



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

**RNN**

MAHDI FARAHANI



# Neural Networks

**01**

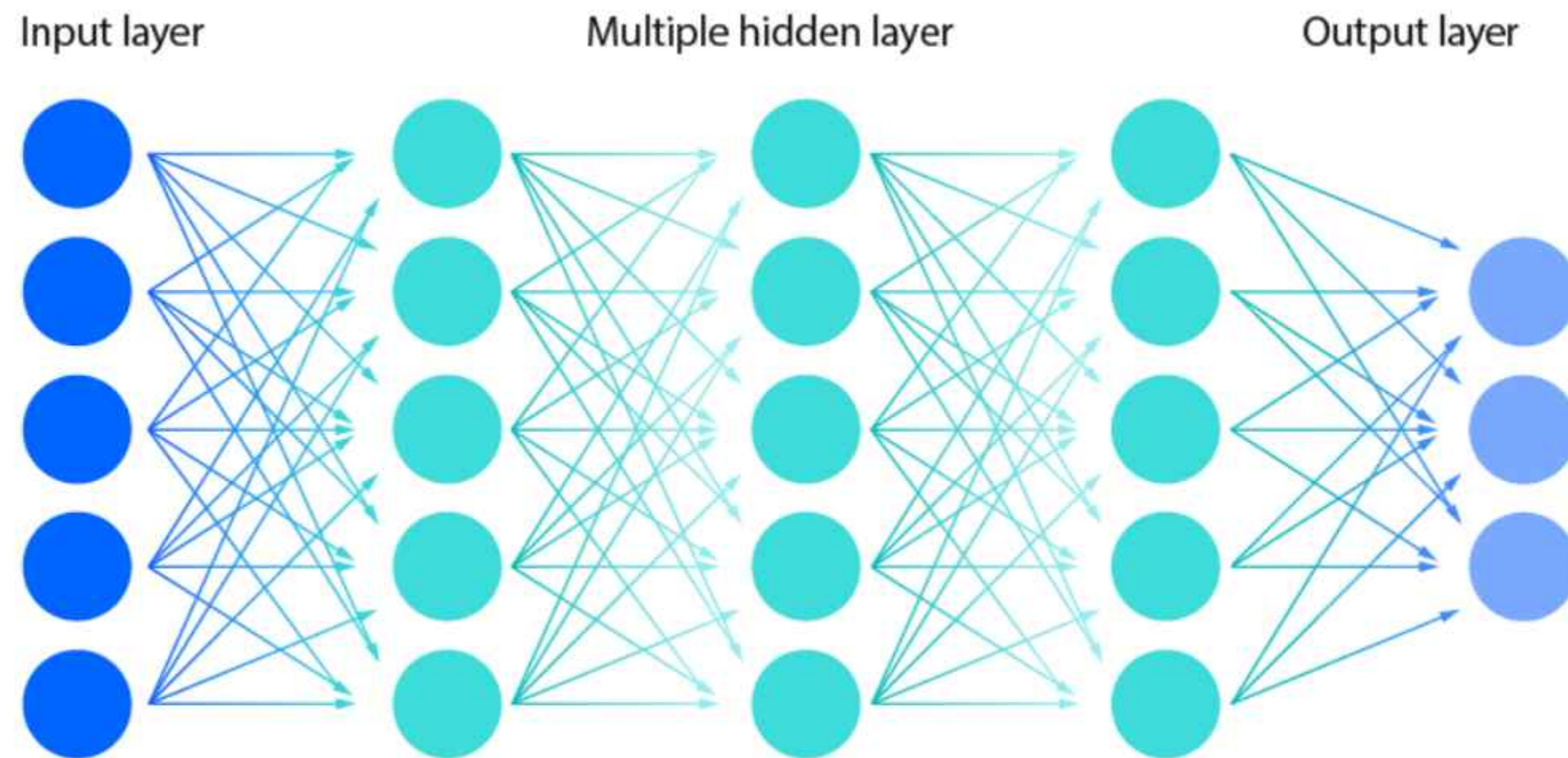
