Recurrent Neural Networks

Machine Learning

(Largely based on slides from Luis Serrano & Fei-Fei Li & Andrej Karpathy & Justin Johnson & Serena Yeung)

Prof. Sandra Avila
Institute of Computing (IC/Unicamp)

MC886, October 30, 2019

Today's Agenda

- Recurrent Neural Networks
  - An Intuitive Explanation
  - A More Formal Explanation
  - Vanilla vs LSTMs

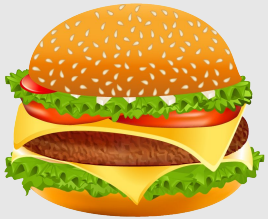
RNNs:
An Intuitive Explanation

Perfect Roommate

Luis Serrano: https://youtu.be/UNmqTiOnRfg

Perfect Roommate

Pasta



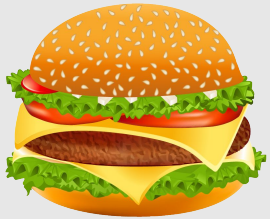
Burger

Salad

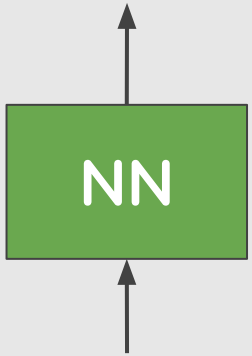
Weather

Weather

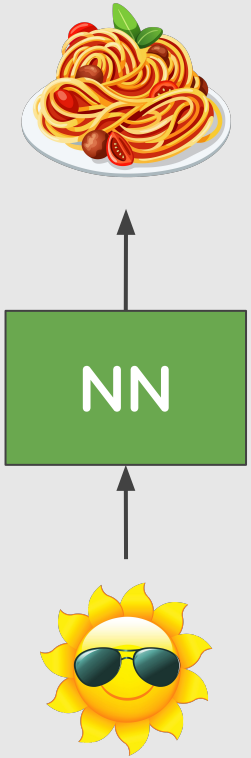
Weather



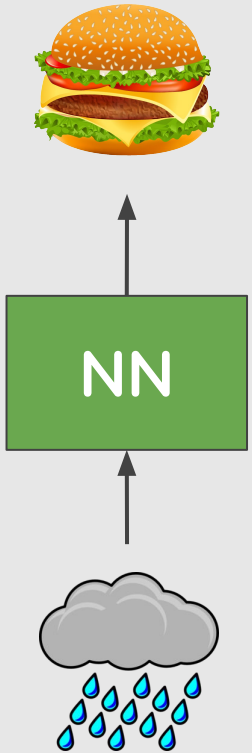
Neural Network



Neural Network

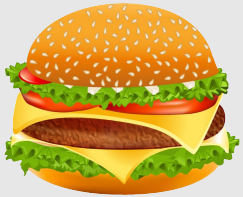


Neural Network



Neural Network

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



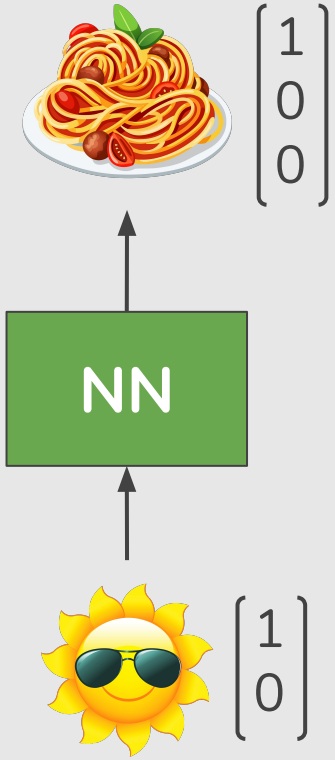
$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

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

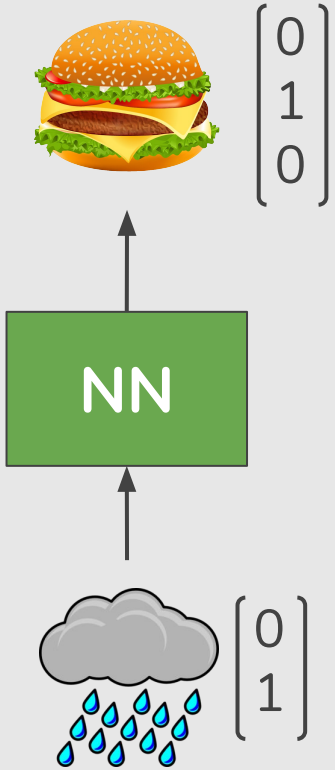
$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

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

Neural Network



Neural Network



Neural Network

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

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

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

Neural Network

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

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

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

Neural Network

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

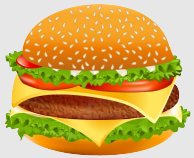
$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

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

Neural Network

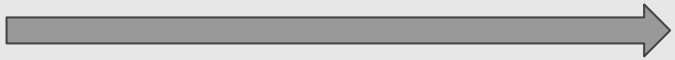
$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

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

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

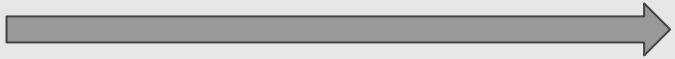
Neural Network

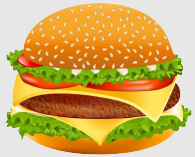
$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$



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

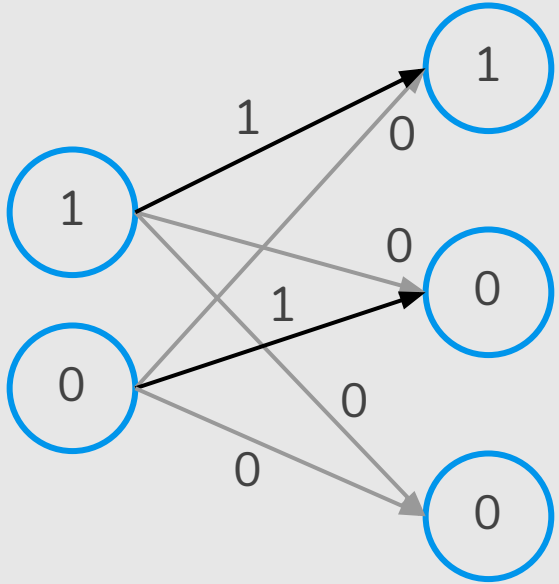
$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$



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

Neural Network

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



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

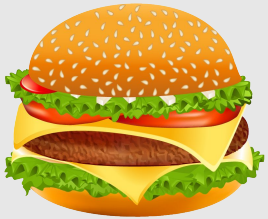
Perfect Roommate

Perfect Roommate

Pasta

Perfect Roommate

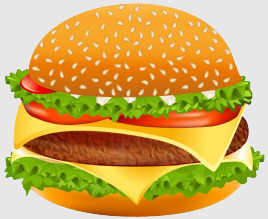
Pasta



Burger

Perfect Roommate

Pasta

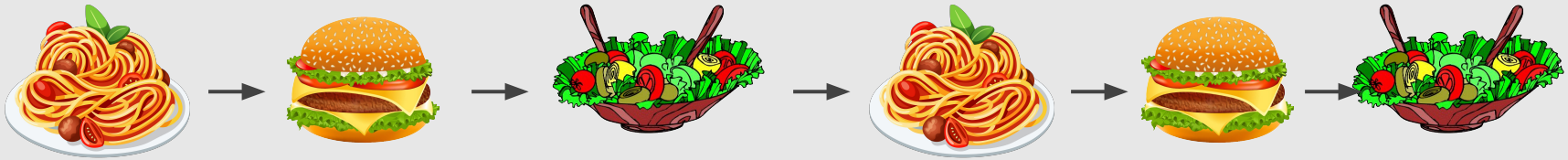


Burger

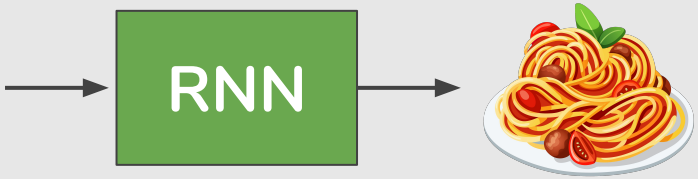
Salad

Cooking Schedule

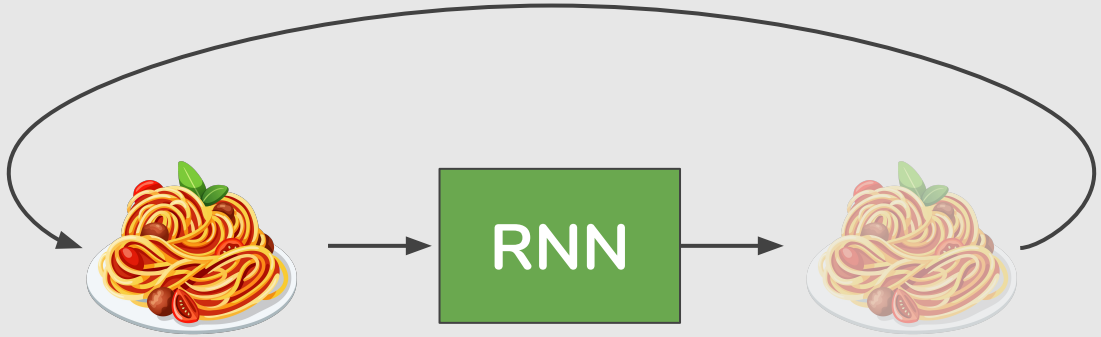
Monday Tuesday Wednesday Thursday Friday Saturday



