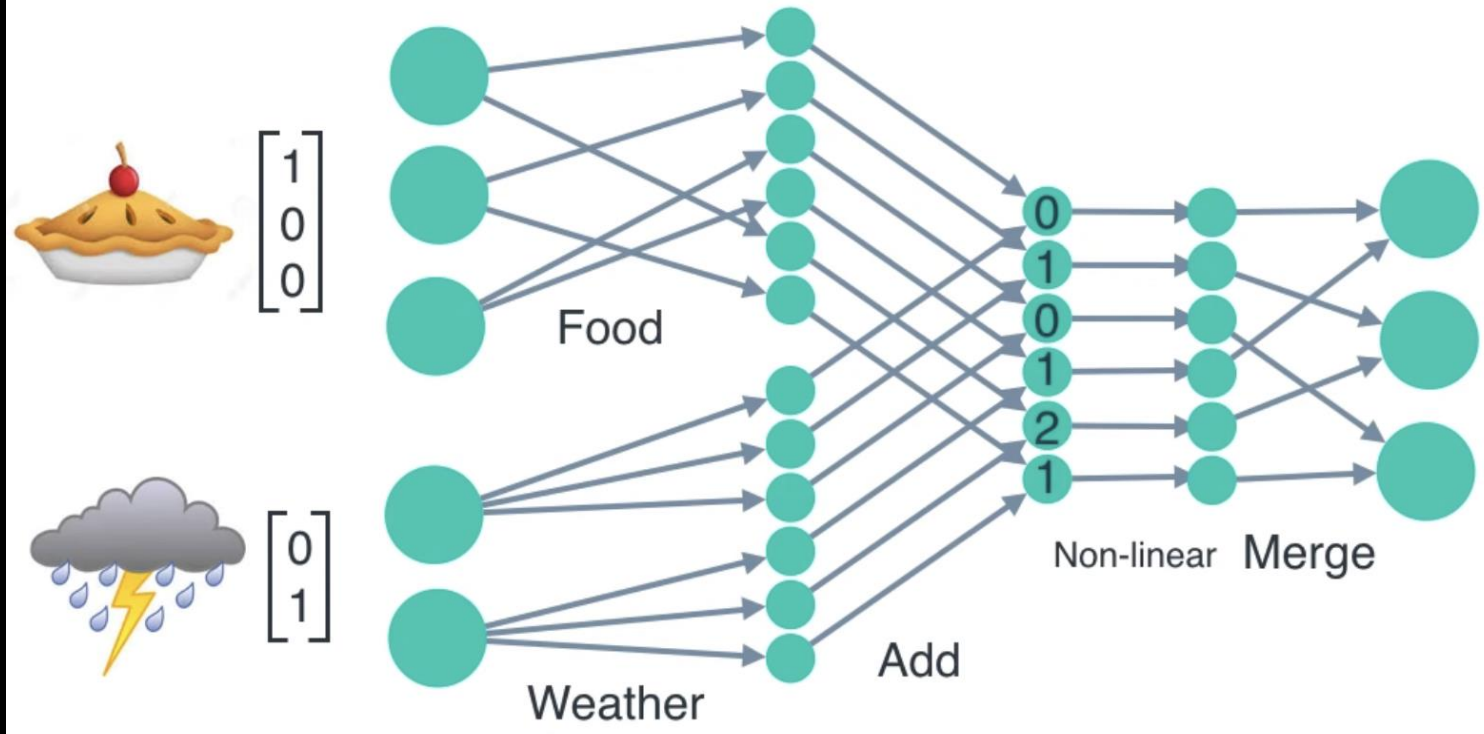
Recurrent Neural Network



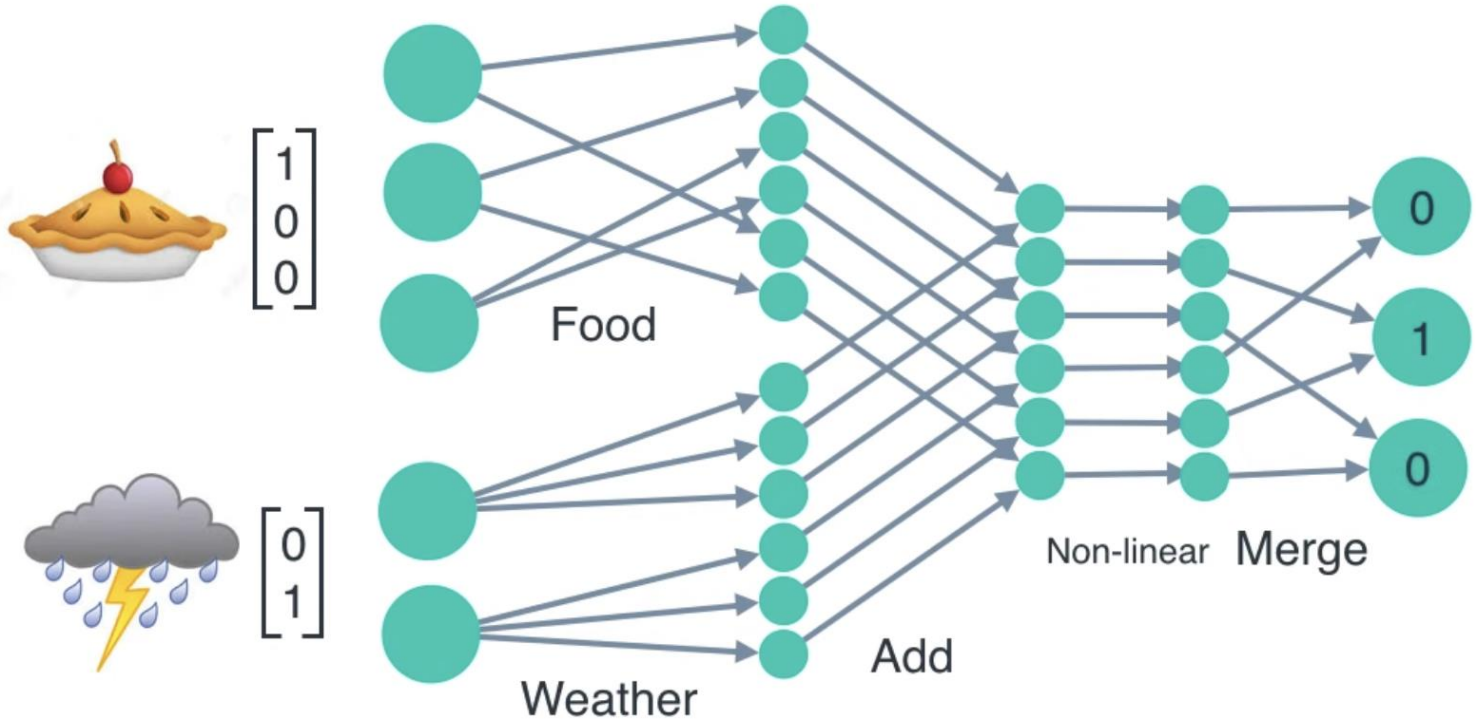
Recurrent Neural Network

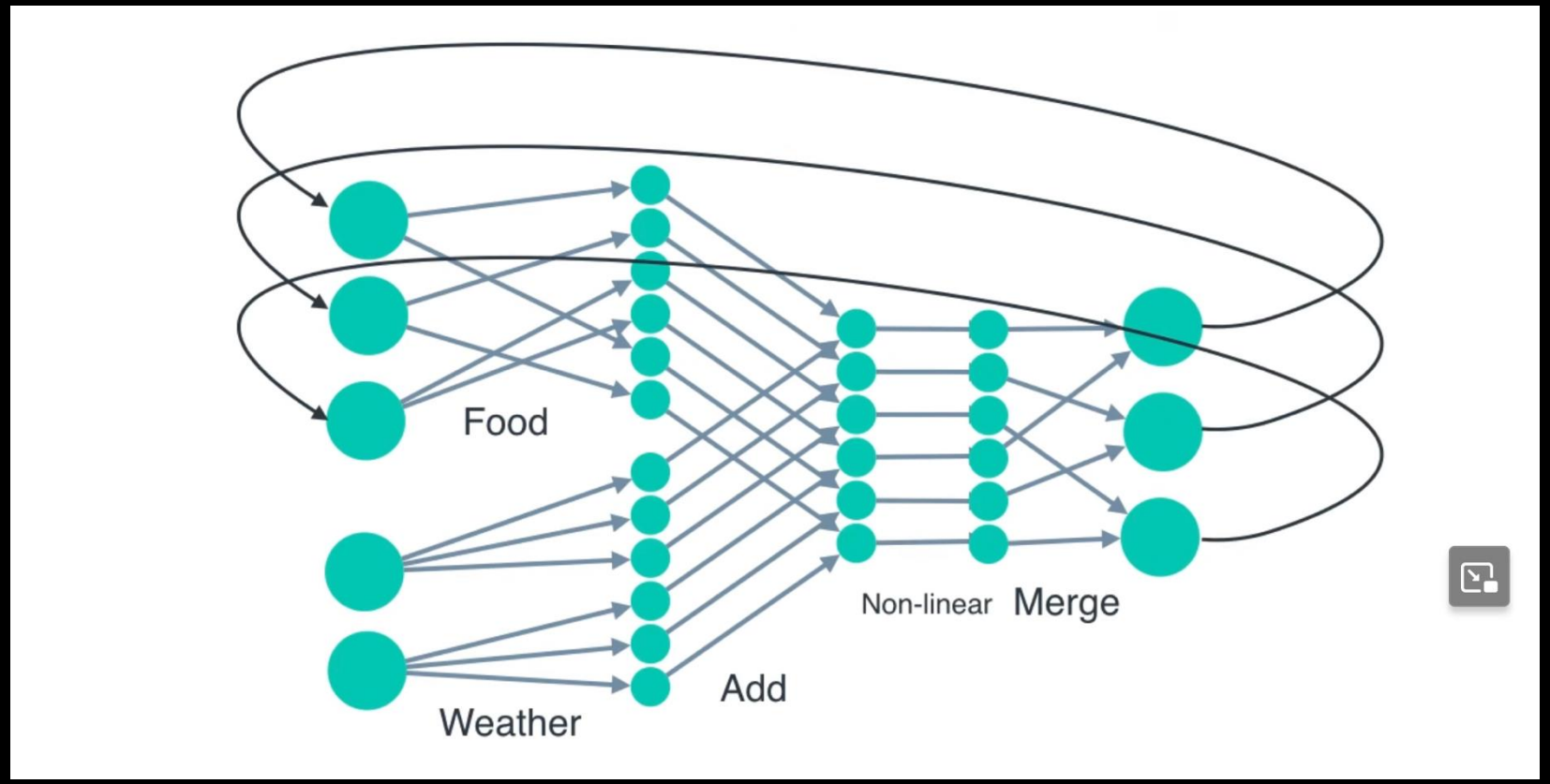


Recurrent Neural Network



Recurrent Neural Network




Food
Weather
Add
Non-linear
Merge

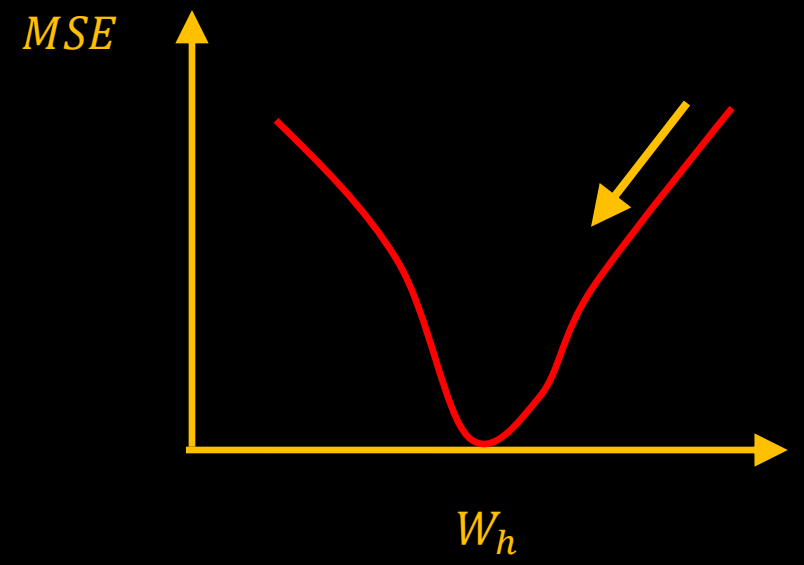