MLP

**02**

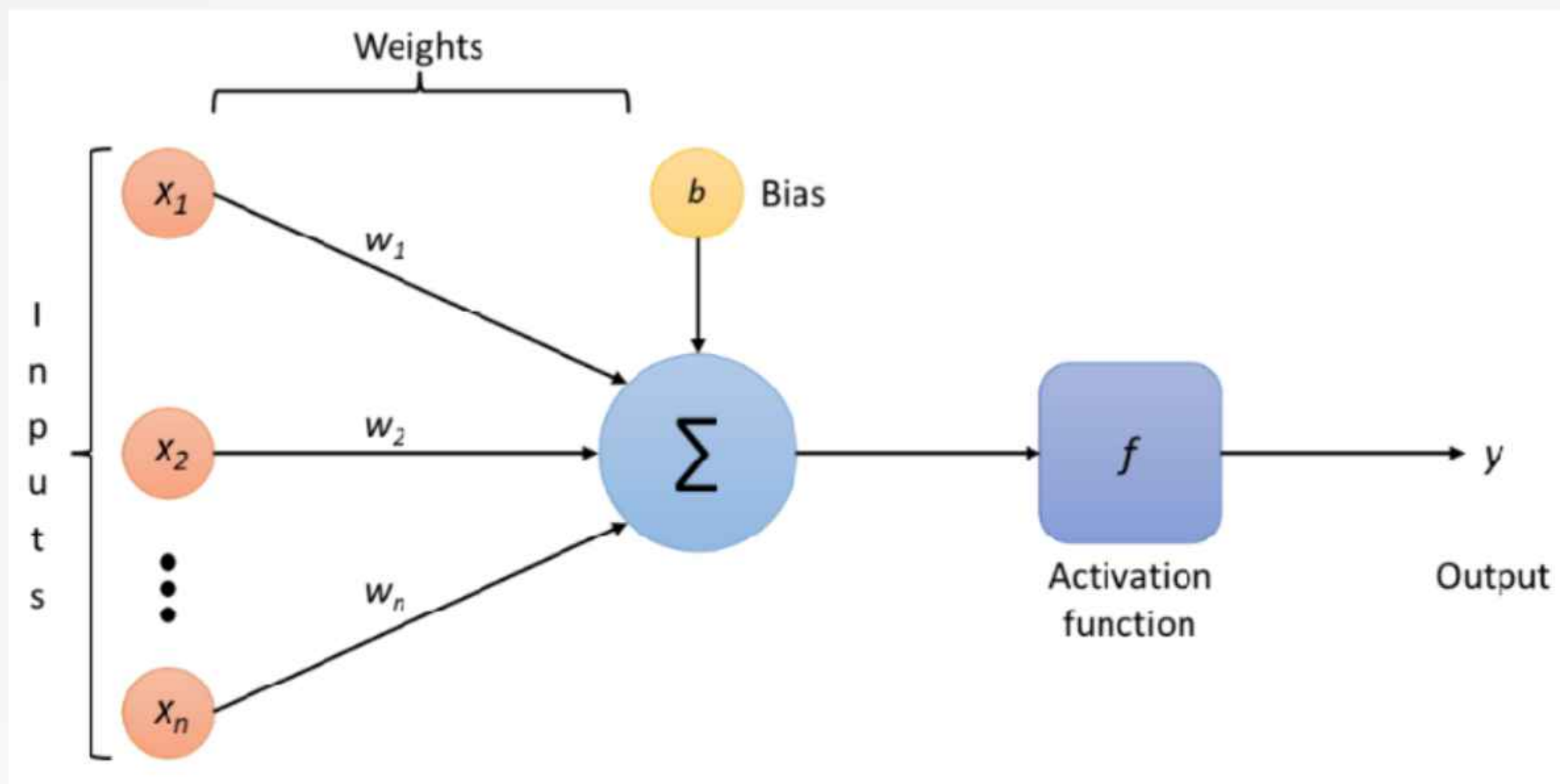
RNN

# MLP

## Deep neural network

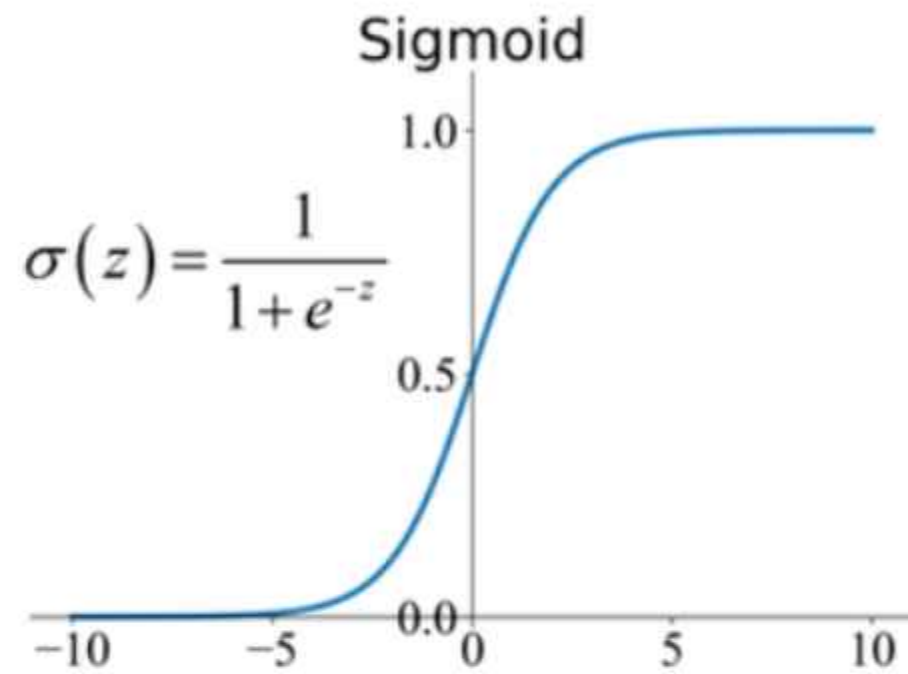