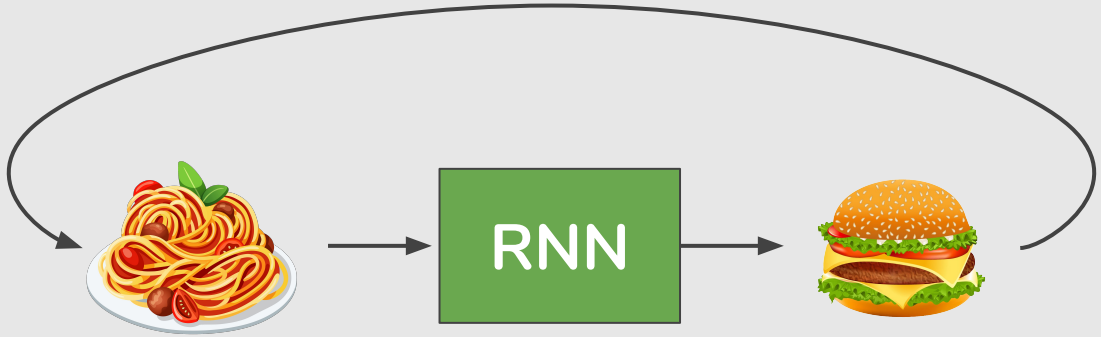
Recurrent Neural Network



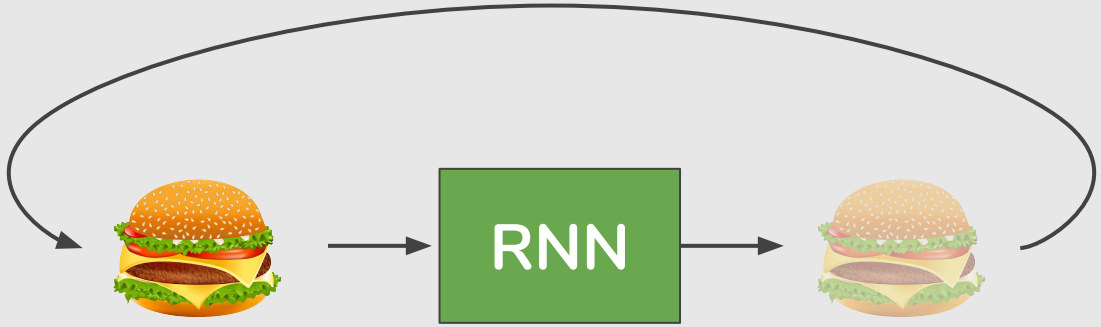
Recurrent Neural Network



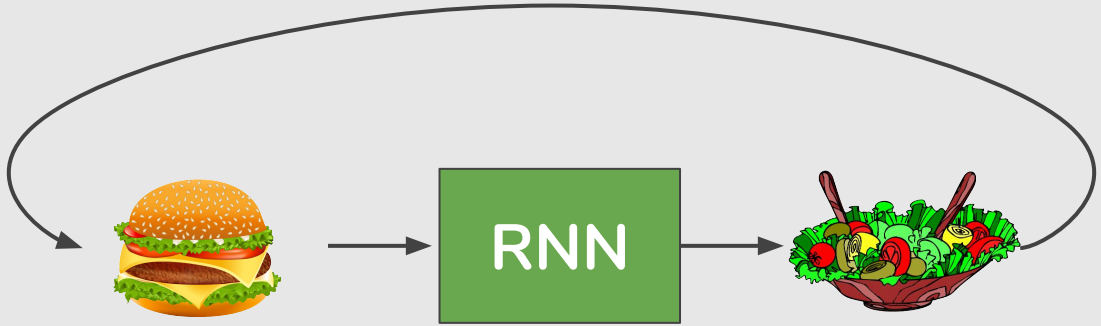
Recurrent Neural Network



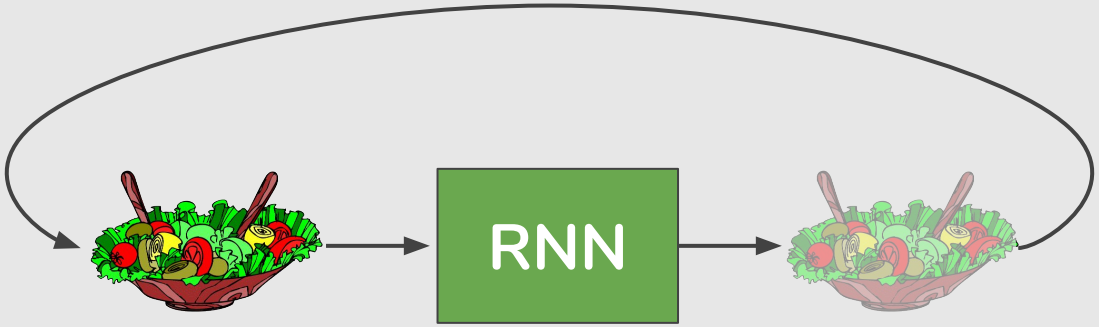
Recurrent Neural Network



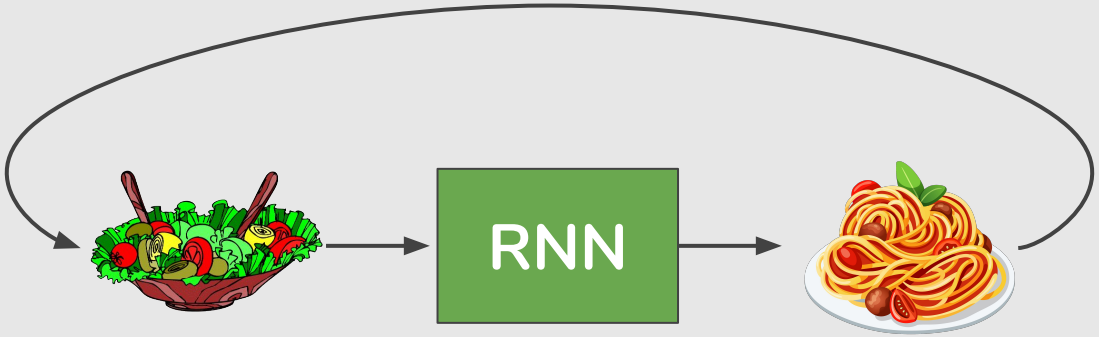
Recurrent Neural Network



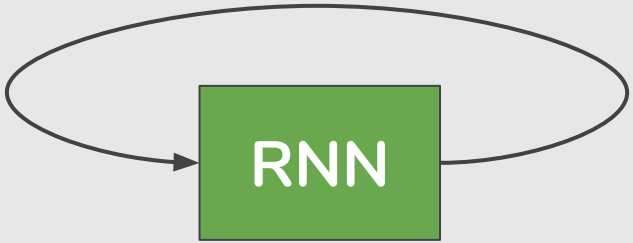
Recurrent Neural Network



Recurrent Neural Network

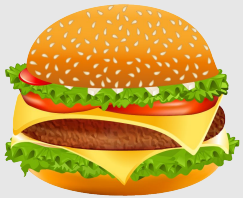


Recurrent Neural Network



Recurrent Neural Network

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



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

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

Recurrent Neural Network

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

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

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

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

Recurrent Neural Network

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

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

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

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

Recurrent Neural Network

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

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

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

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

Recurrent Neural Network

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

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

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

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

Recurrent Neural Network

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

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

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

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

Recurrent Neural Network

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

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

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

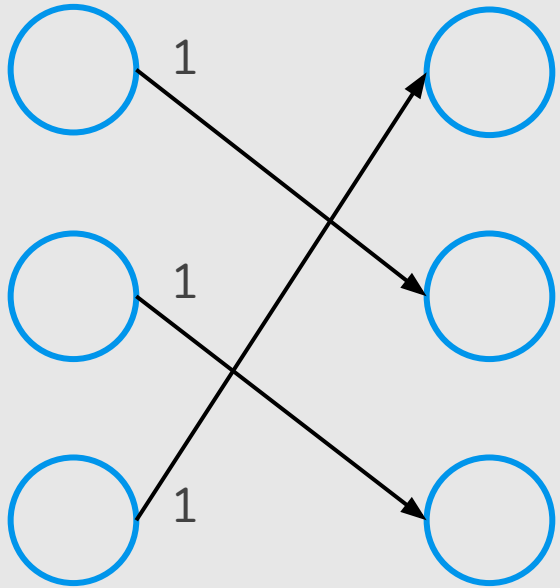
$\begin{bmatrix}0\\0\\1\end{bmatrix}$

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

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

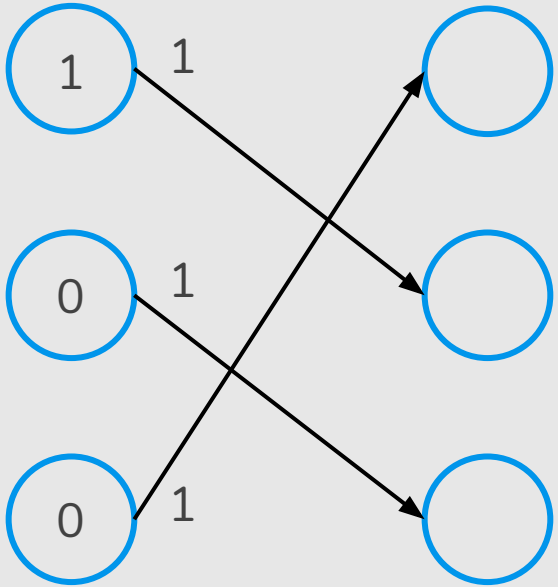
Simple (Recurrent) Neural Network

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



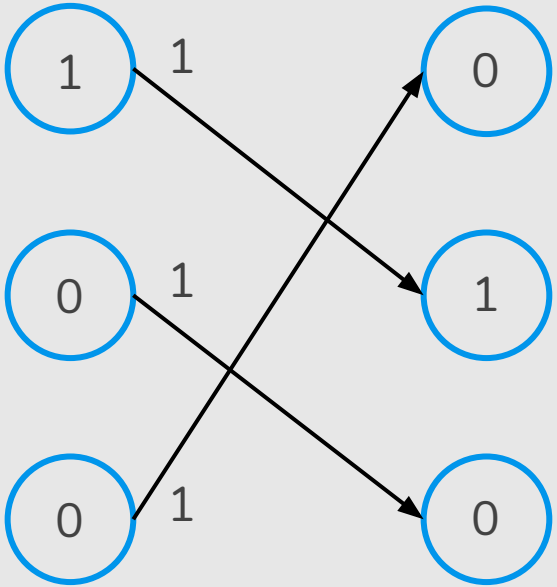
Simple (Recurrent) Neural Network

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



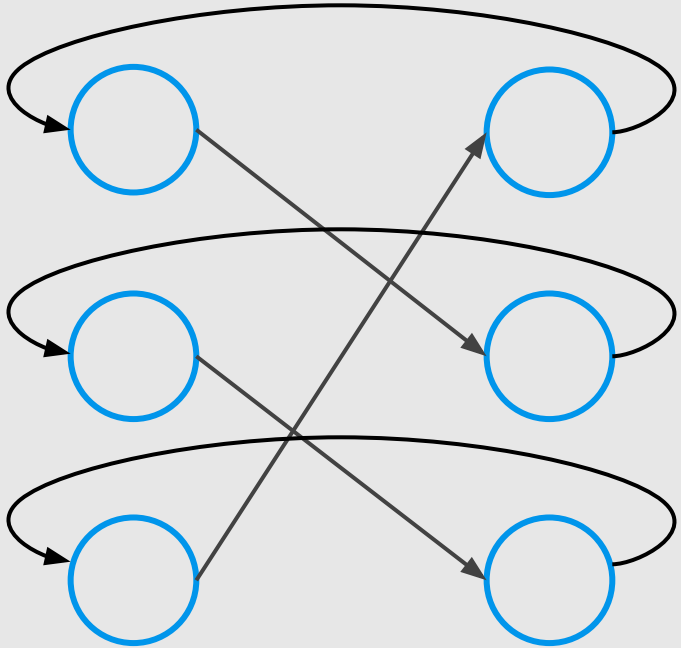
Simple (Recurrent) Neural Network

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



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

Simple (Recurrent) Neural Network



Weather

Weather

Sunny
Same as yesterday

Rain
Next dish

Cooking Schedule

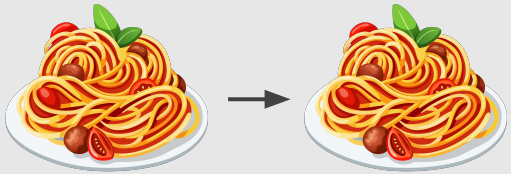
Monday Tuesday Wednesday Thursday Friday Saturday