En este ejemplo, los pesos se diseñaron a mano.

Lo que sucede es que los pesos se entrenan a partir de una serie

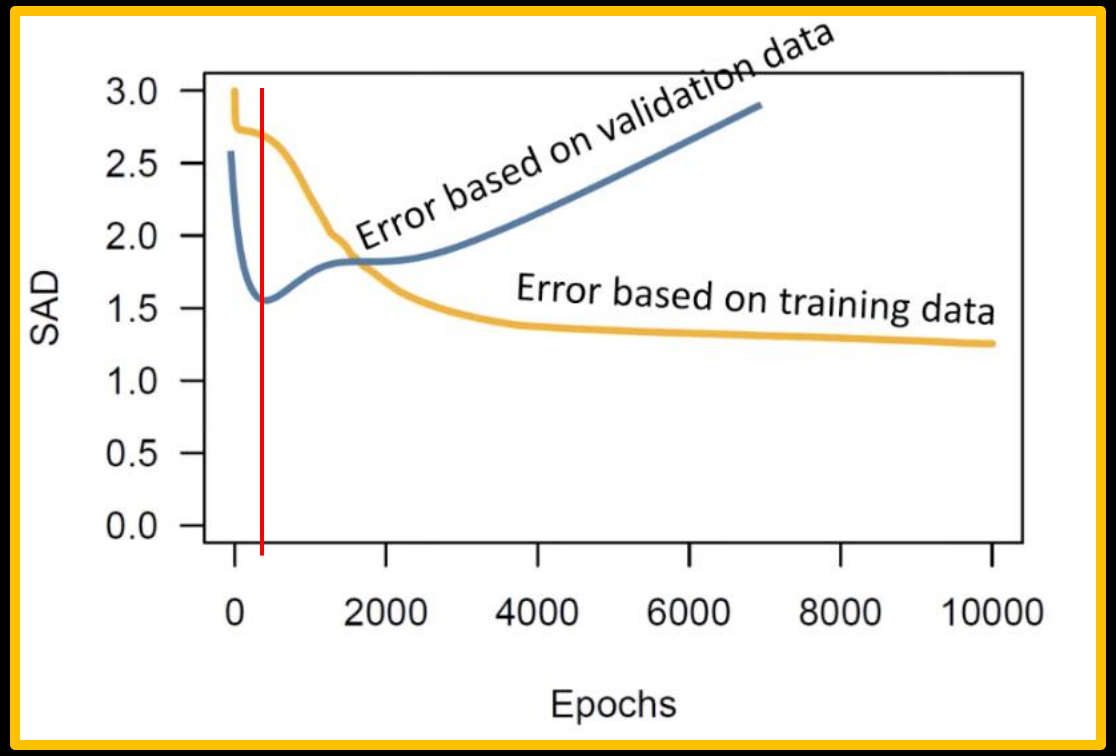
Como?

Con backpropagation, obvio.

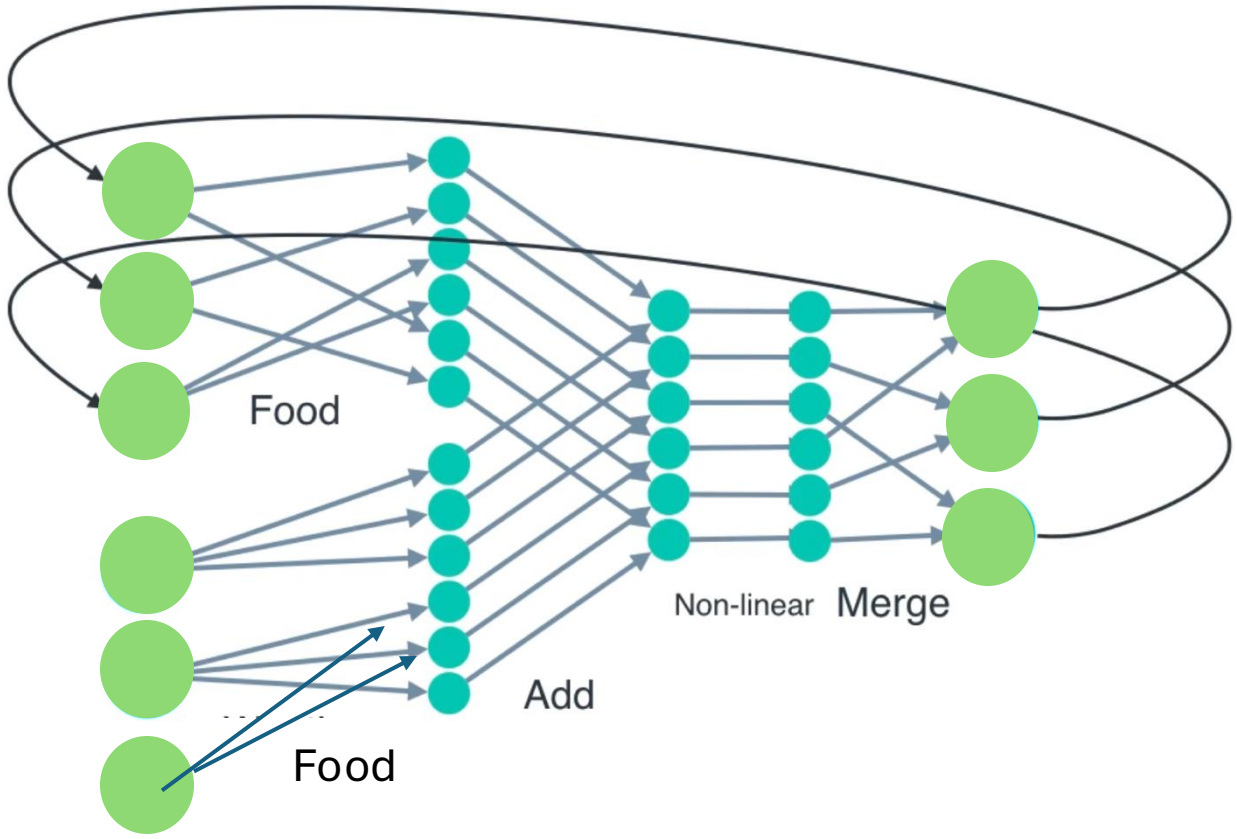
$$MSE = \sum_{1}^{9} (\hat{y}_i - y_i)^2$$