# NEURON

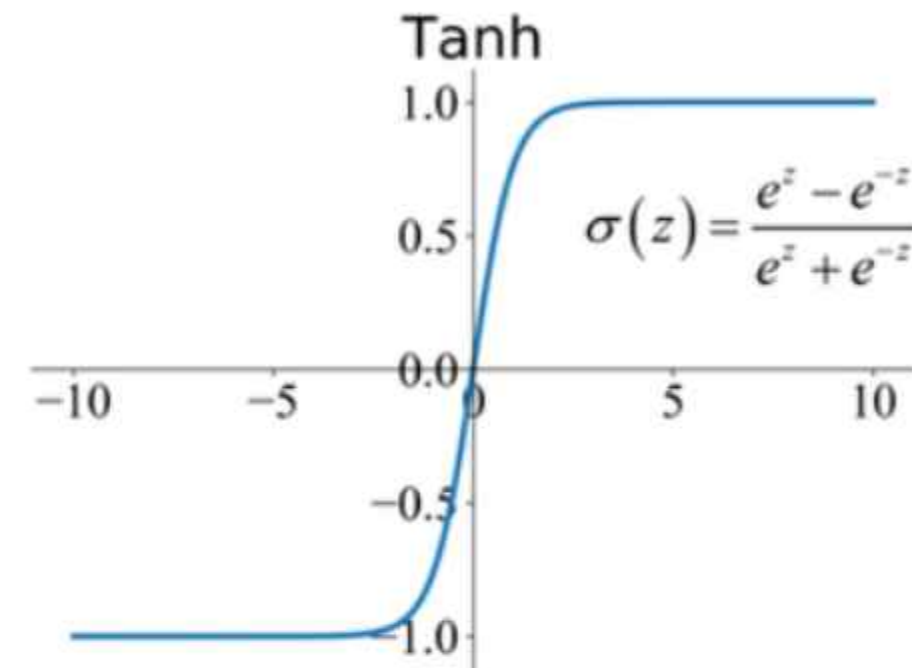




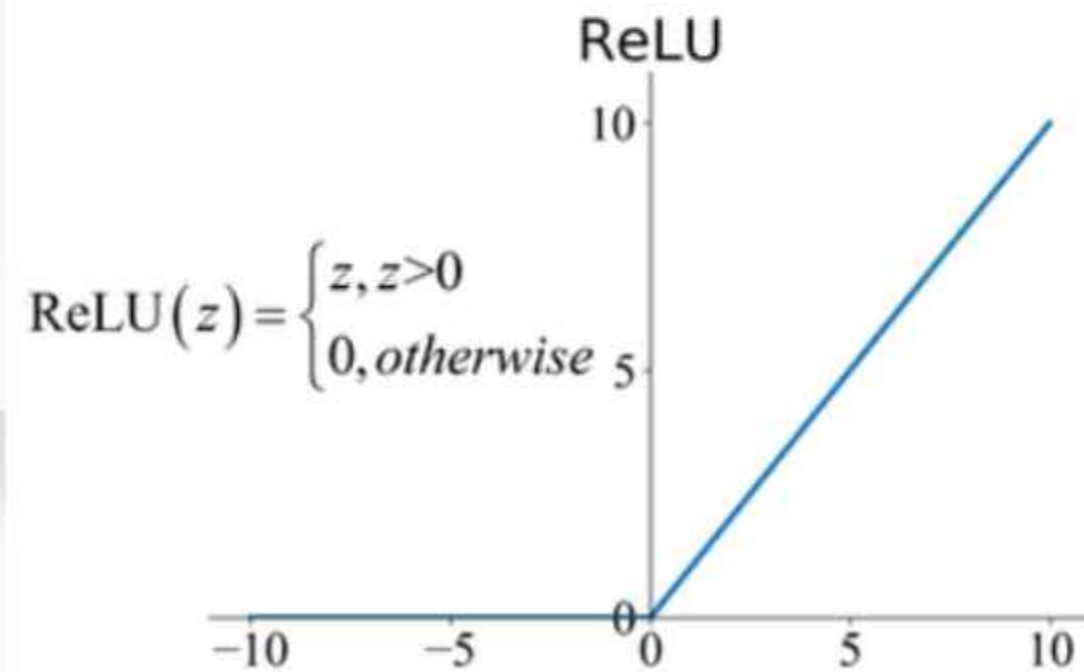
# ACTIVATION FUNCTION



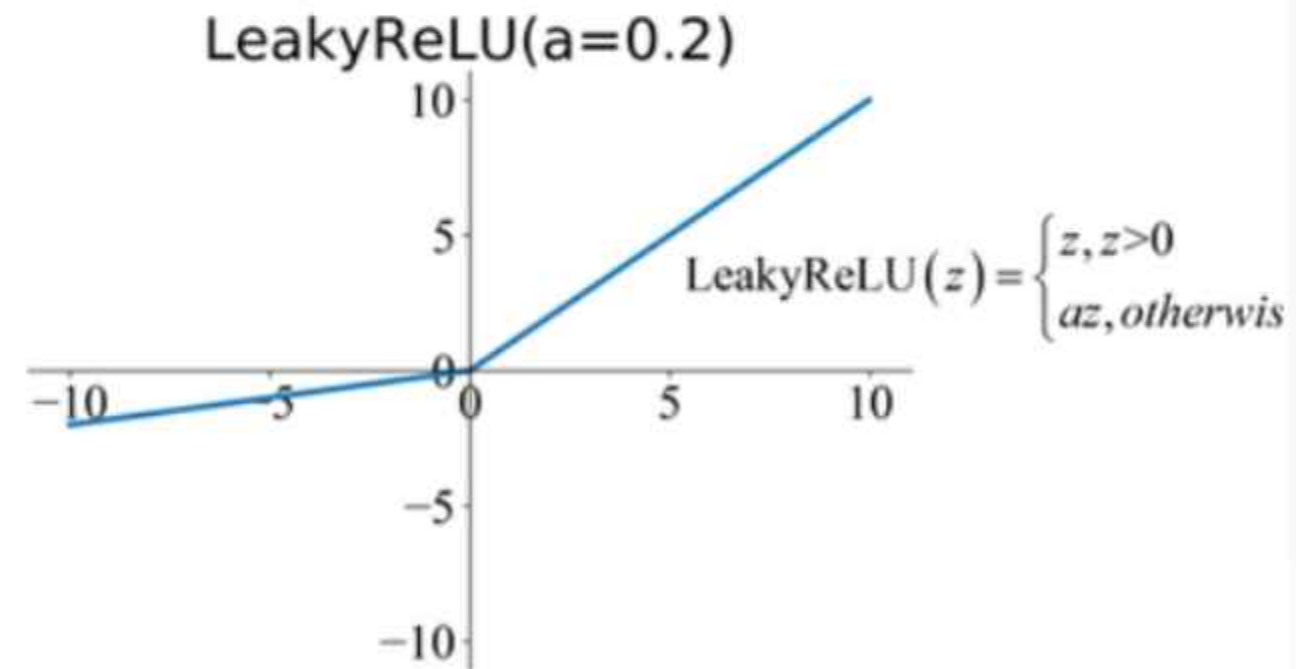
(a)



(b)