Cooking Schedule

Monday Tuesday Wednesday Thursday Friday Saturday

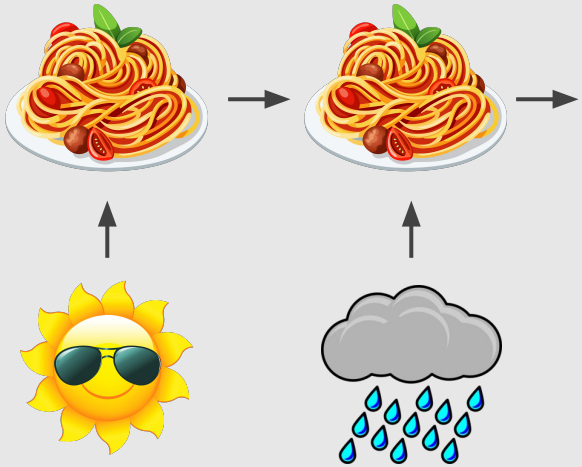
Cooking Schedule

Monday Tuesday Wednesday Thursday Friday Saturday



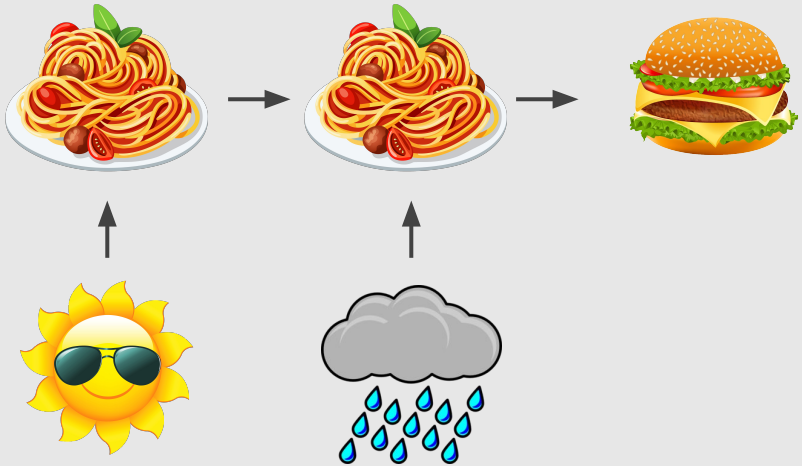
Cooking Schedule

Monday Tuesday Wednesday Thursday Friday Saturday



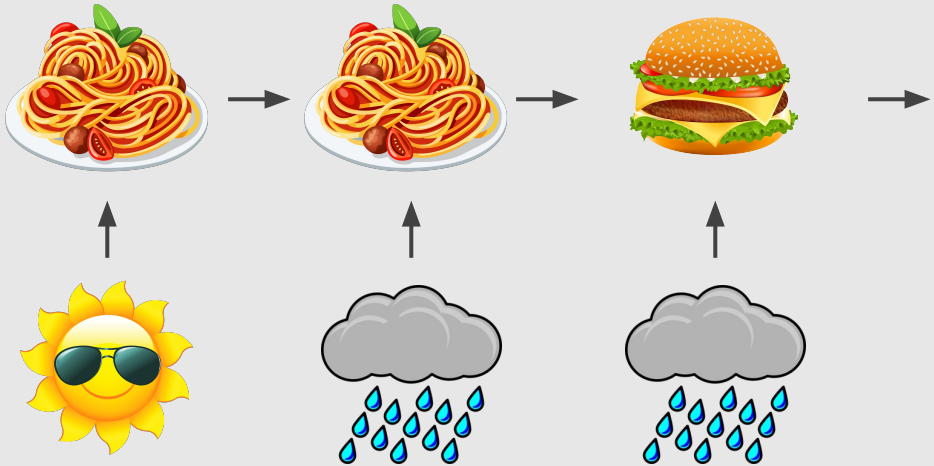
Cooking Schedule

Monday Tuesday Wednesday Thursday Friday Saturday



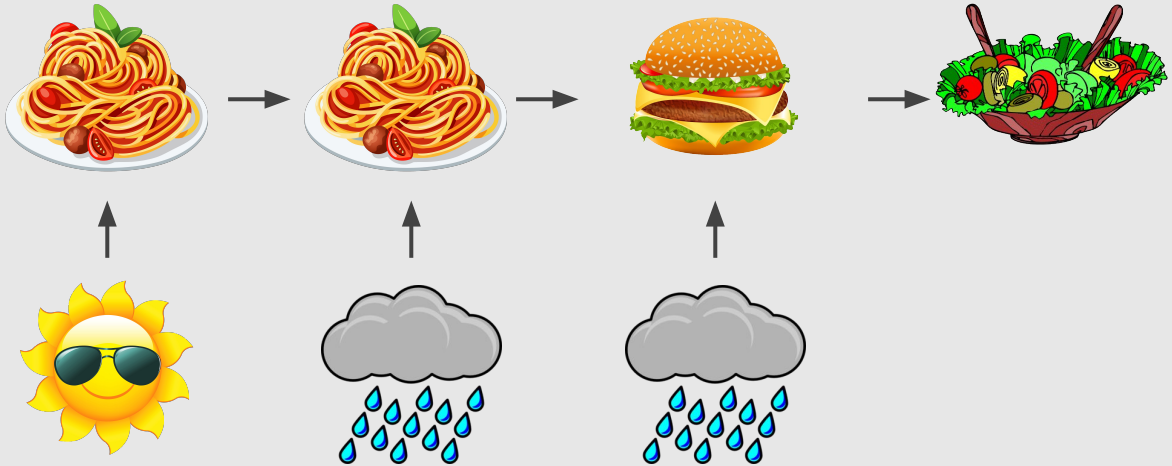
Cooking Schedule

Monday Tuesday Wednesday Thursday Friday Saturday



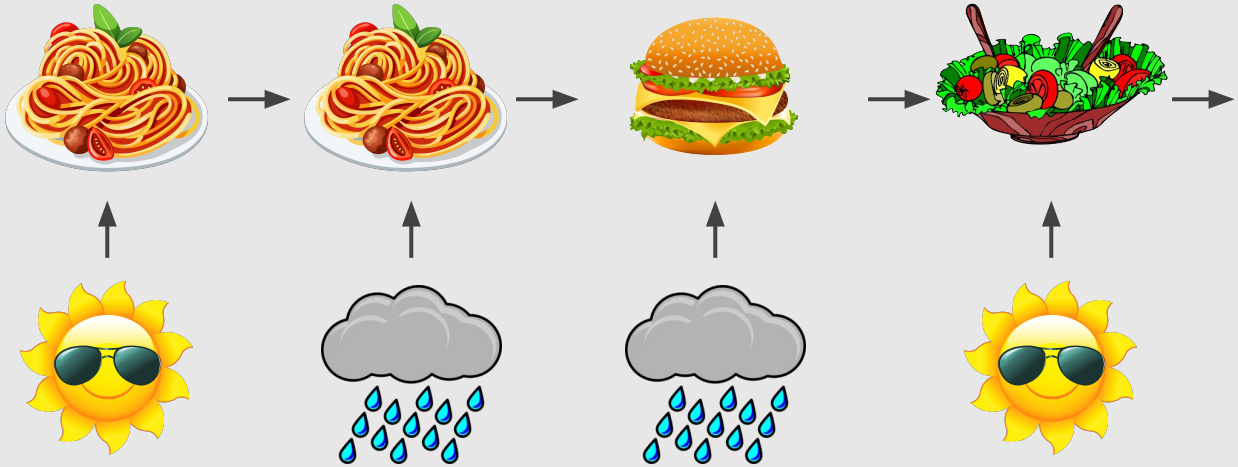
Cooking Schedule

Monday Tuesday Wednesday Thursday Friday Saturday



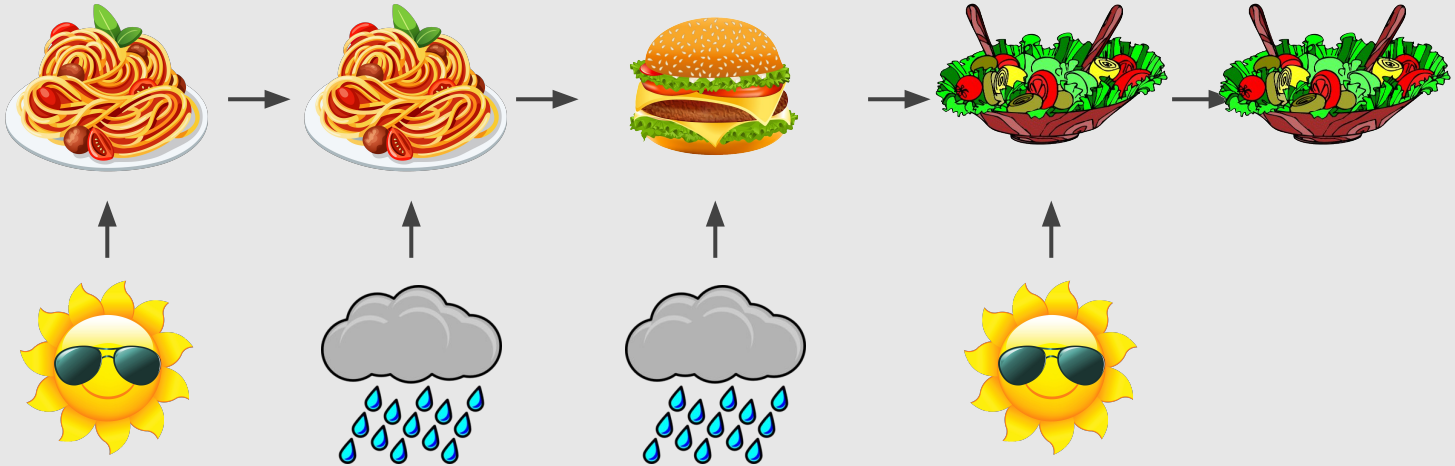
Cooking Schedule

Monday Tuesday Wednesday Thursday Friday Saturday



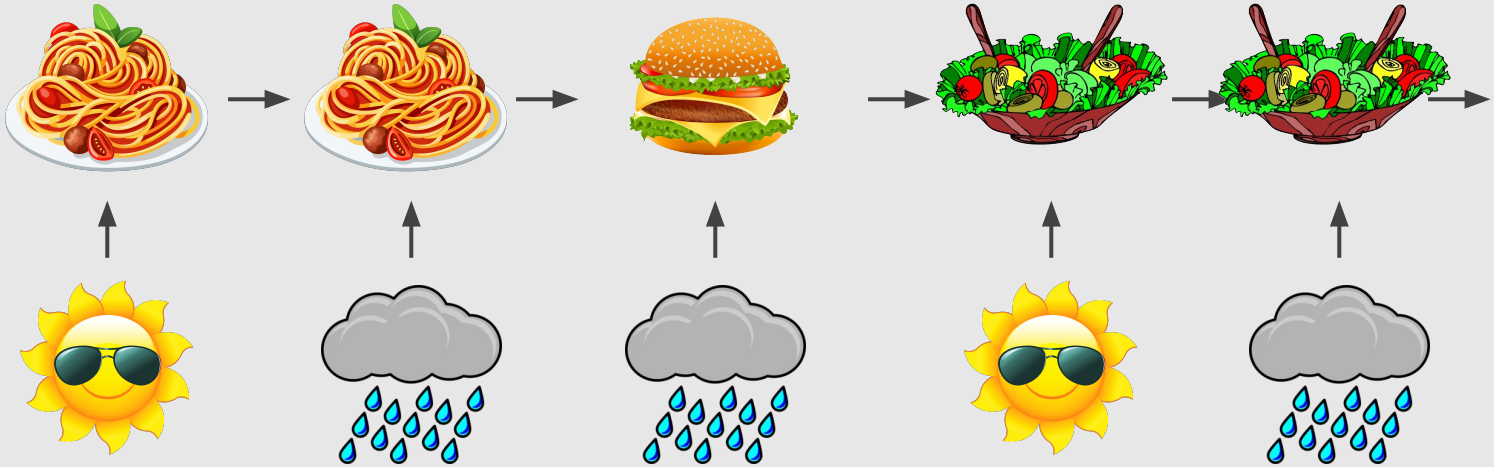
Cooking Schedule

Monday Tuesday Wednesday Thursday Friday Saturday

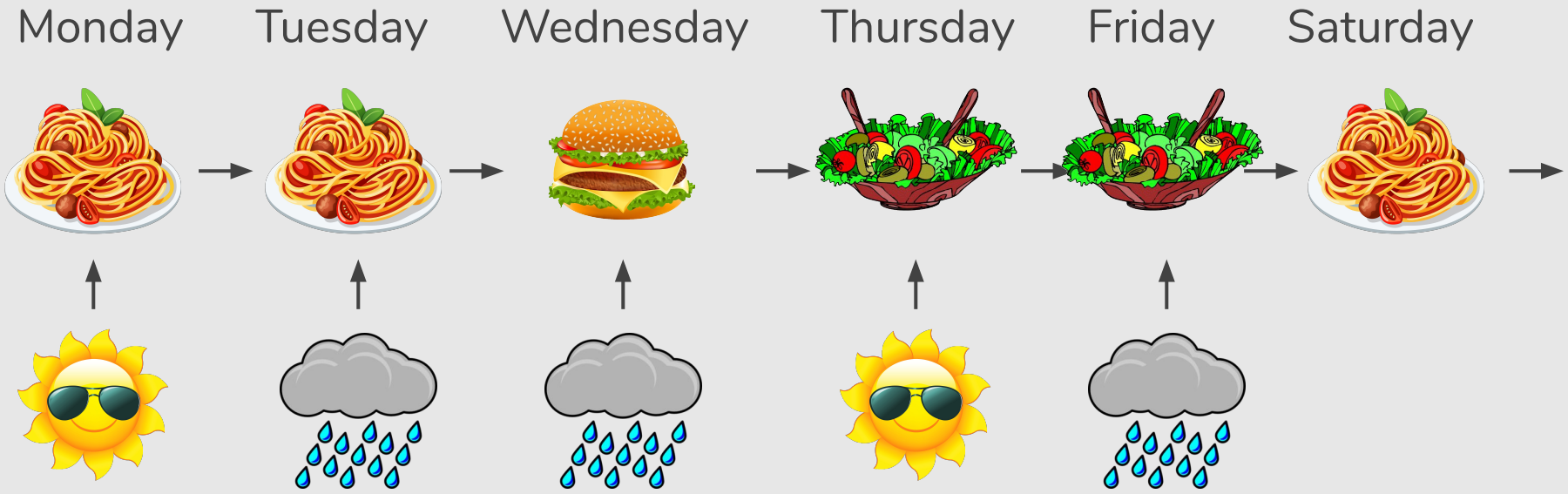


Cooking Schedule

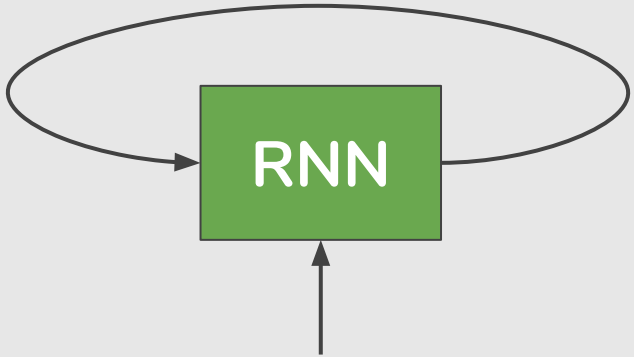
Monday Tuesday Wednesday Thursday Friday Saturday