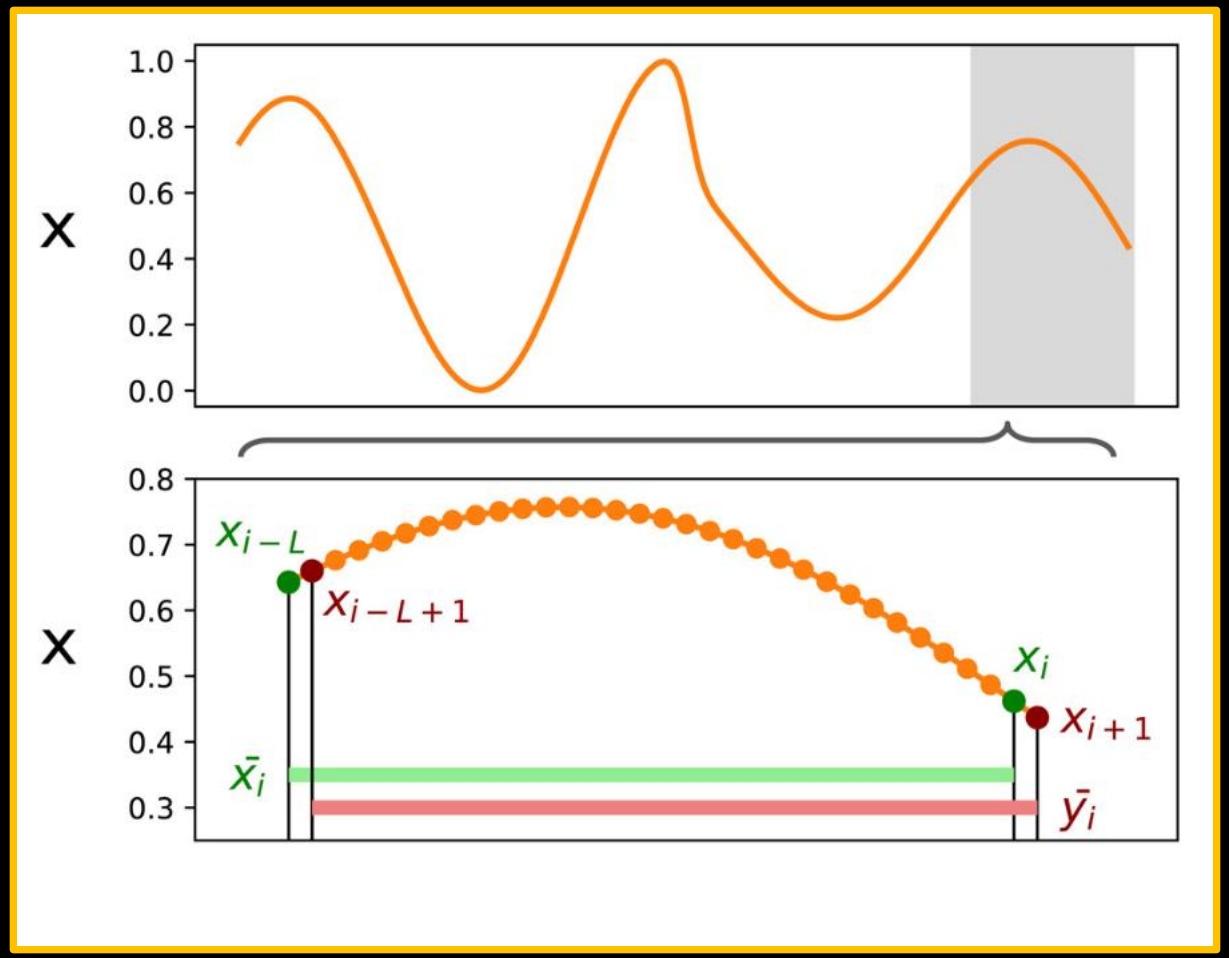


No tiene sentido y mas alla de la linea roja



Una señal temporal (vectorial) y punto...



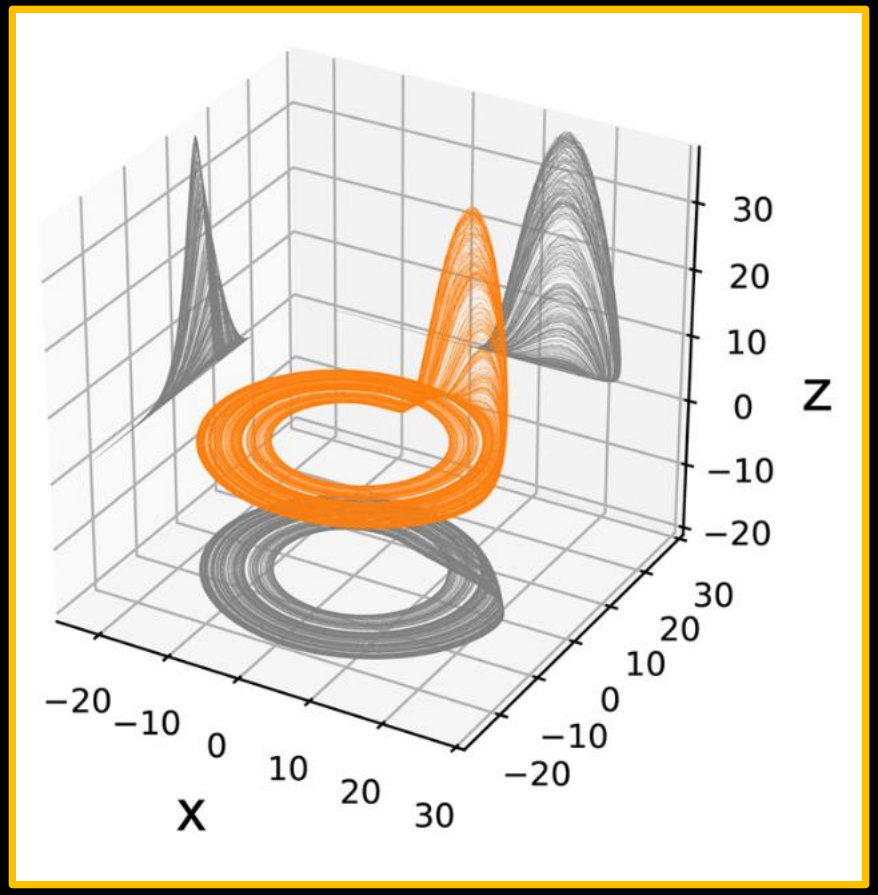


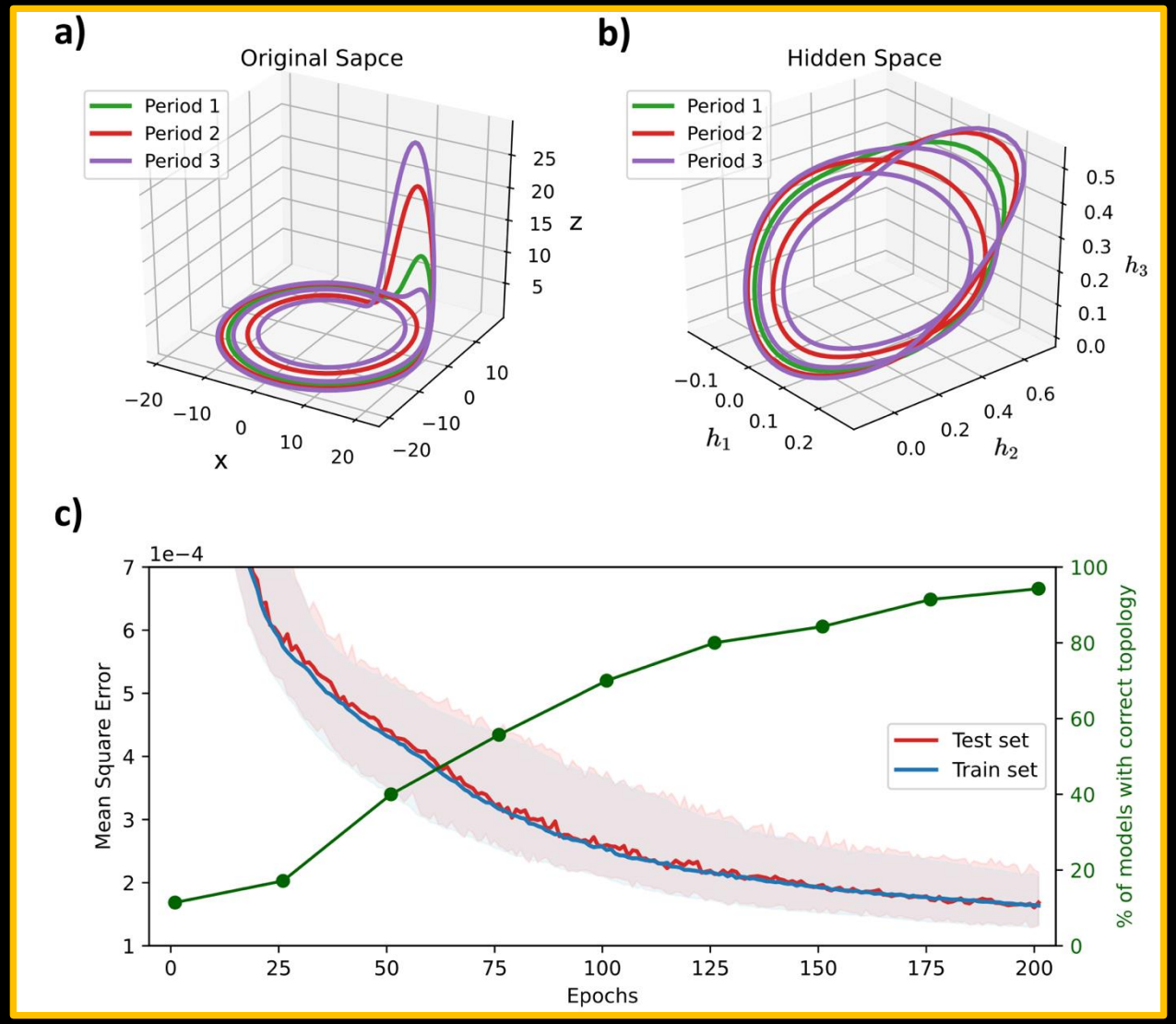
Mas bien unas cuarenta coordenadas, no 3, en este ejemplo...