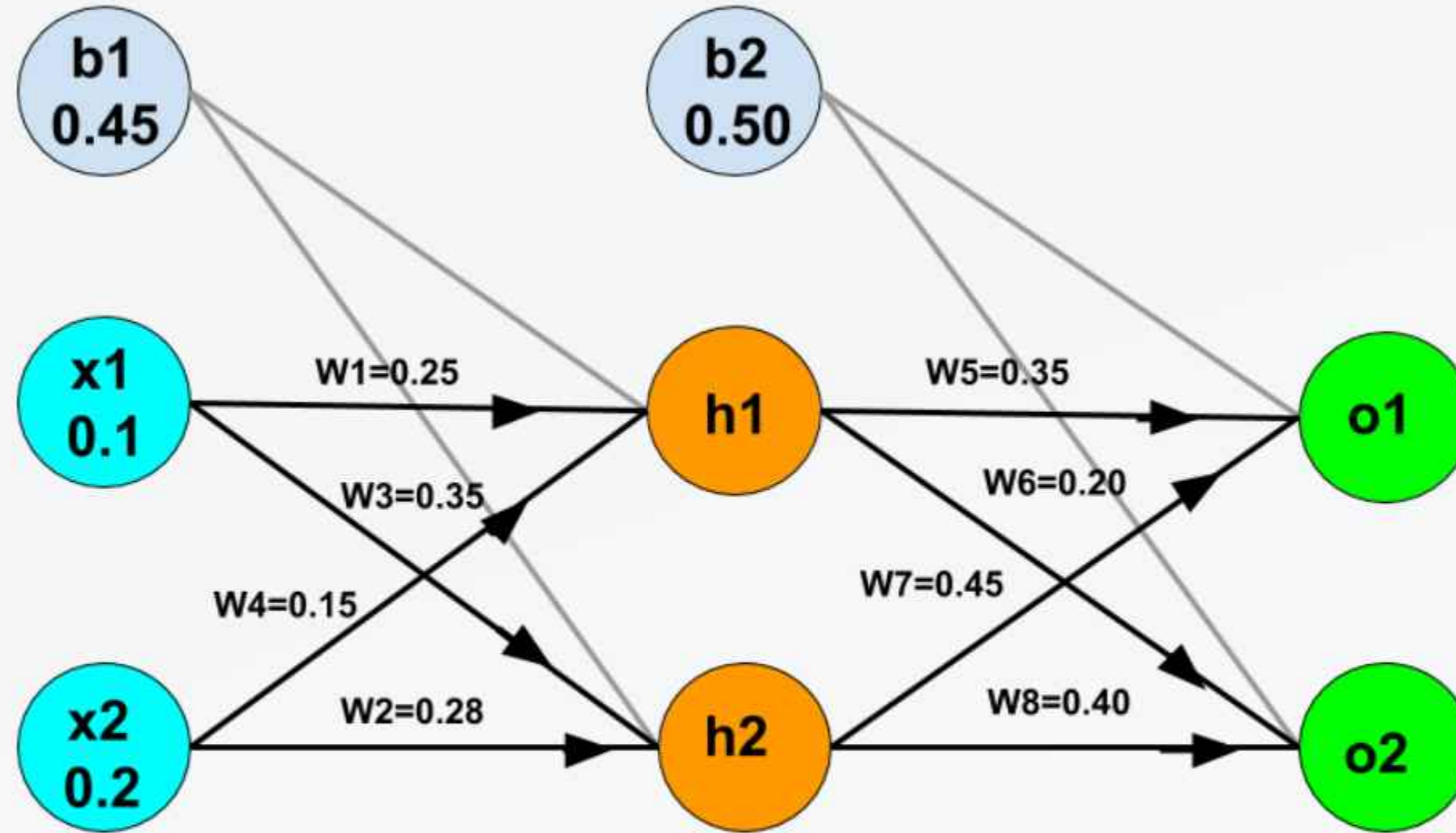


(c)



(d)

# FEEDFORWARD



**LOSS  
FUNCTIONS**

# BACKPROPAGATION





# LOSS FUNCTIONS

## Regression Losses

Mean Bias Error	Captures average bias in prediction. But is rarely used for training.	$\mathcal{L}_{MBE} = \frac{1}{N} \sum_{i=1}^N (y_i - f(x_i))$
Mean Absolute Error	Measures absolute average bias in prediction. Also called L1 Loss.	$\mathcal{L}_{MAE} = \frac{1}{N} \sum_{i=1}^N  y_i - f(x_i) $
Mean Squared Error	Average squared distance between actual and predicted. Also called L2 Loss.	$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - f(x_i))^2$
Root Mean Squared Error	Square root of MSE. Loss and dependent variable have same units.	$\mathcal{L}_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - f(x_i))^2}$
Huber Loss	A combination of MSE and MAE. It is parametric loss function.	$\mathcal{L}_{Huberloss} = \begin{cases} \frac{1}{2}(y_i - f(x_i))^2 & :  y_i - f(x_i)  \leq \delta \\ \delta( y_i - f(x_i)  - \frac{1}{2}\delta) & : otherwise \end{cases}$
Log Cosh Loss	Similar to Huber Loss + non-parametric. But computationally expensive.	$\mathcal{L}_{LogCosh} = \frac{1}{N} \sum_{i=1}^N \log(\cosh(f(x_i) - y_i))$

## Classification Losses (Binary + Multi-class)

Binary Cross Entropy (BCE)	Loss function for binary classification tasks.	$\mathcal{L}_{BCE} = \frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(x_i)) + (1 - y_i) \cdot \log(1 - p(x_i))$
Hinge Loss	Penalizes wrong and right (but less confident) predictions. Commonly used in SVMs.	$\mathcal{L}_{Hinge} = \max(0, 1 - (f(x) \cdot y))$
Cross Entropy Loss	Extension of BCE loss to multi-class classification.	$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \cdot \log(f(x_{ij}))$ <small><math>N</math> : samples; <math>M</math> : classes</small>
KL Divergence	Minimizes the divergence between predicted and true probability distribution	$\mathcal{L}_{KL} = \sum_{i=1}^N y_i \cdot \log\left(\frac{y_i}{f(x_i)}\right)$