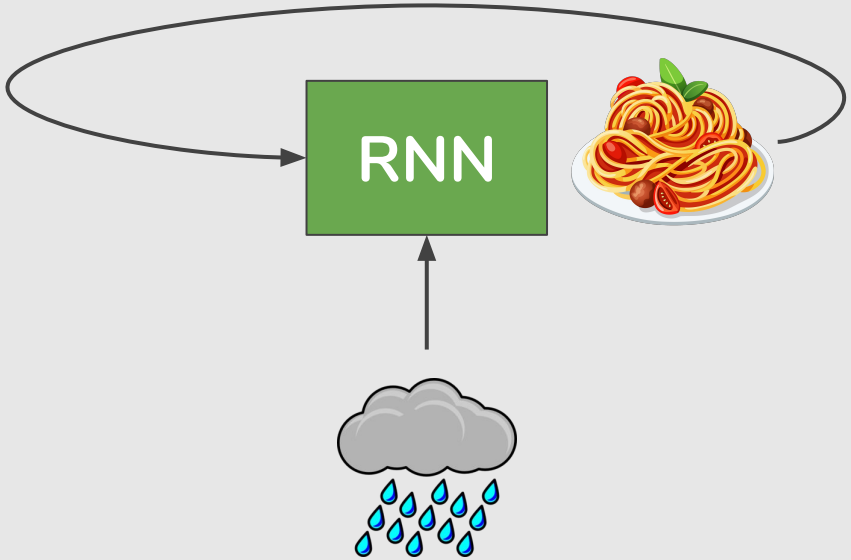
Cooking Schedule



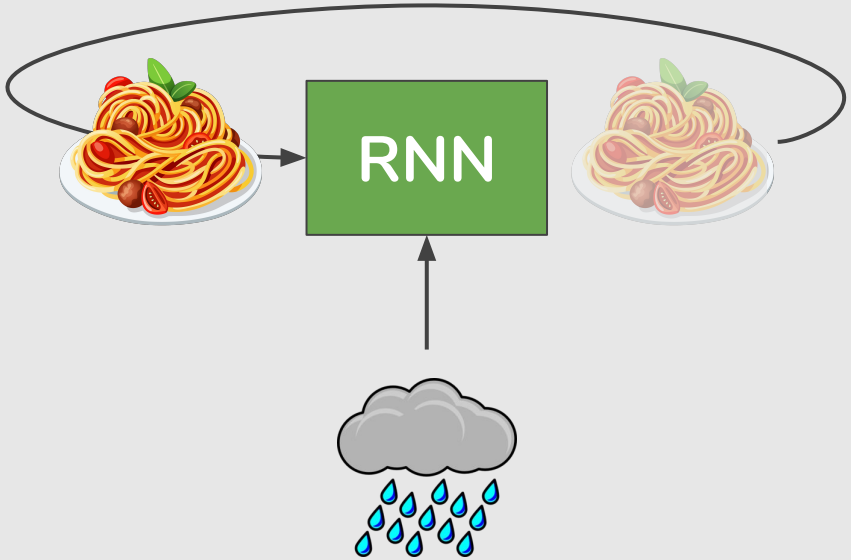
Recurrent Neural Network



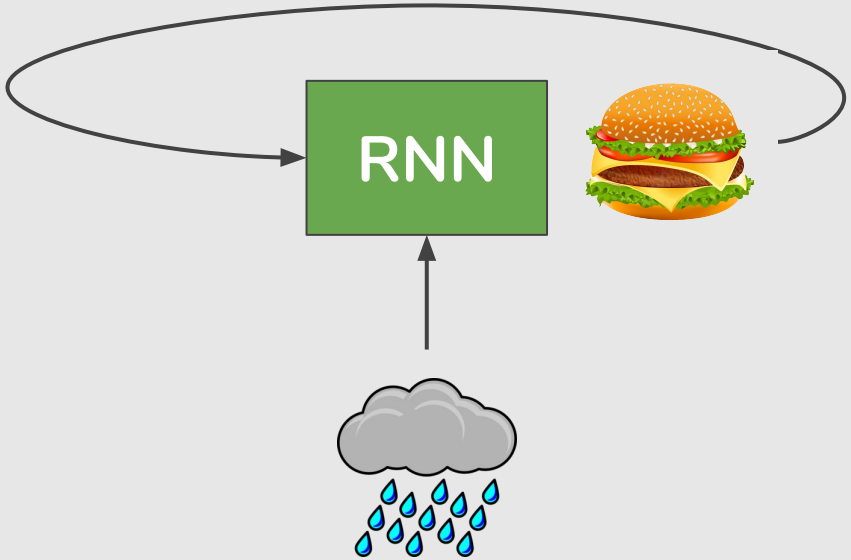
Recurrent Neural Network



Recurrent Neural Network

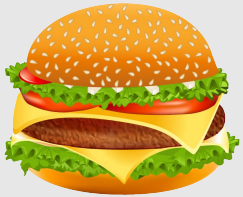


Recurrent Neural Network



Recurrent Neural Network

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



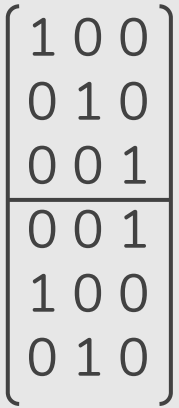
$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

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

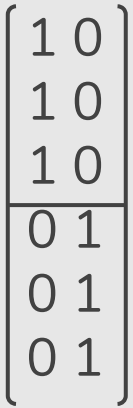
$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

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

Recurrent Neural Network



Food



Weather

Recurrent Neural Network (Food)

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

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

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

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

Food

Recurrent Neural Network (Food)

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

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

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

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

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

Food

Recurrent Neural Network (Food)

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

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

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

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

Food

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

$$=$$

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

same

next day

Recurrent Neural Network (Food)

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

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

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

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

Food

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

$$=$$

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

same

next day

Recurrent Neural Network (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

Recurrent Neural Network (Weather)

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

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

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

Weather

Recurrent Neural Network (Weather)

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

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

$$\left[\begin{array}{c|c}\begin{matrix}1&0\\1&0\\1&0\end{matrix} & \begin{matrix}0&1\\0&1\\0&1\end{matrix}\end{array}\right] \begin{bmatrix}1\\0\end{bmatrix}$$

Weather

$$=$$

$$\left[\begin{array}{c|c}\begin{matrix}1\\1\\1\end{matrix} & \begin{matrix}0\\0\\0\end{matrix}\end{array}\right]$$

same

next day

Recurrent Neural Network (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

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

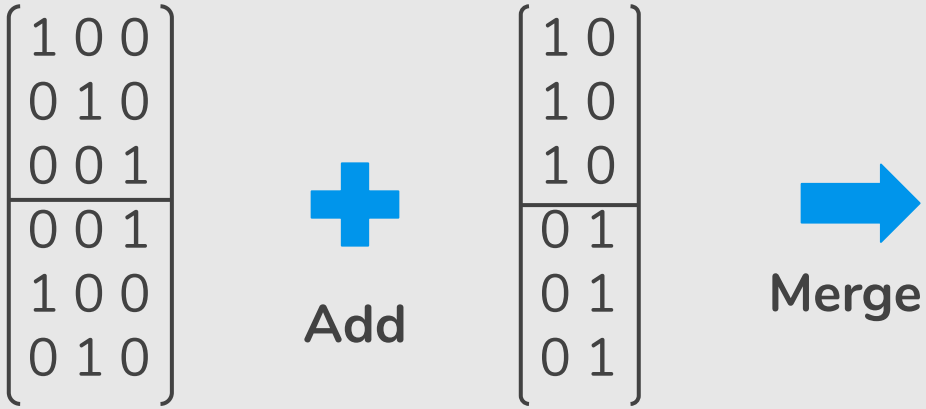
$$=$$

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

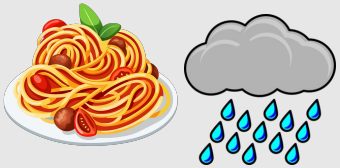
same

next day

Recurrent Neural Network



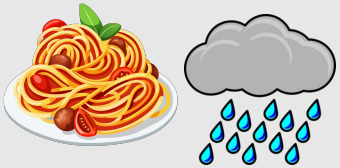
Recurrent Neural Network



(Add)

Recurrent Neural Network

(Add)



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

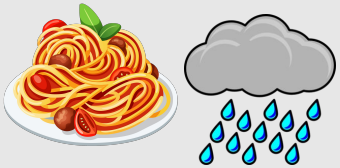
same

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

next day

Recurrent Neural Network

(Add)



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

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

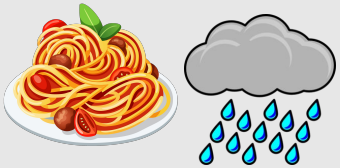
Add

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

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$ next day

Recurrent Neural Network

(Add)



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

same

next day

Add

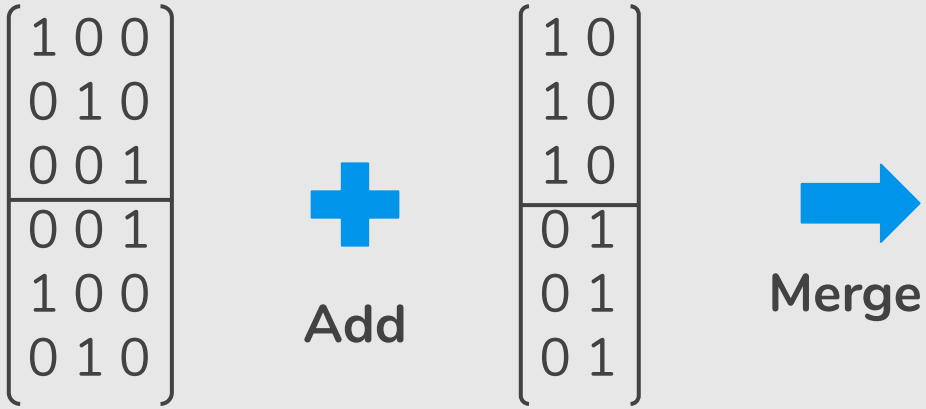
$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ \hline 1 \\ 1 \\ 1 \end{bmatrix}$$

same

next day

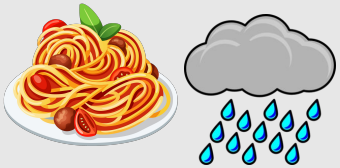
$$= \begin{bmatrix} 1 \\ 0 \\ 0 \\ \hline 1 \\ 2 \\ 1 \end{bmatrix}$$

Recurrent Neural Network



Recurrent Neural Network

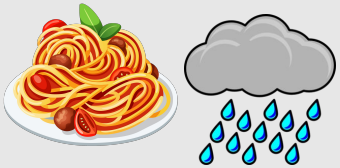
(Merge)

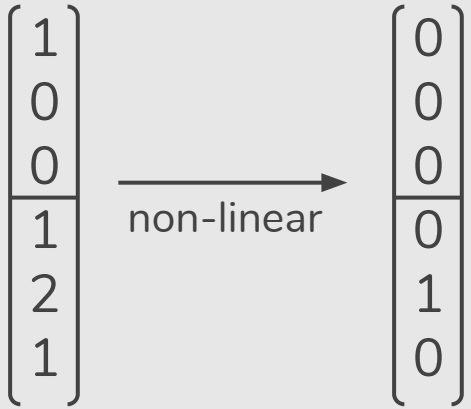


$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ \hline 1 \\ 2 \\ 1 \end{bmatrix} \xrightarrow{\text{non-linear}}$$

Recurrent Neural Network

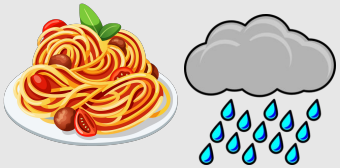
(Merge)

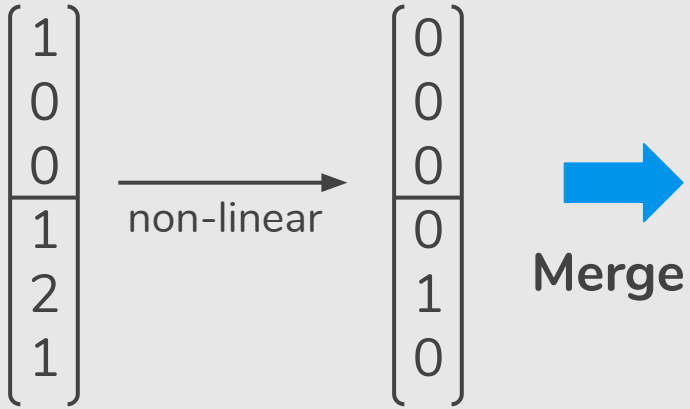




Recurrent Neural Network

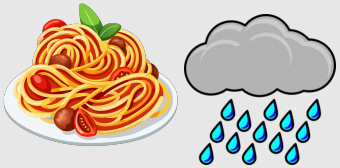
(Merge)





Recurrent Neural Network

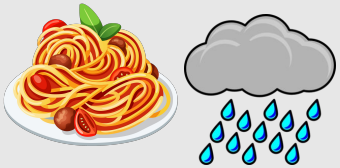
(Merge)