$$\frac{dx}{dt} = -y - z$$

$$\frac{dy}{dt} = x + ay$$

$$\frac{dz}{dt} = b + z(x - c)$$




a)
Original Sapce
Period 1
Period 2
Period 3
25
20
15
Z
10
5
−20
−10
0
10
20
X
10
0
−10
−20
b)
Hidden Space
Period 1
Period 2
Period 3
0.5
0.4
0.3
0.2
0.1
0.0
h_3
−0.1
0.0
0.1
0.2
h_1
0.0
0.2
0.4
0.6
h_2
c)
1e−4
7
6
5
4
3
2
1
Mean Square Error
100
80
60
40
20
0
% of models with correct topology
Test set
Train set
0
25
50
75
100
125
150
175
200
Epochs

Emerge una pintura muy interesante,
en la cual tambien en estas redes que
predicen la dinamica, se mapea el flujo
preservando la topologia. En este caso,
no tengo demostracion de por que esto funciona.

1) RNN Simple

Keras

```
model = Sequential()
model.add(SimpleRNN(5, activation='tanh', input_shape=(look_back, n_features)))
model.add(Dense(1,activation='linear'))
model.compile(optimizer='adam', loss='mse')
```

Importante: Los datos de entrada deben tener la forma:

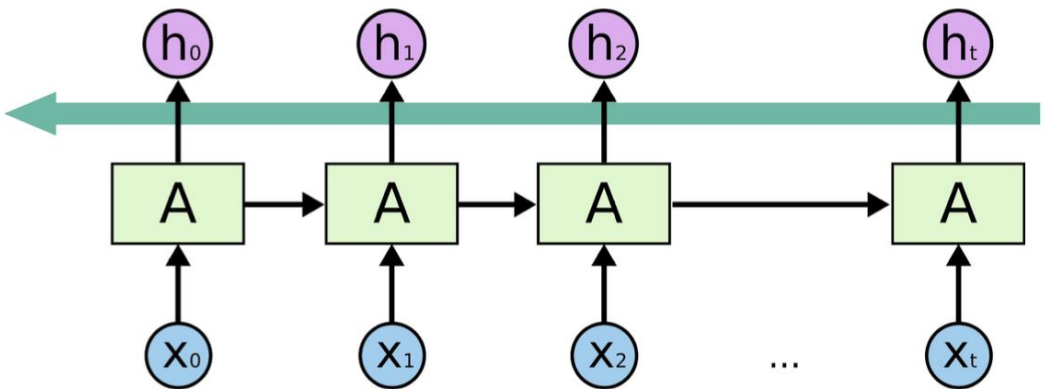
X.shape = [num_samples, look_back , n_features]

En la señal anterior, num_samples es del orden de 40000, los look back unos 40, y las features 1

1) RNN Simple

Vanishing gradient problem

Problema! Les cuesta aprender dependencias de largo plazo.



Backpropagation through time

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

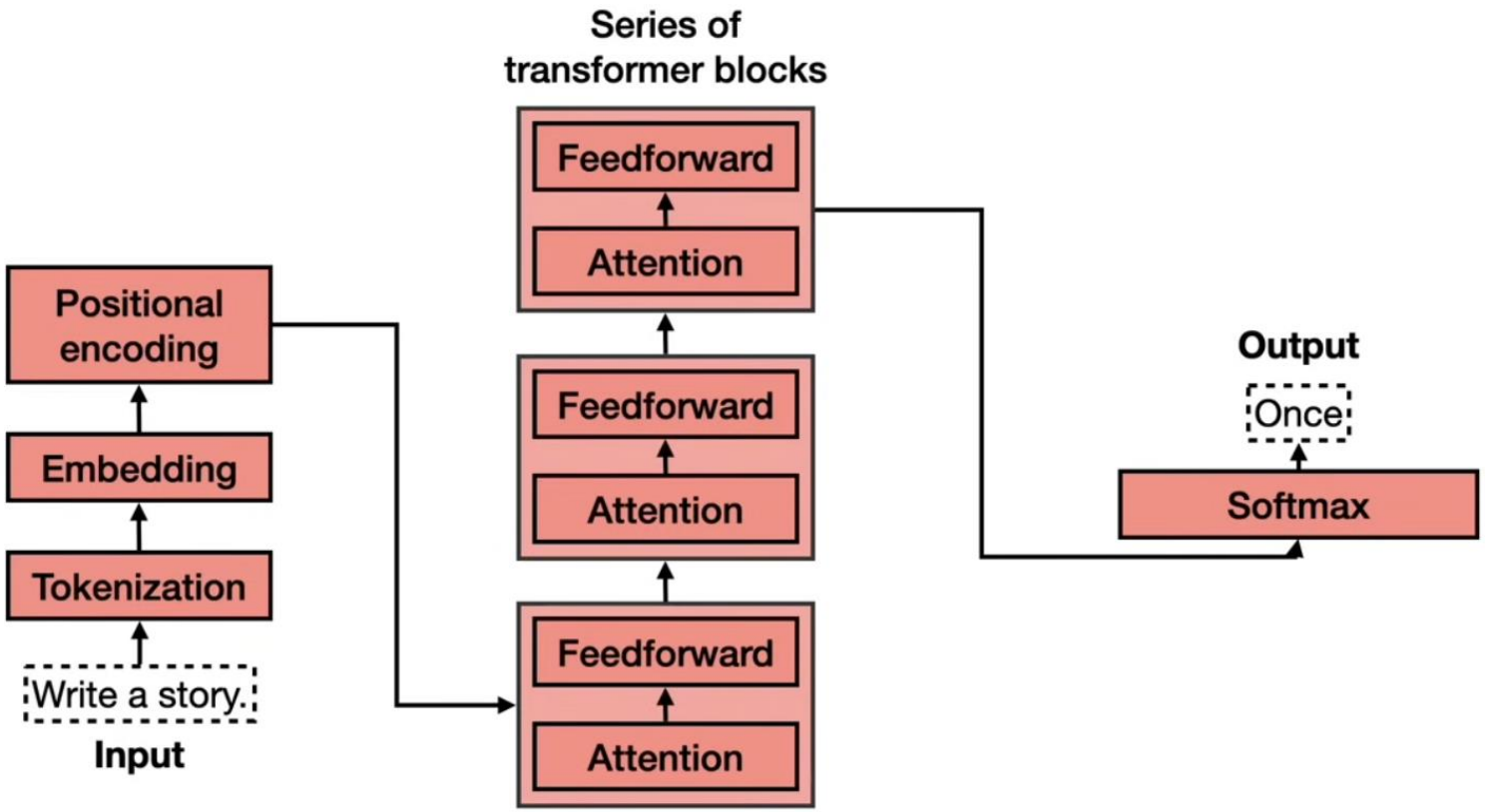
Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Łukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