# OPTIMIZER

**STOCHASTIC GRADIENT DESCENT (SGD)**

**ADAM**

**RMSPROP**

**ADAGRAD**

**ADADELTA**

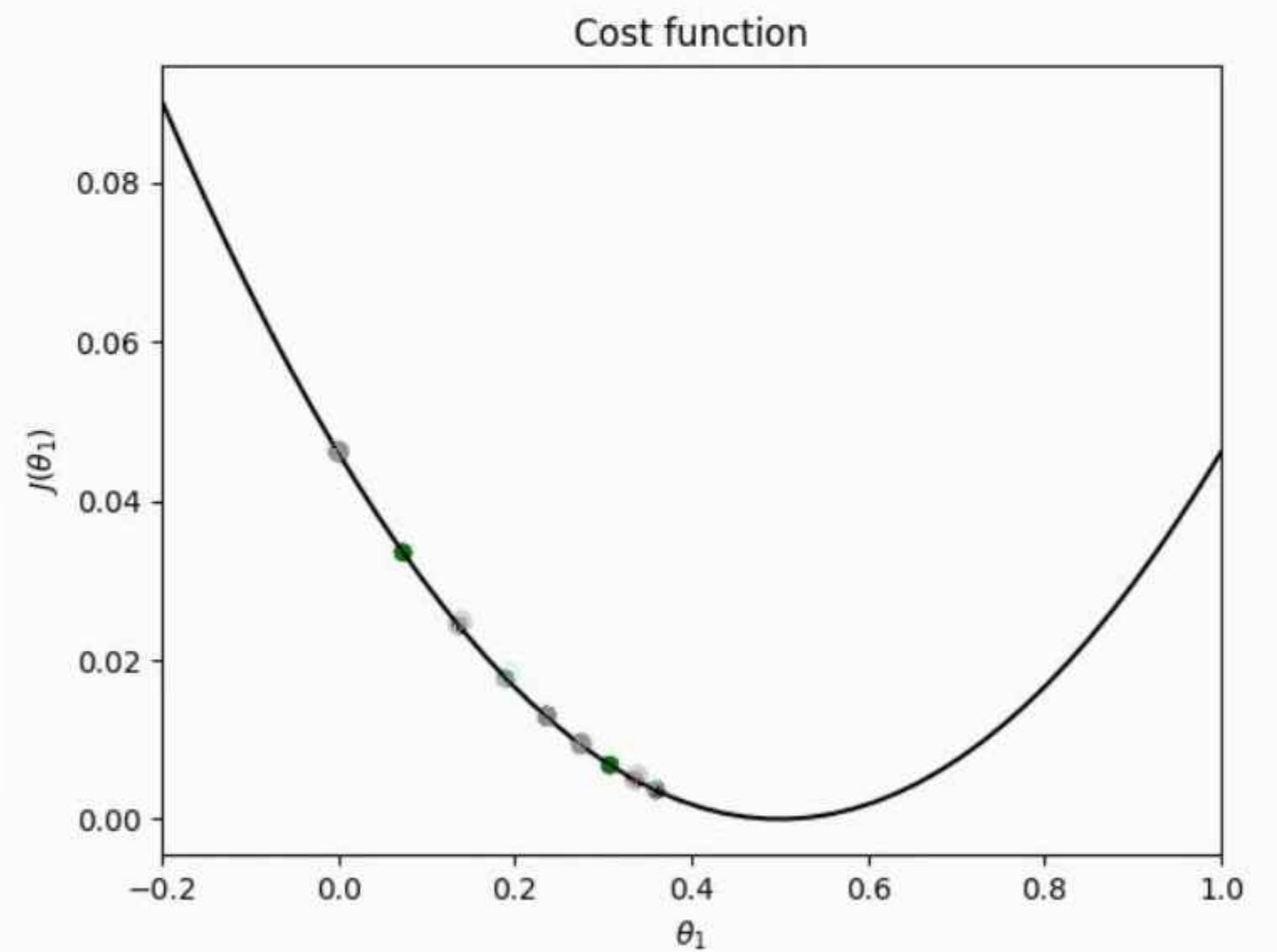
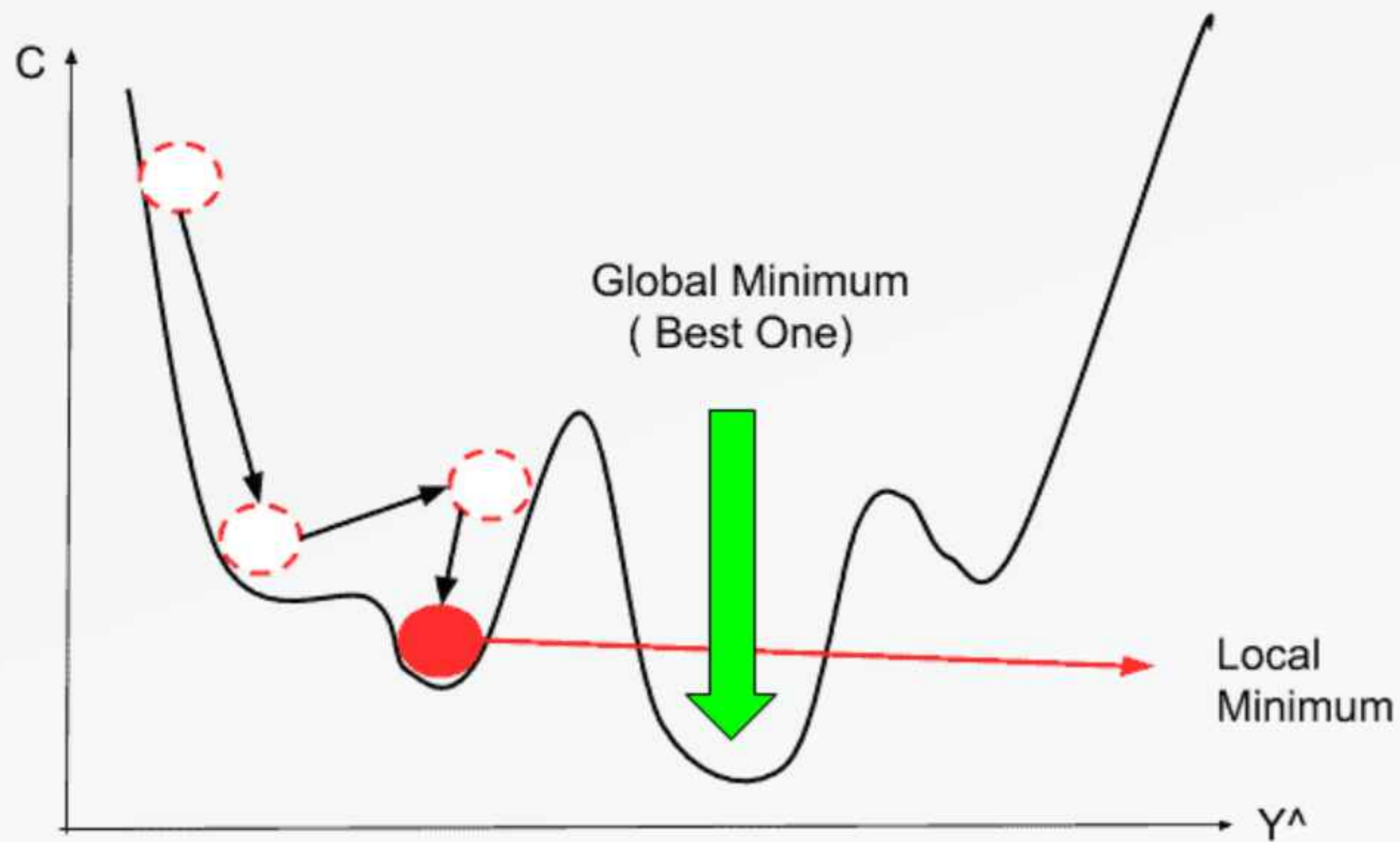
**NADAM**

**ADAMW**

▪  
▪  
▪



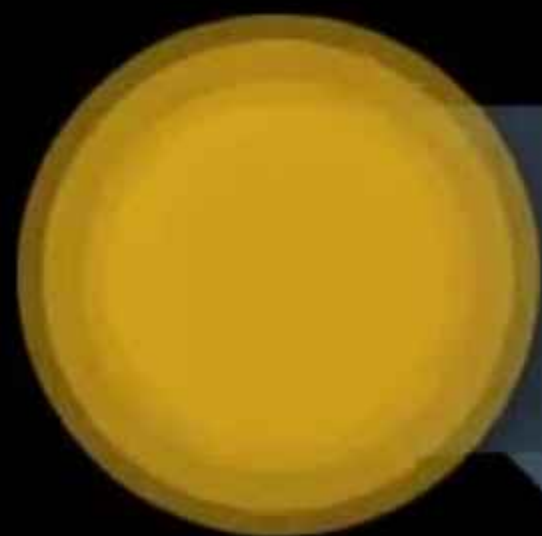
# STOCHASTIC GRADIENT DESCENT (SGD)