$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ \hline 1 \\ 2 \\ 1 \end{bmatrix} \xrightarrow{\text{non-linear}} \begin{bmatrix} 0 \\ 0 \\ 0 \\ \hline 0 \\ 1 \\ 0 \end{bmatrix} \Rightarrow \text{Merge} \quad \begin{bmatrix} 0+0 \\ 0+1 \\ 0+0 \end{bmatrix} =$$

Recurrent Neural Network

(Merge)

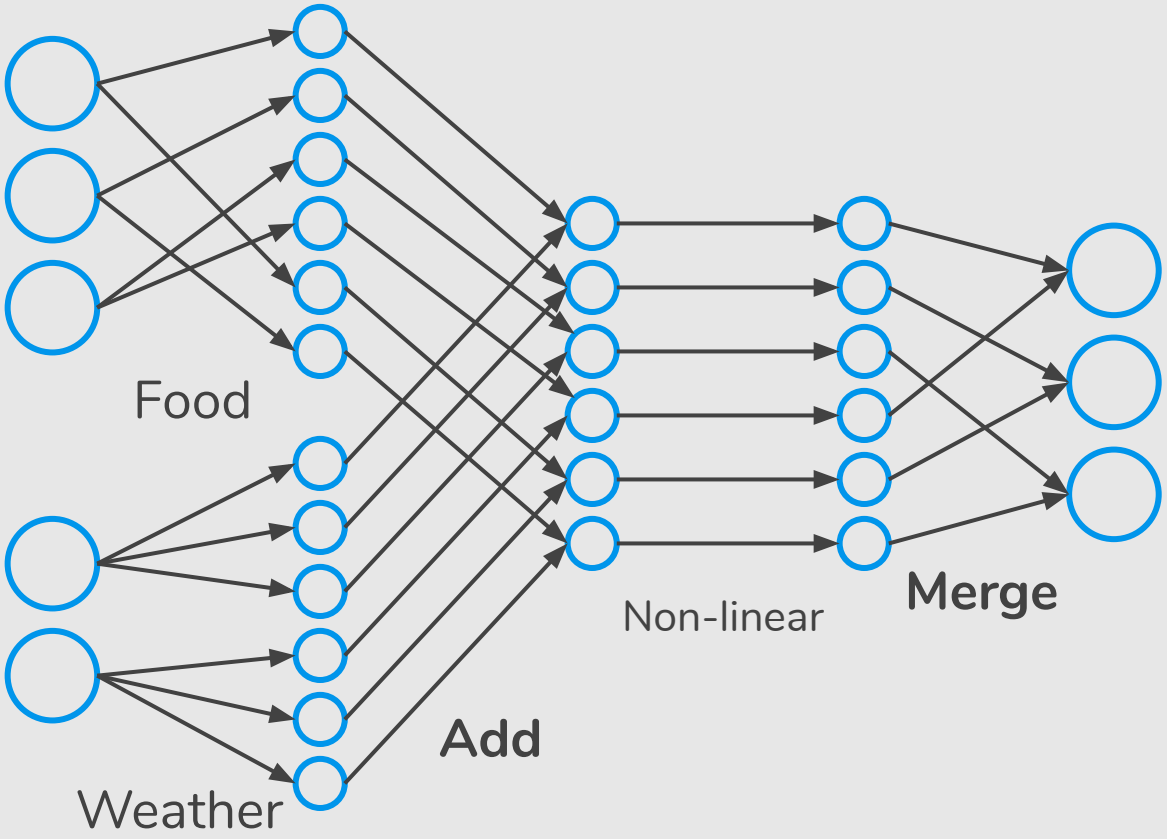


$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ \hline 1 \\ 2 \\ 1 \end{bmatrix} \xrightarrow{\text{non-linear}} \begin{bmatrix} 0 \\ 0 \\ 0 \\ \hline 0 \\ 1 \\ 0 \end{bmatrix}$$