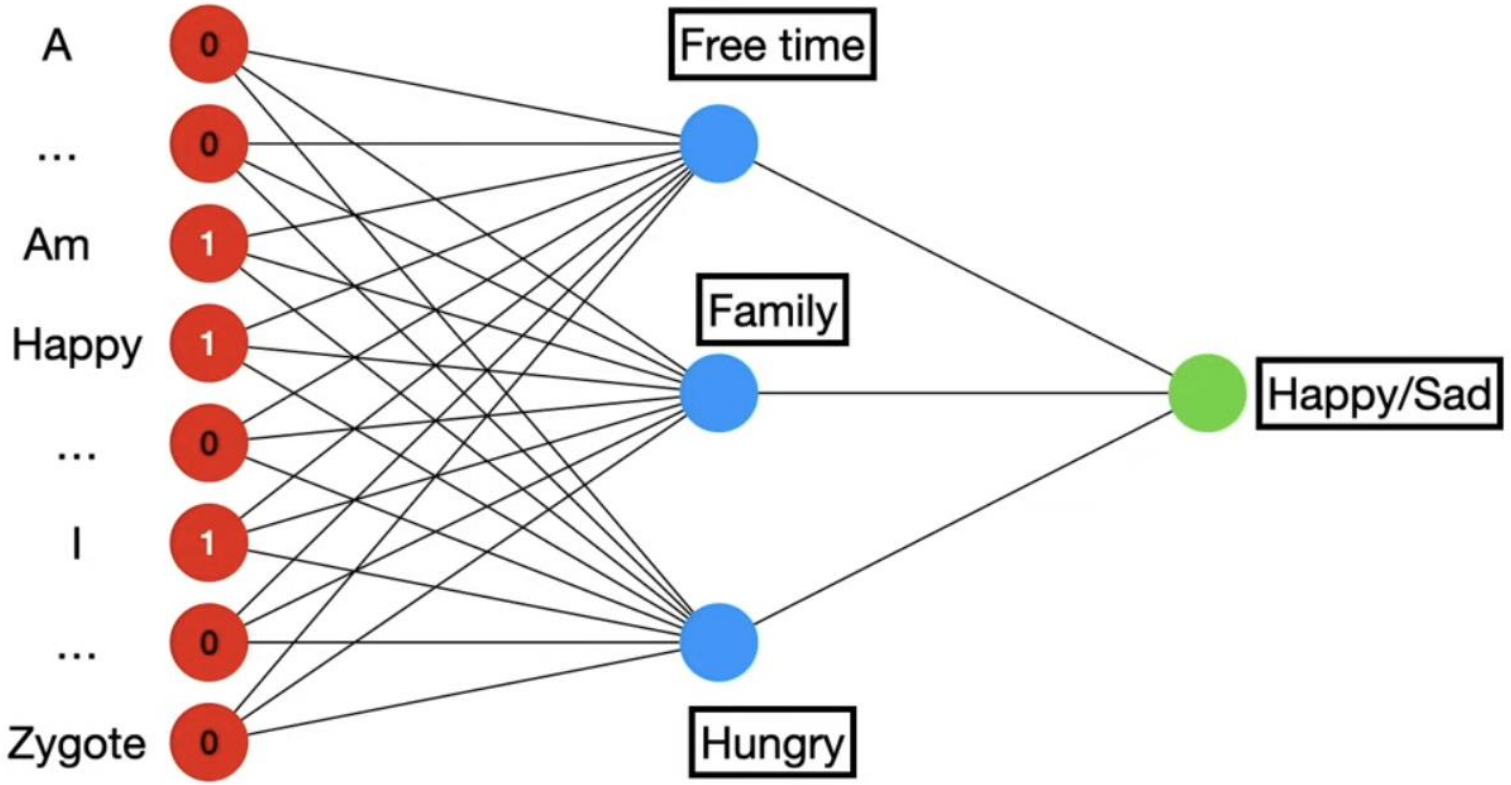
Transformers... prediccion, palabra por palabra.



La construccion de un embedding “semantico”

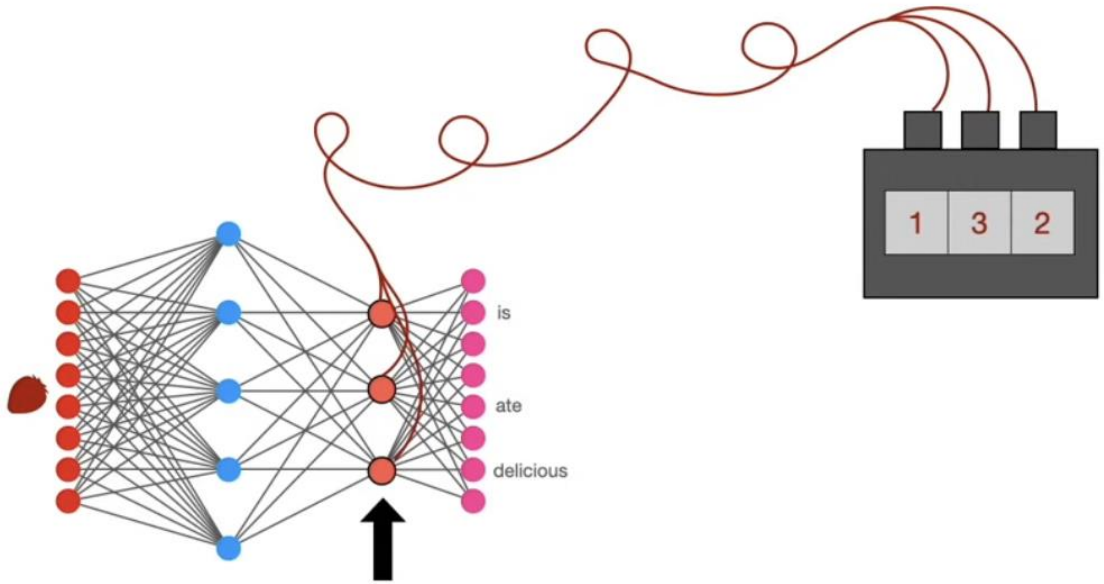
Sea una red “all to all”,
preparada para predecir la proxima palabra

I
Have
Spent
Time
With
My
Family
Therefore
I
am



Un modo de hacer embeddings

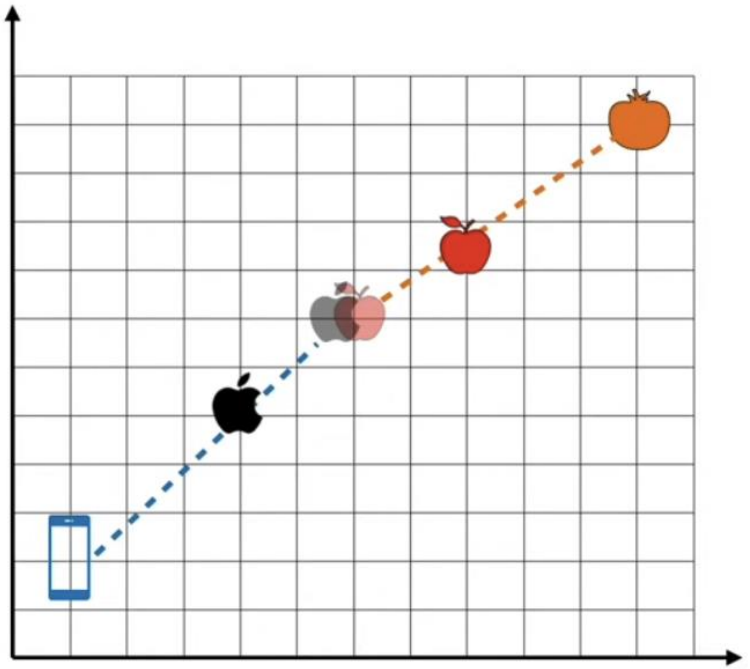
Representacion via los valores numericos de la capa anteultima

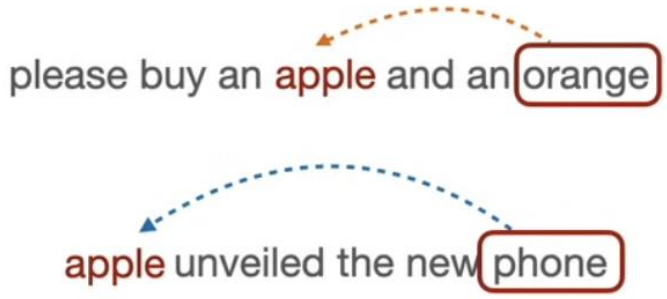


Word	Numbers		
Strawberry	1	3	2
Apple	1.1	2.9	2.2

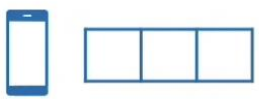
Attention

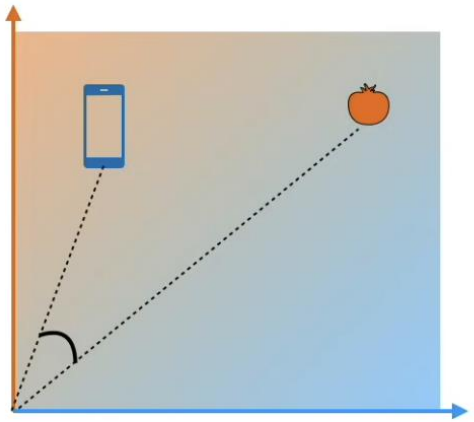
Un modo de que las palabras sean afectadas por el contexto:
Alterar el embedding de las palabras segun las demas del contexto