داستان شگفت انگیز

A.N.N

خنک دوست داشتنی



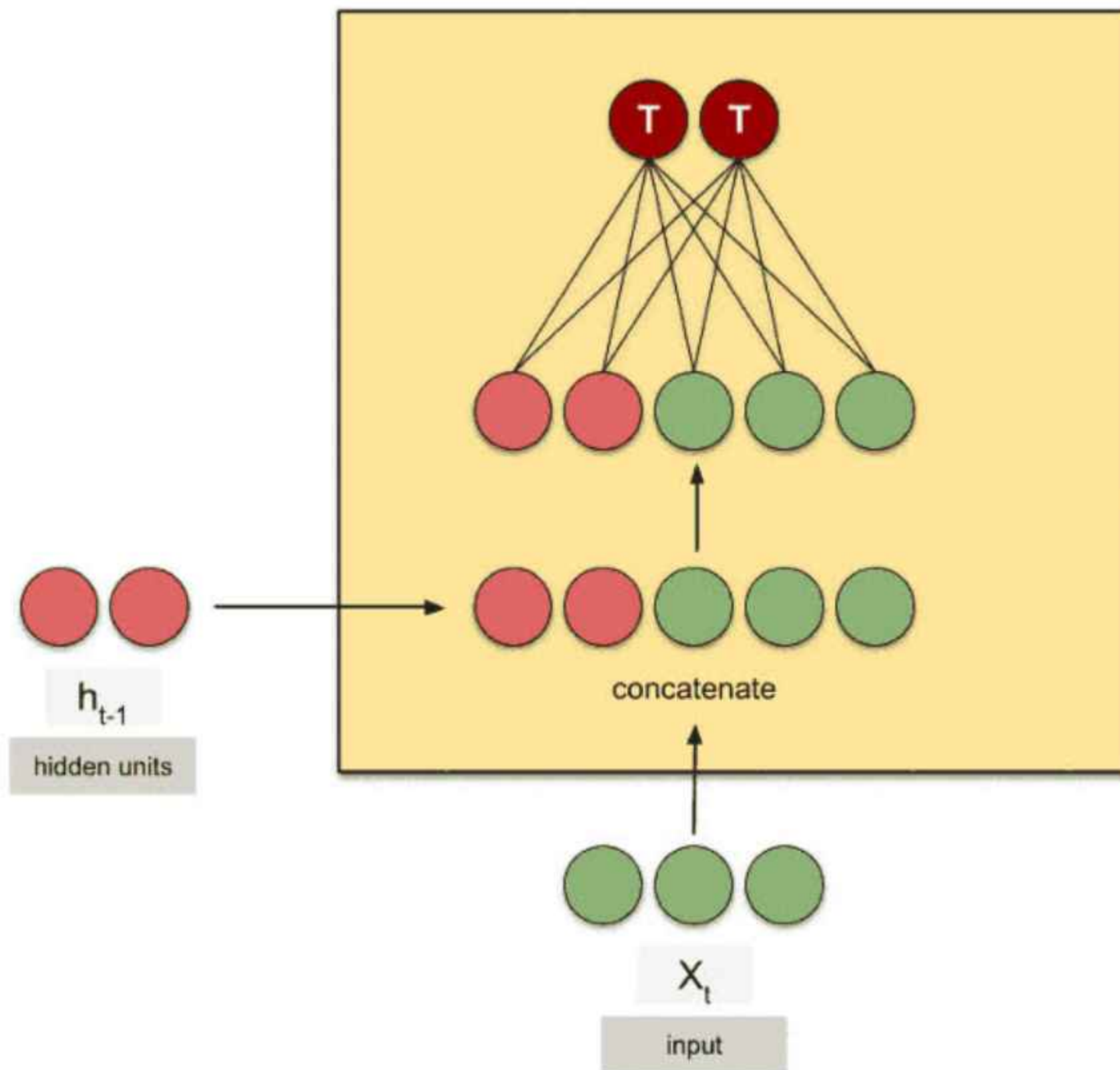
مجموعه

www.ann