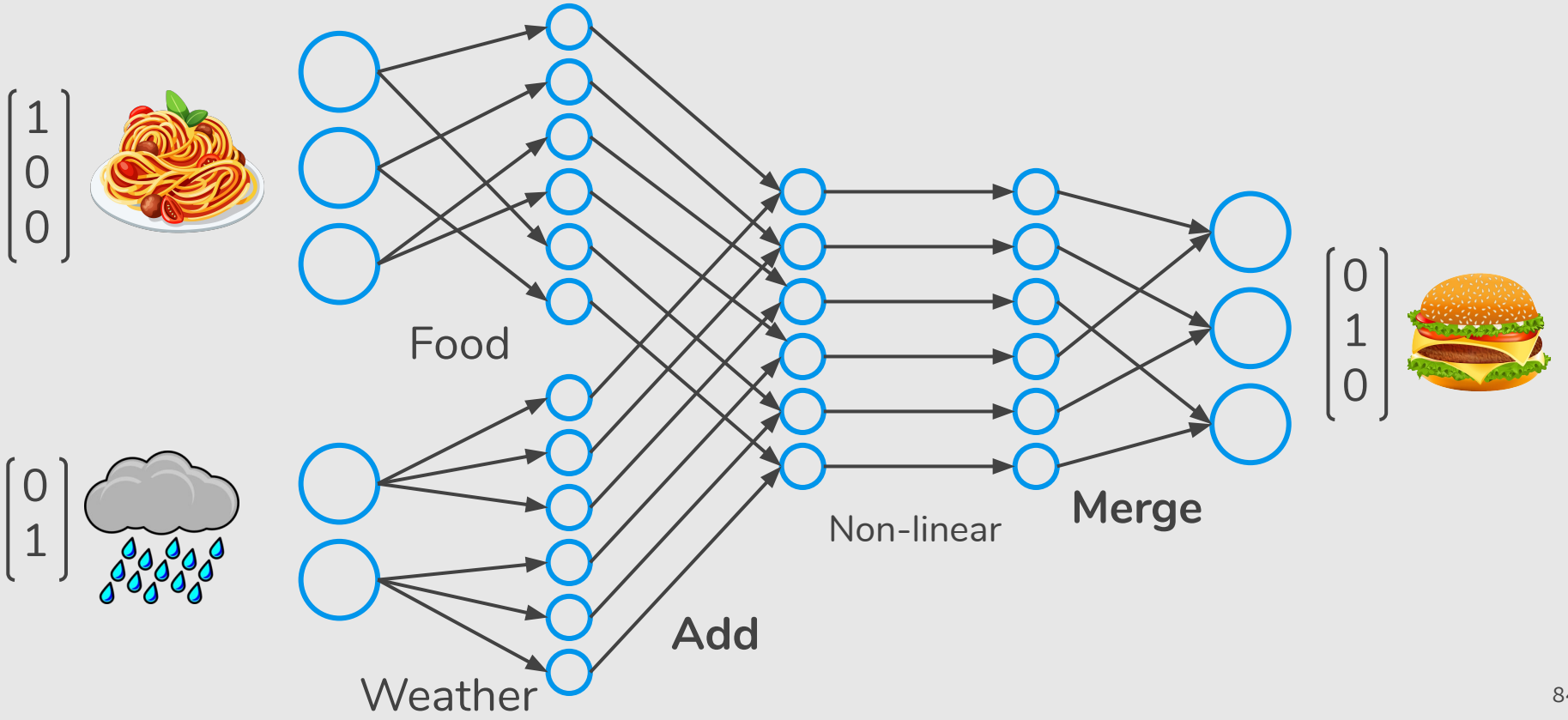
Merge

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

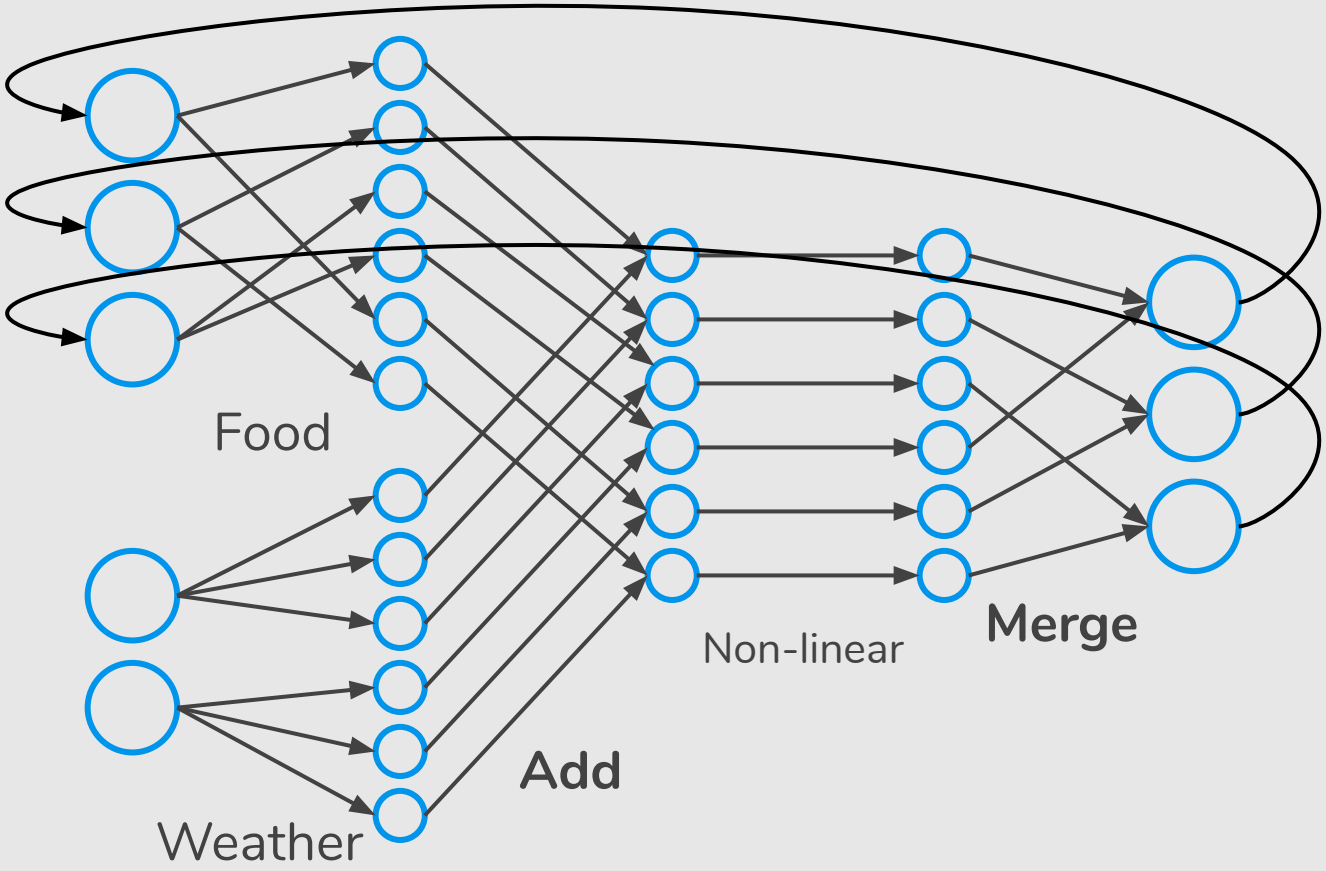
Recurrent Neural Network