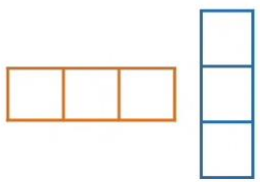




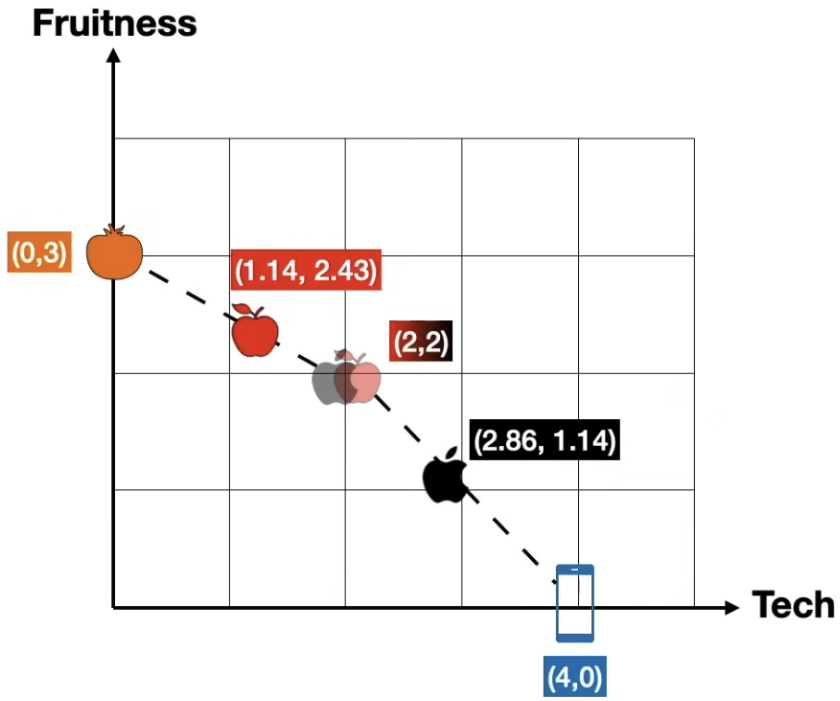
Similarity





Similarity  = 

Attention refiere a moverse en el espacio de embedding
en la direccion que sugiera la similitud



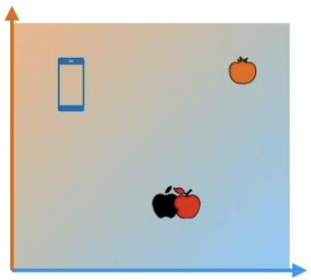
an apple and an orange

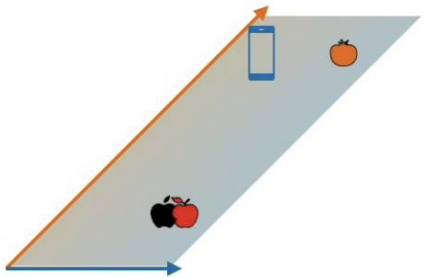
Apple → 0.43 Orange + 0.57 Apple

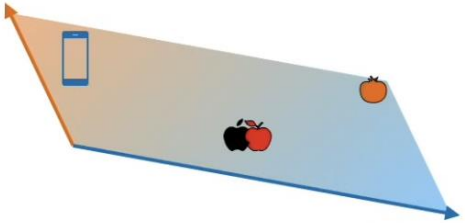
an apple phone

Apple → 0.43 Phone + 0.57 Apple

Ahora, para mover segun el contexto,
es mejor realizar una transformacion lineal

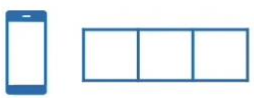


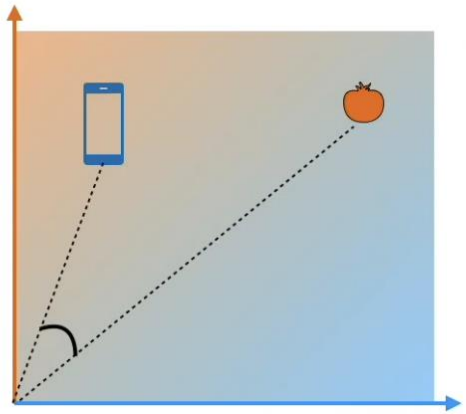


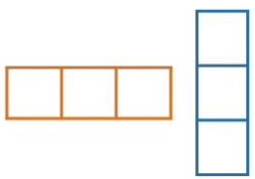


(donde el corrimiento funcione mejor)