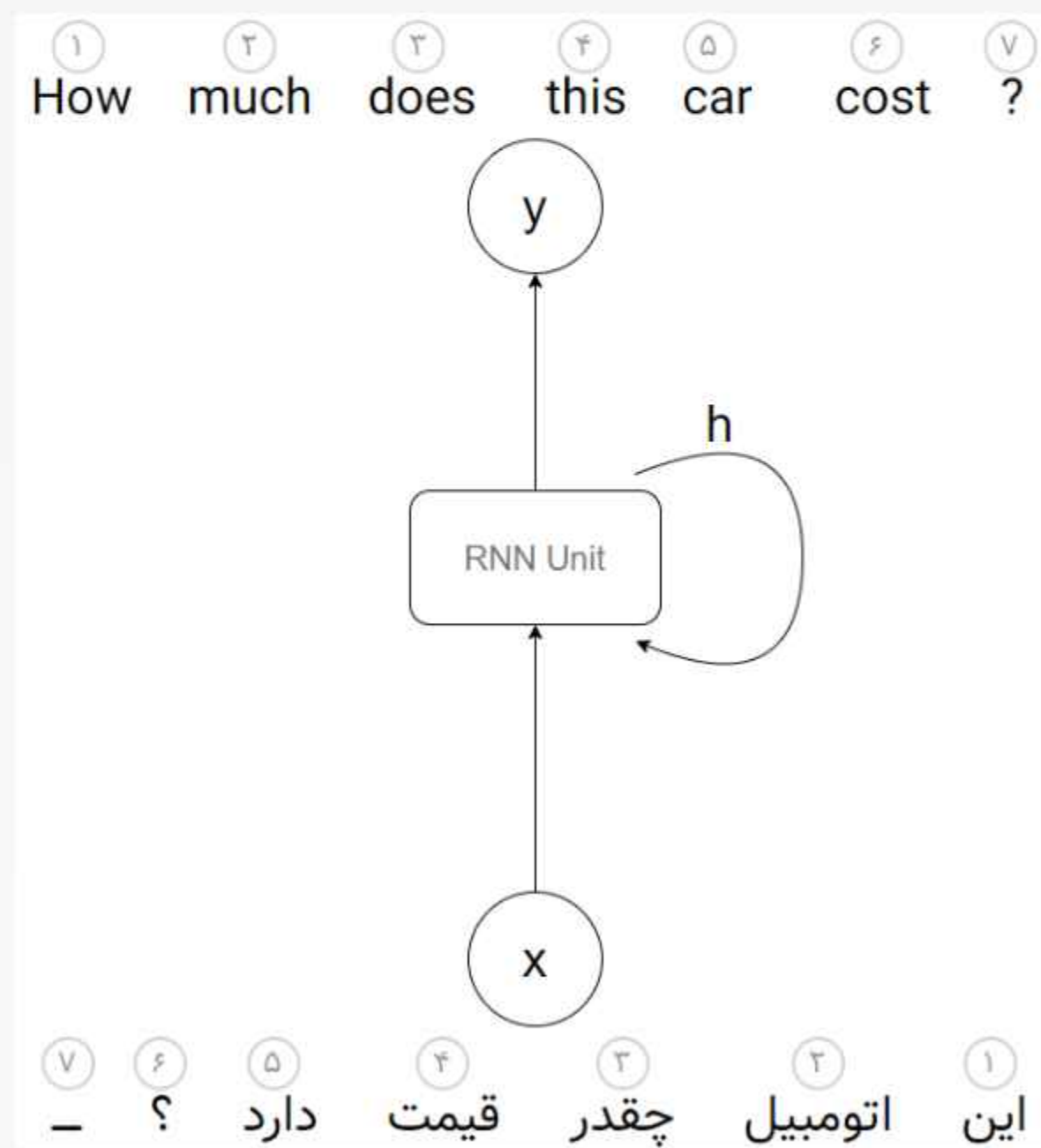
anarst.com/onlinehmc



# RECURRENT NEURAL NETWORK



# RNN

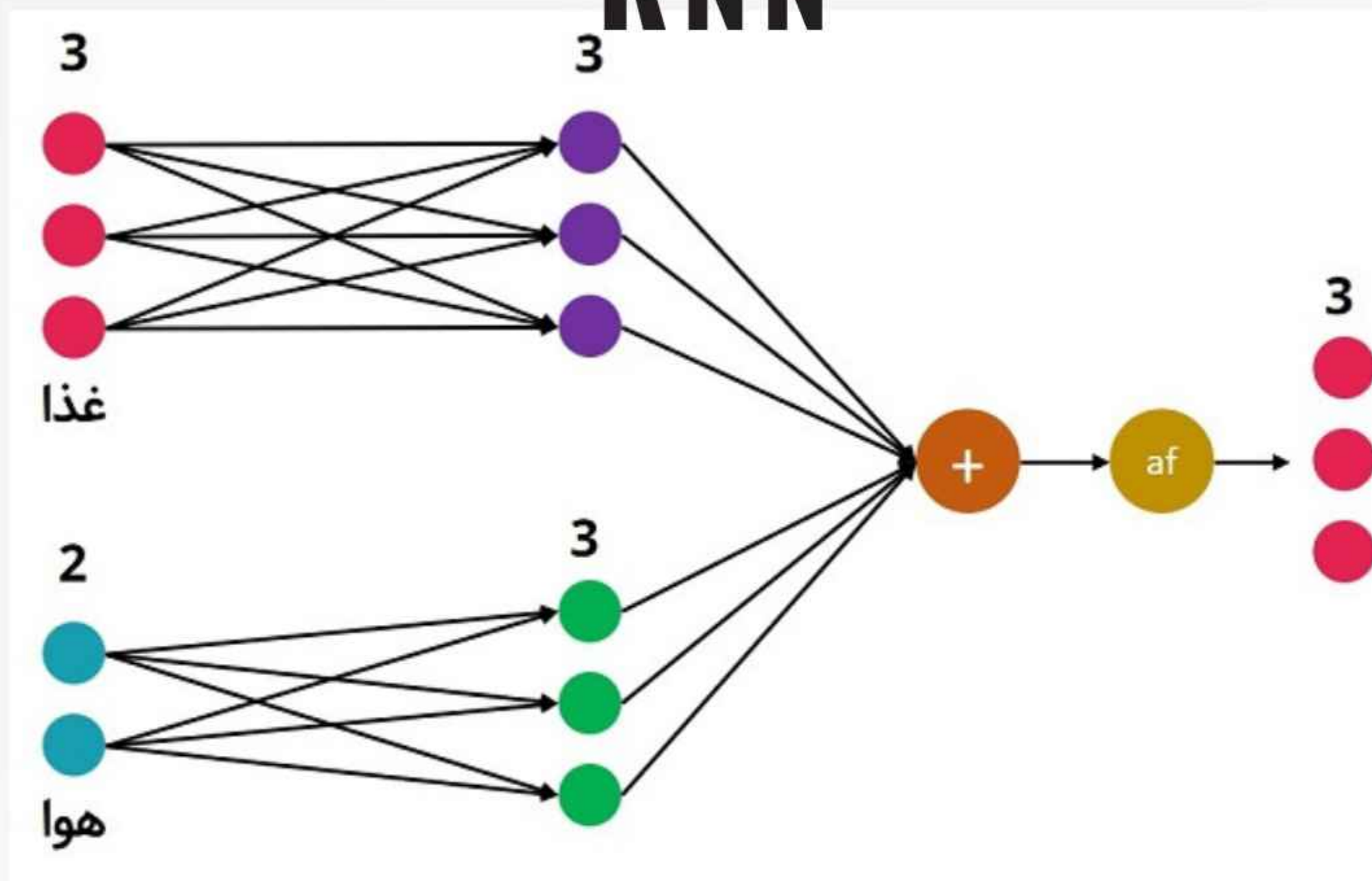