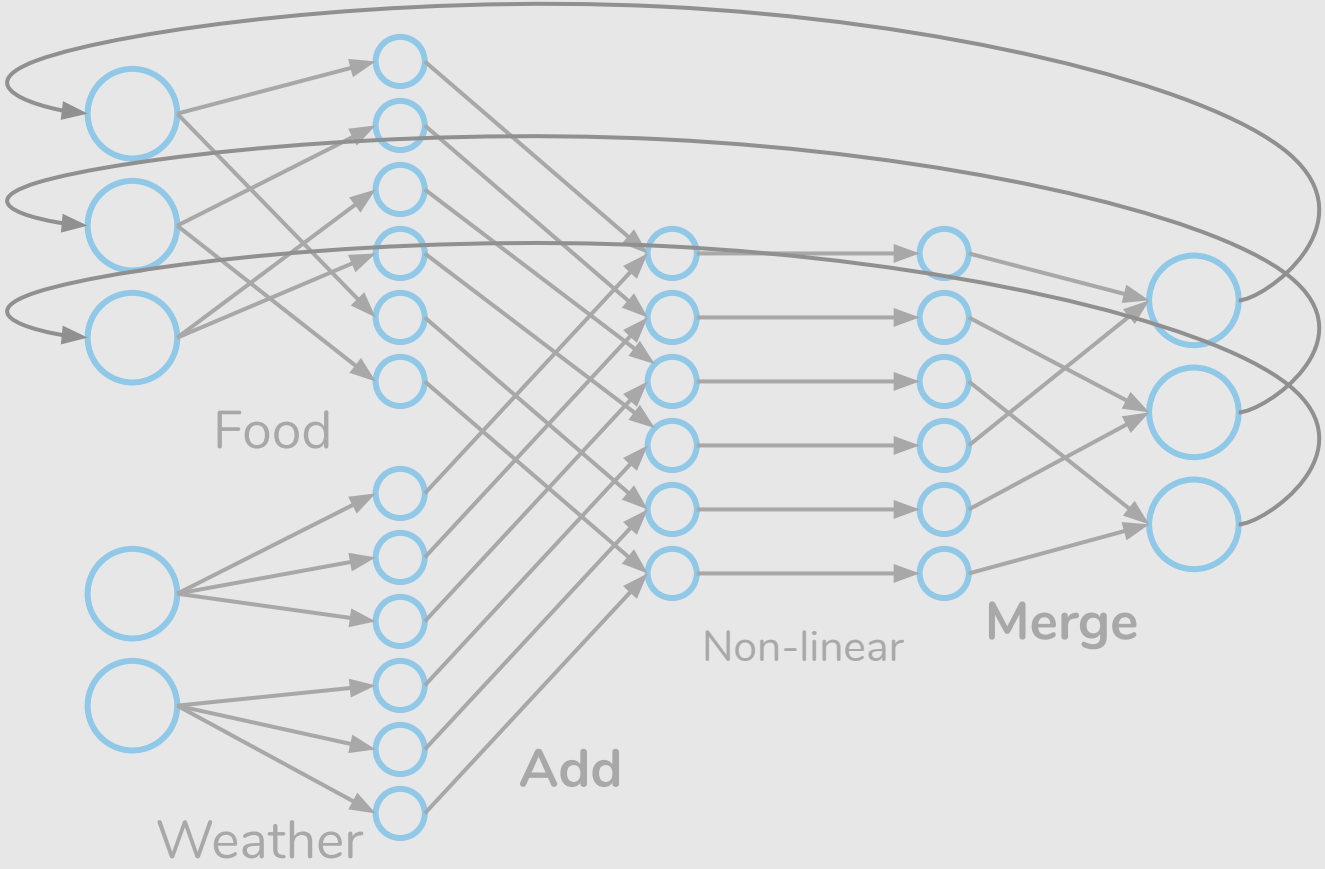
Recurrent Neural Network



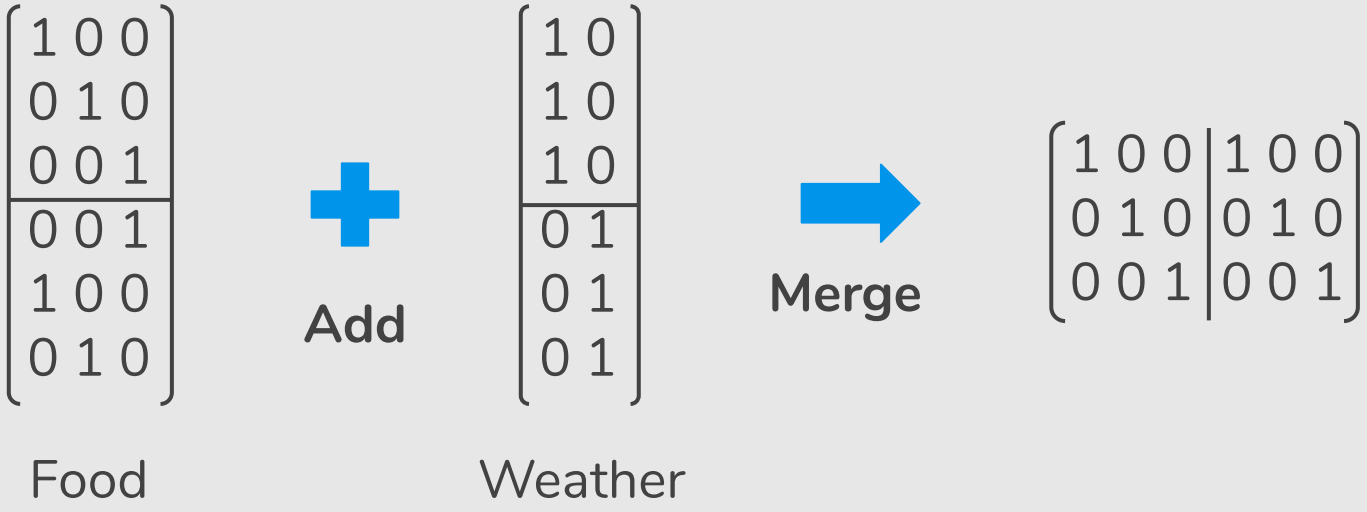
Recurrent Neural Network



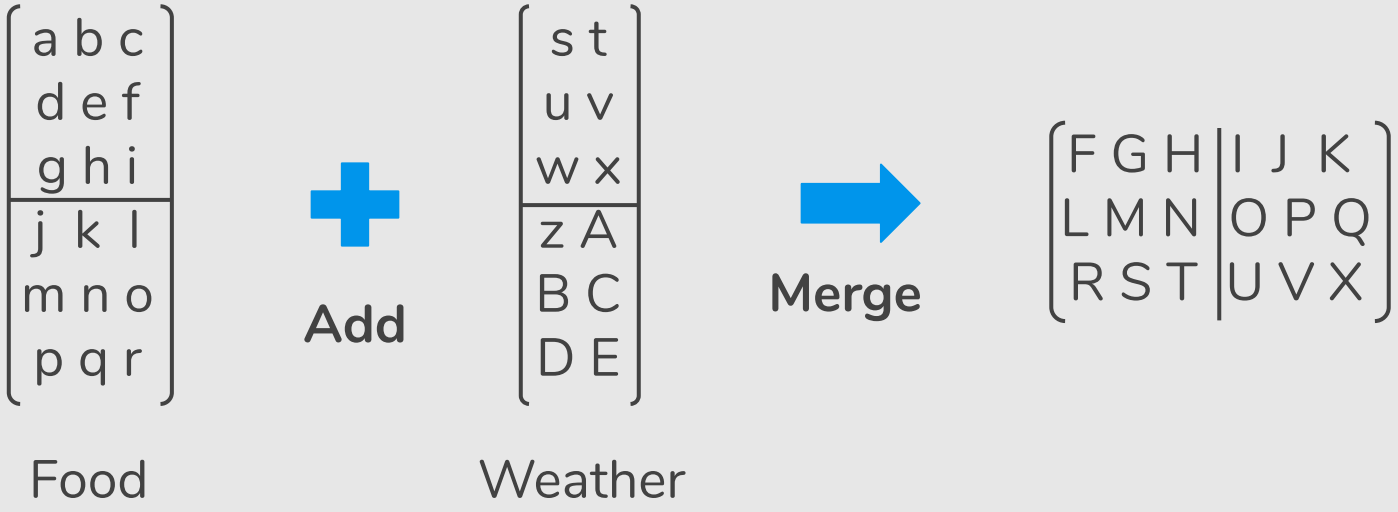
How to train the RNN?