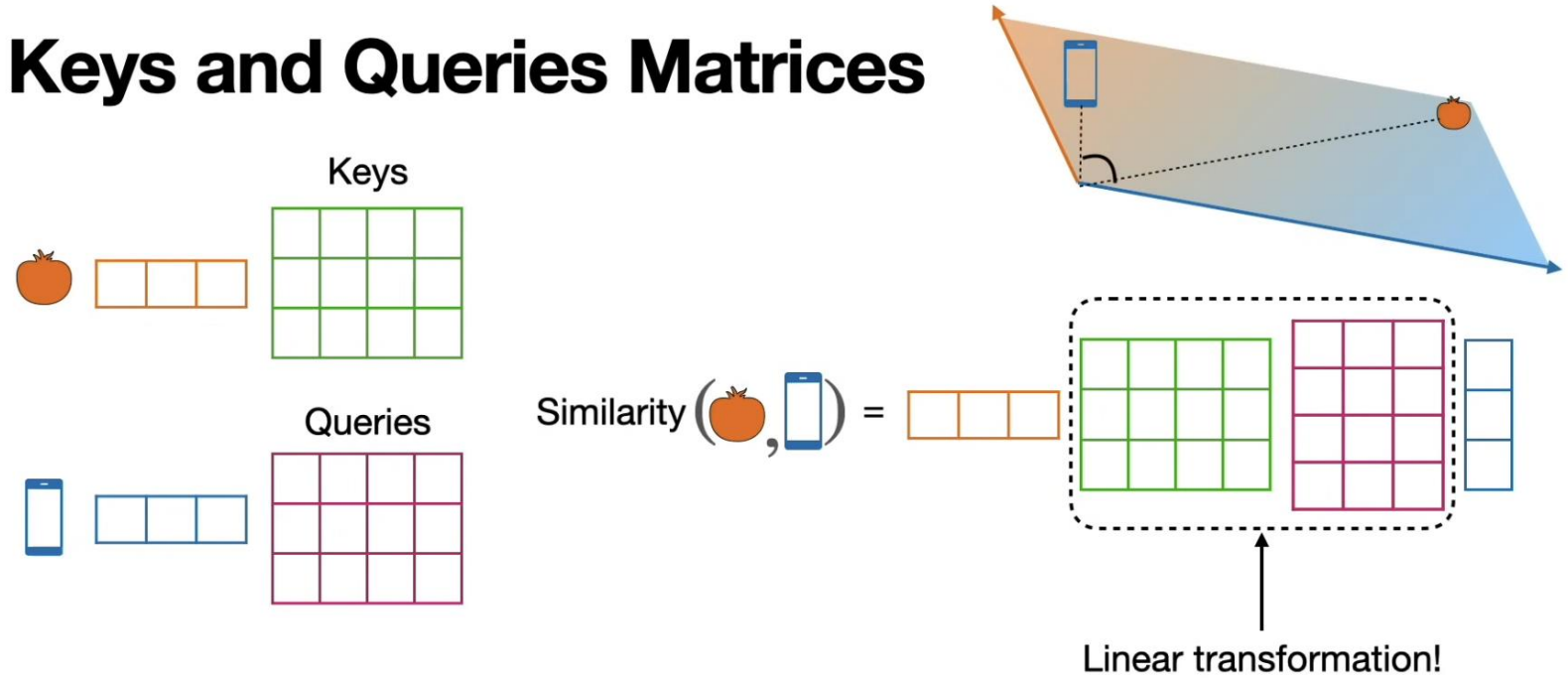
Similarity



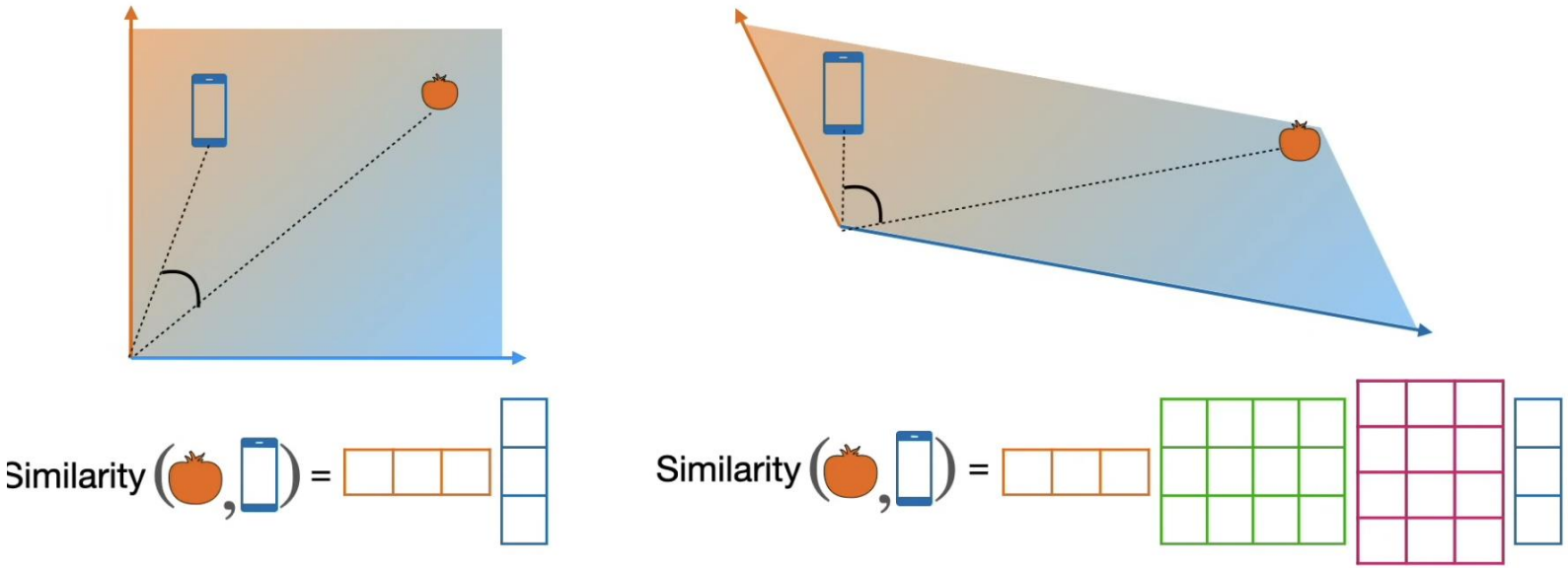


Similarity  = 

Un nuevo embedding

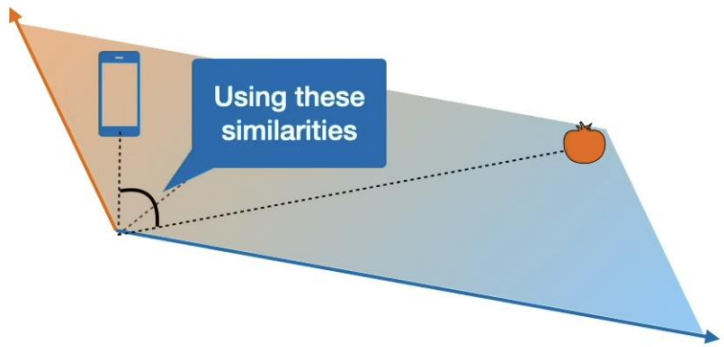


Similarity on a transformed embedding

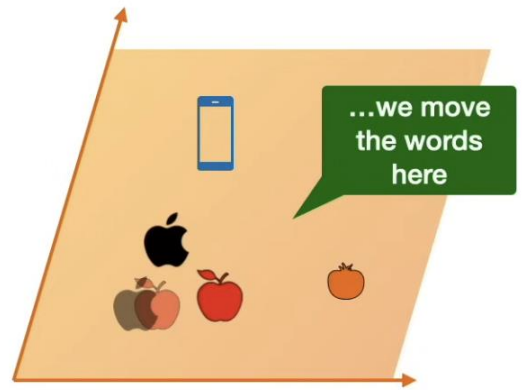


Y la idea es que attention funcione mejor en este nuevo embedding

Values matrix

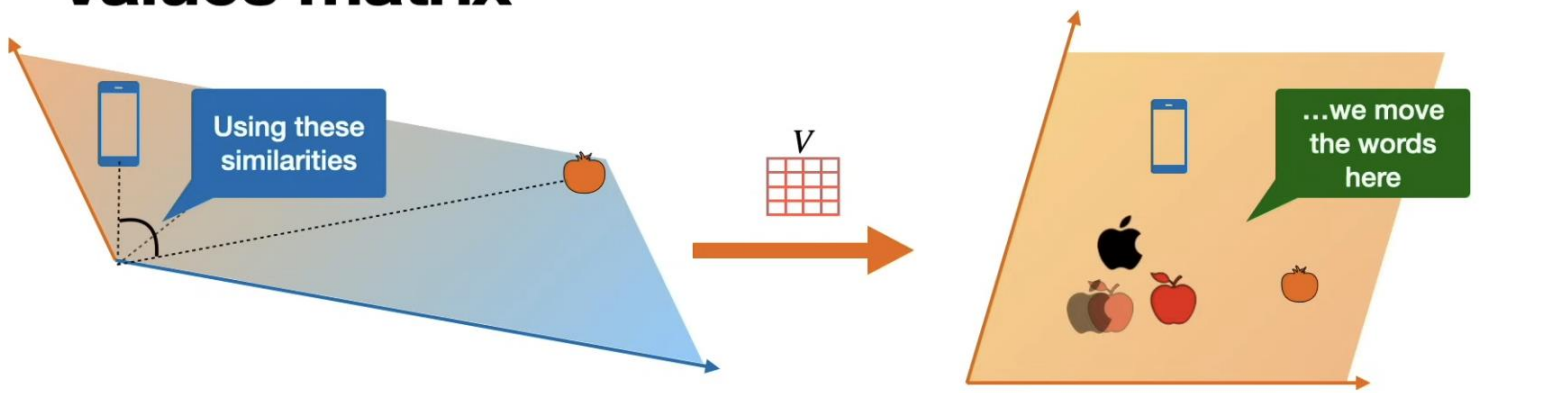


Best embedding for finding similarities



Best embedding for finding the next word

Values matrix



Best embedding for finding similarities

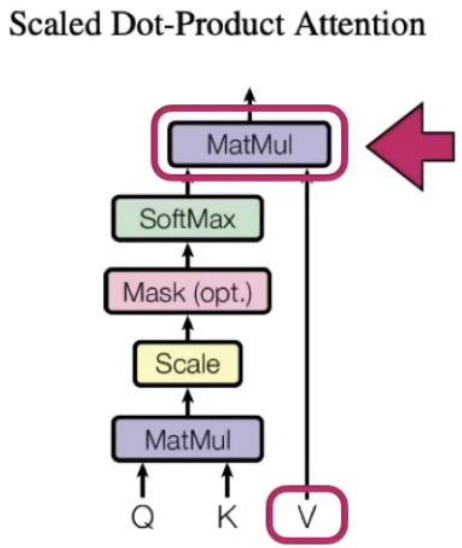
Best embedding for finding the next word