RNN


1 0 0
0 1 0
0 0 1
0 0 1
1 0 0
0 1 0
Food
Add
1 0
1 0
1 0
0 1
0 1
0 1
Weather
Merge
1 0 0 | 1 0 0
0 1 0 | 0 1 0
0 0 1 | 0 0 1

Start with Random Weights



Food

Weather

Today's Agenda

- Recurrent Neural Networks
  - An Intuitive Explanation
  - **A More Formal Explanation**
  - Vanilla vs LSTMs

RNNs:
A More Formal Explanation

Recurrent Neural Networks: Process Sequences

one to one

Vanilla Neural Networks

Recurrent Neural Networks: Process Sequences

one to one

Vanilla Neural Networks

one to many

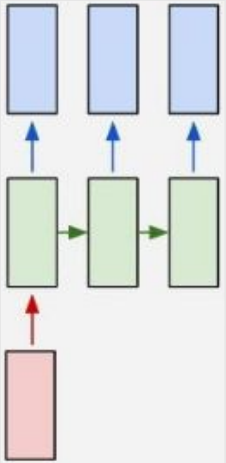


Image Captioning
image ⇒ seq. words

Recurrent Neural Networks: Process Sequences

one to one

Vanilla Neural Networks

one to many

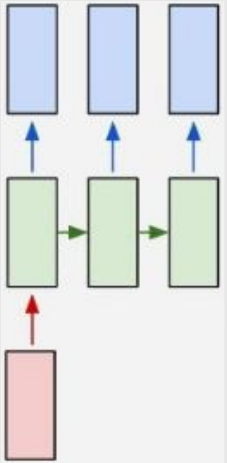
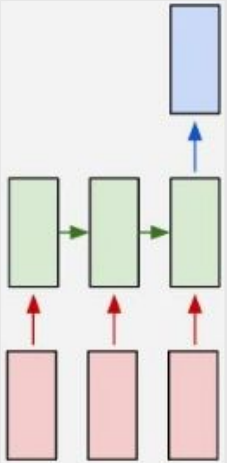


Image Captioning
image ⇒ seq. words

many to one



Sentiment Classification
seq. words ⇒ sentiment

Recurrent Neural Networks: Process Sequences

one to one

Vanilla Neural Networks

one to many

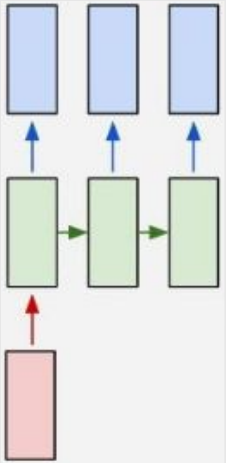
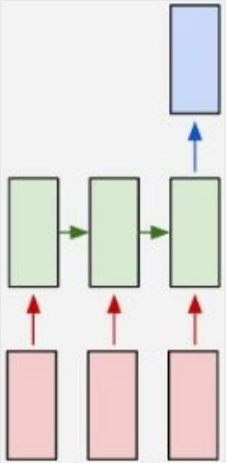


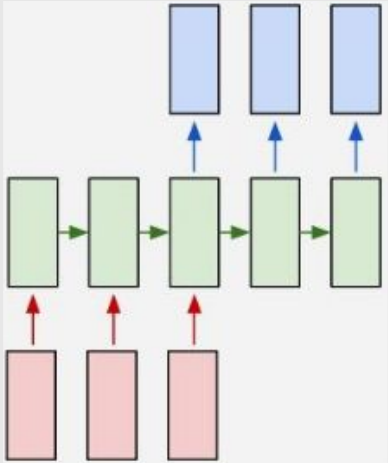
Image Captioning
image ⇒ seq. words

many to one



Sentiment Classification
seq. words ⇒ sentiment

many to many



Machine Translation
seq. words ⇒ seq. of words

Recurrent Neural Networks: Process Sequences

one to one

Vanilla Neural Networks

one to many

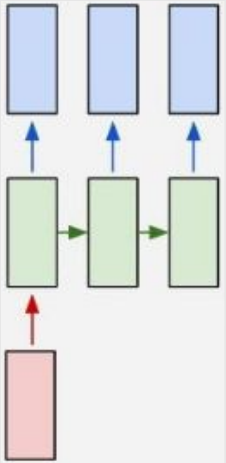
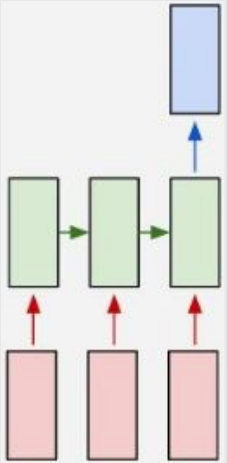


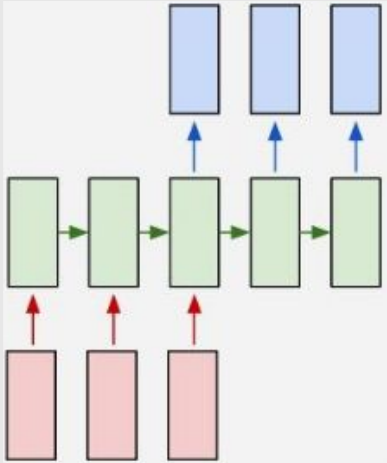
Image Captioning
image ⇒ seq. words

many to one



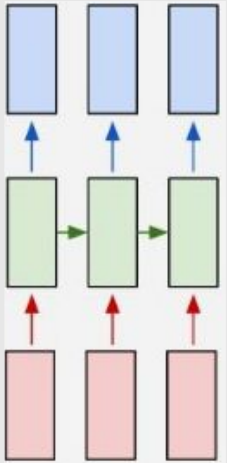
Sentiment Classification
seq. words ⇒ sentiment

many to many



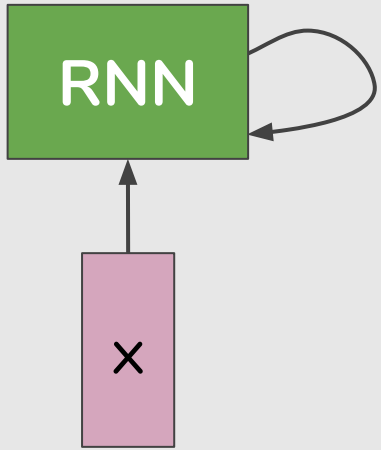
Machine Translation
seq. words ⇒ seq. of words

many to many

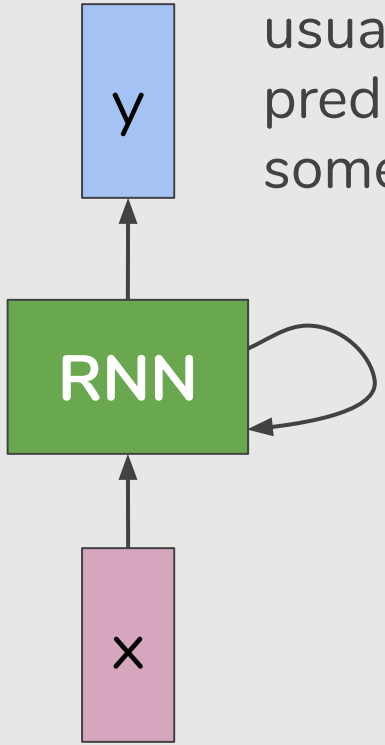


Video classification on frame level

Recurrent Neural Network



Recurrent Neural Network



usually want to predict a vector at some time steps

Recurrent Neural Network

We can process a sequence of vectors **x** by applying a **recurrence formula** at every time step:

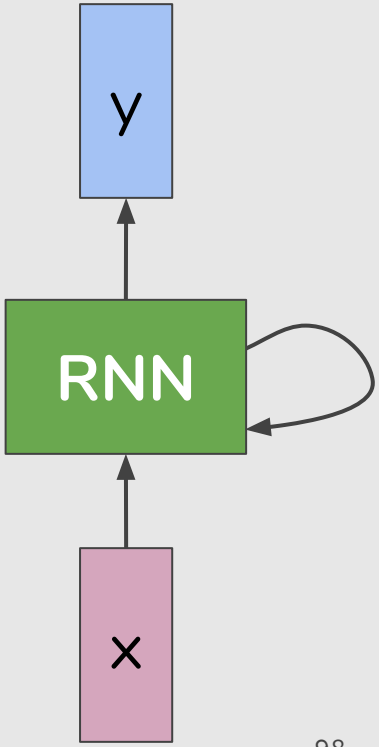
$$h_t = f_W(h_{t-1}, x_t)$$

new state

old state

input vector at some time step

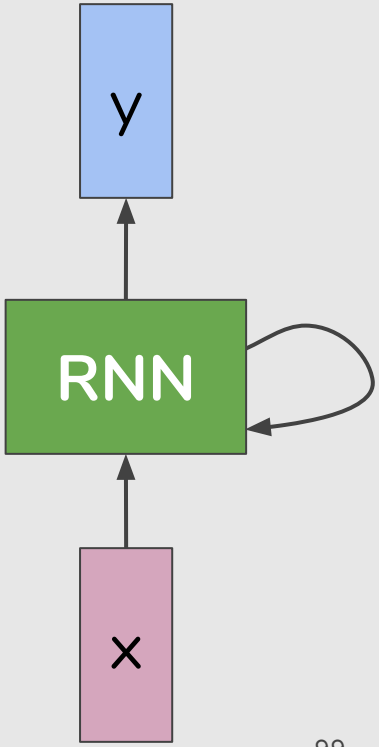
some function with parameters W



Recurrent Neural Network

We can process a sequence of vectors **x** by applying a **recurrence formula** at every time step:

$$h_t = f_W(h_{t-1}, x_t)$$



Notice: the same function and the same set of parameters are used at every time step.

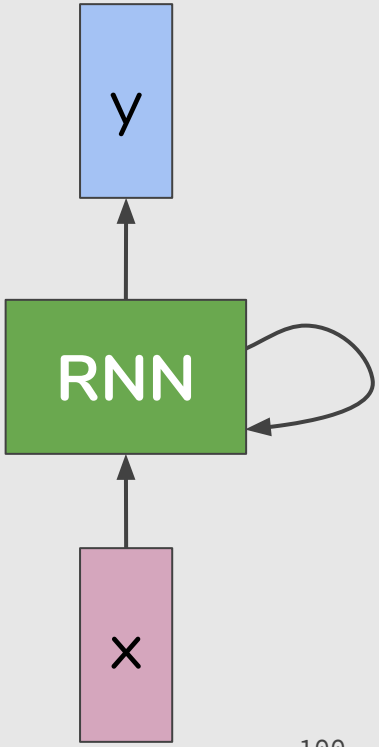
Recurrent Neural Network

The state consists of a single “hidden” vector **h**:

$$h_t = f_W(h_{t-1}, x_t)$$

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

$$y_t = W_{hy}h_t$$



To be continued ....