Keys

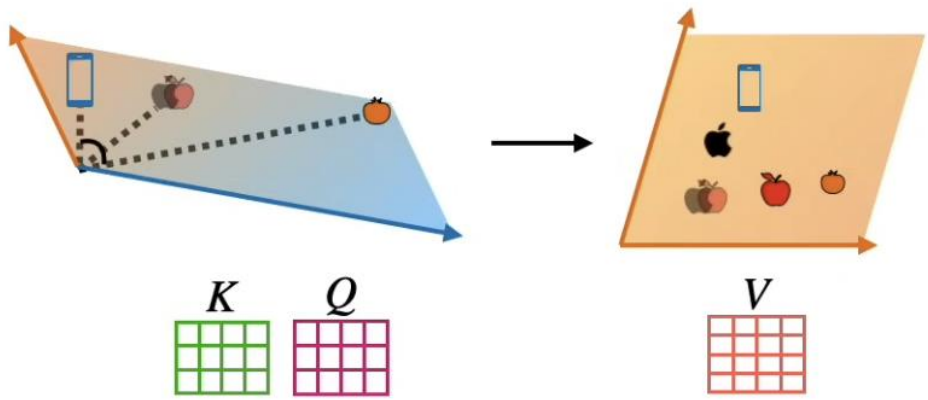
Queries

Values

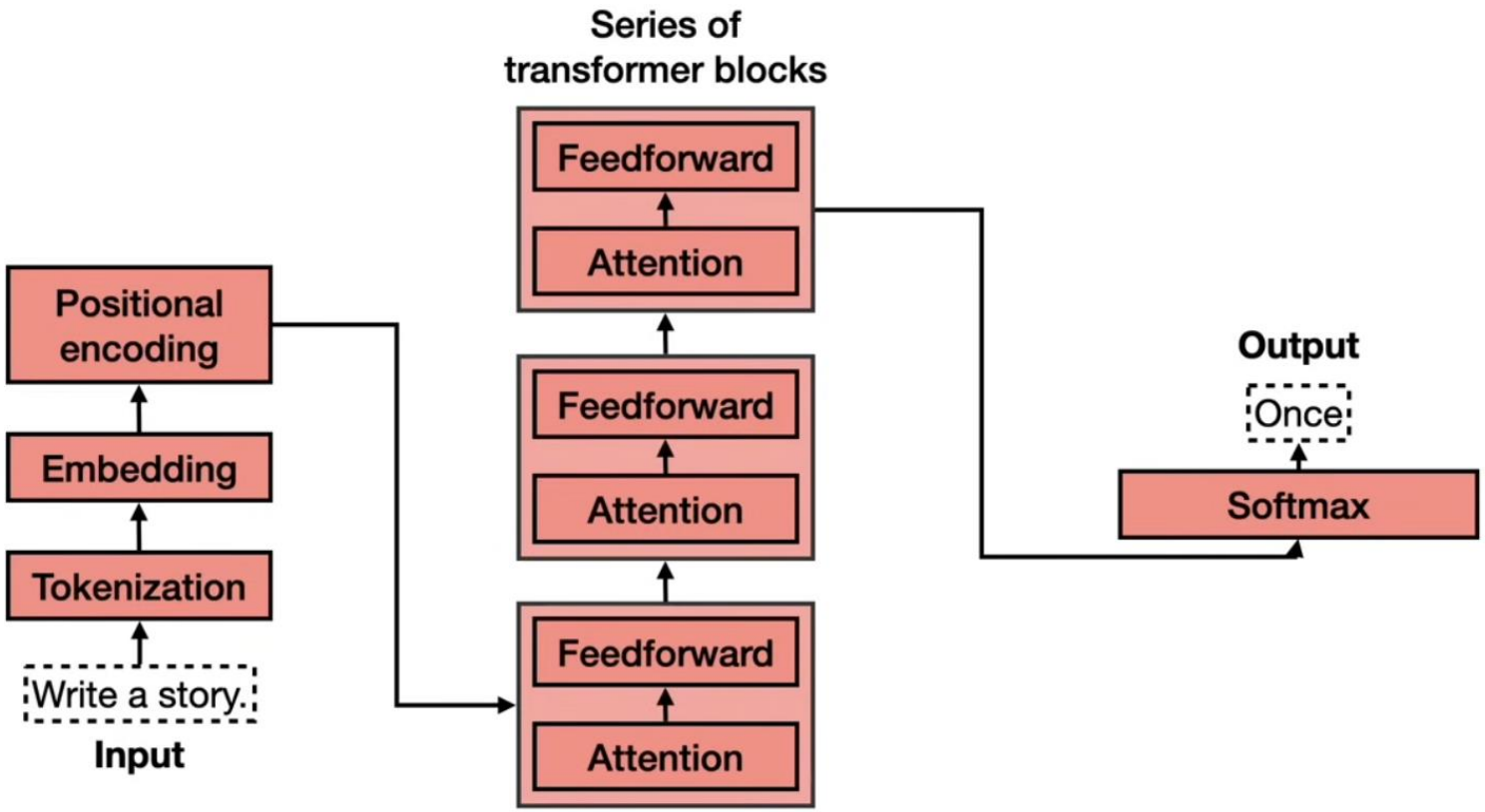
Self-attention

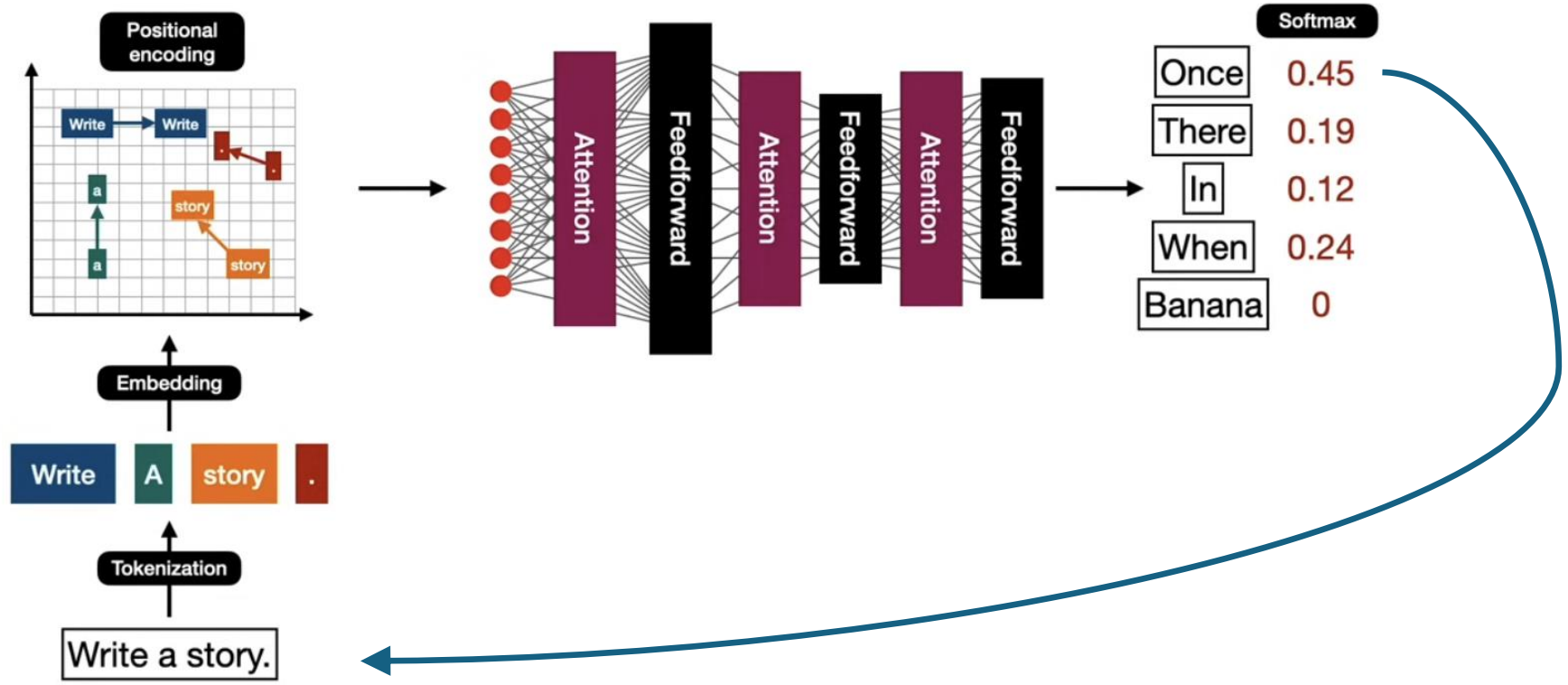


$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$



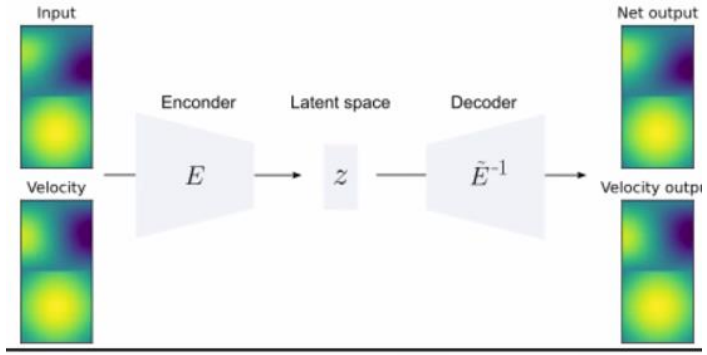
Transformers... prediccion, palabra por palabra.

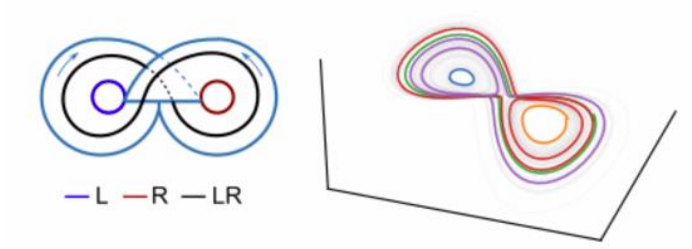



Positional encoding
Write
Write
a
story
a
story
Embedding
Write
A
story
.
Tokenization
Write a story.
Attention
Feedforward
Attention
Feedforward
Attention
Feedforward
Softmax
Once 0.45
There 0.19
In 0.12
When 0.24
Banana 0

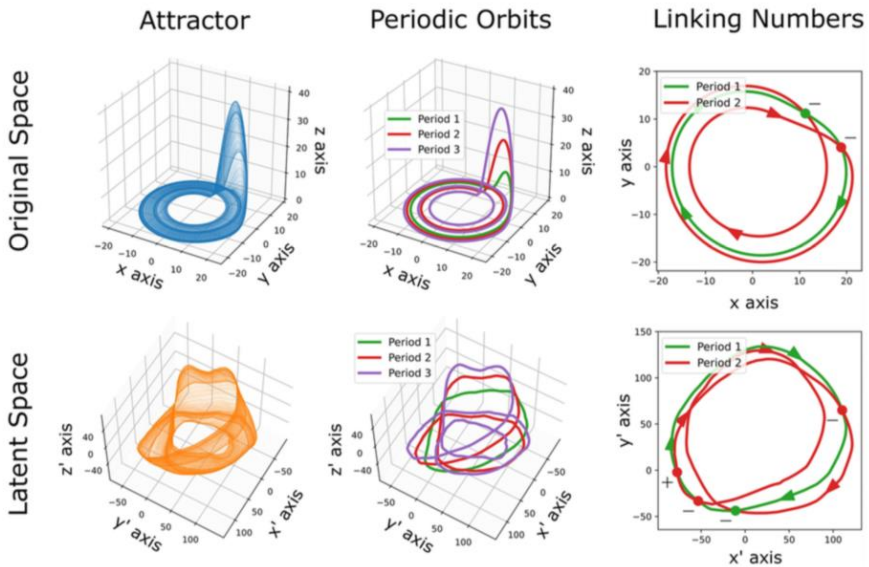
The picture that emerges is that the neural networks used in different implementations of IA projects, do recreate the original flow of the dynamics.

AE trained with movie





AE trained with temporal data



RNN trained with temporal data