# RNN

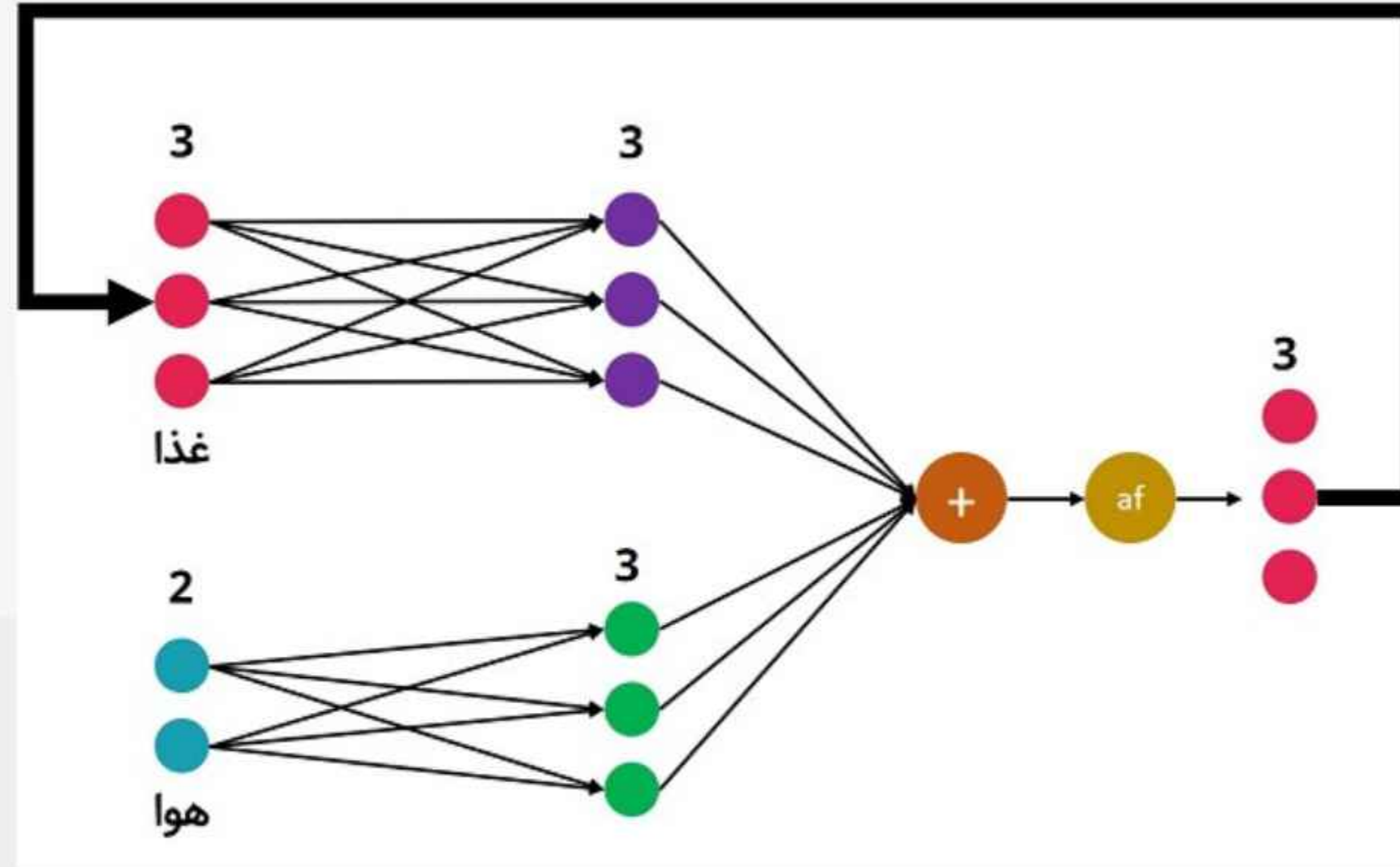


# RNN

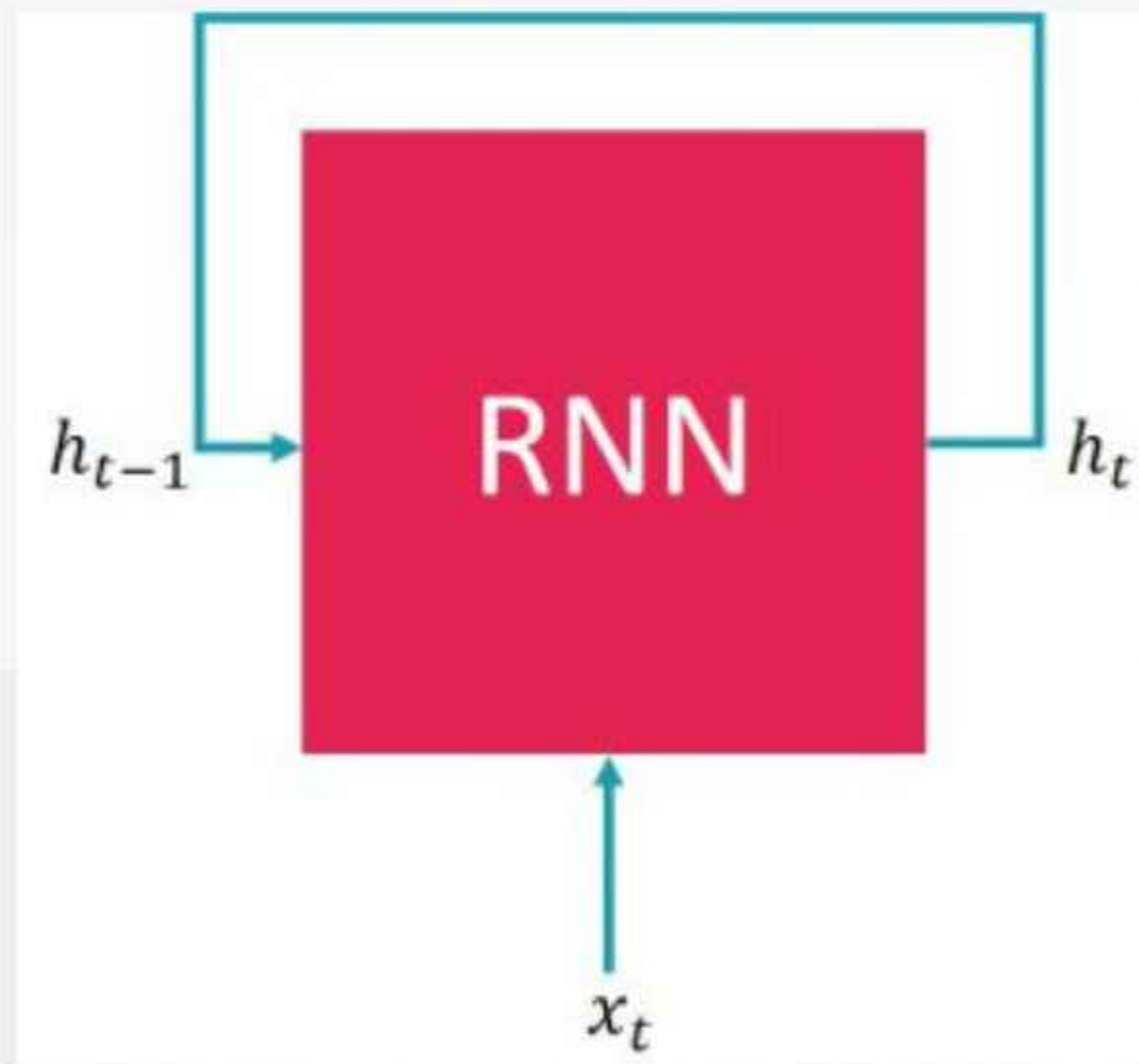




# RNN

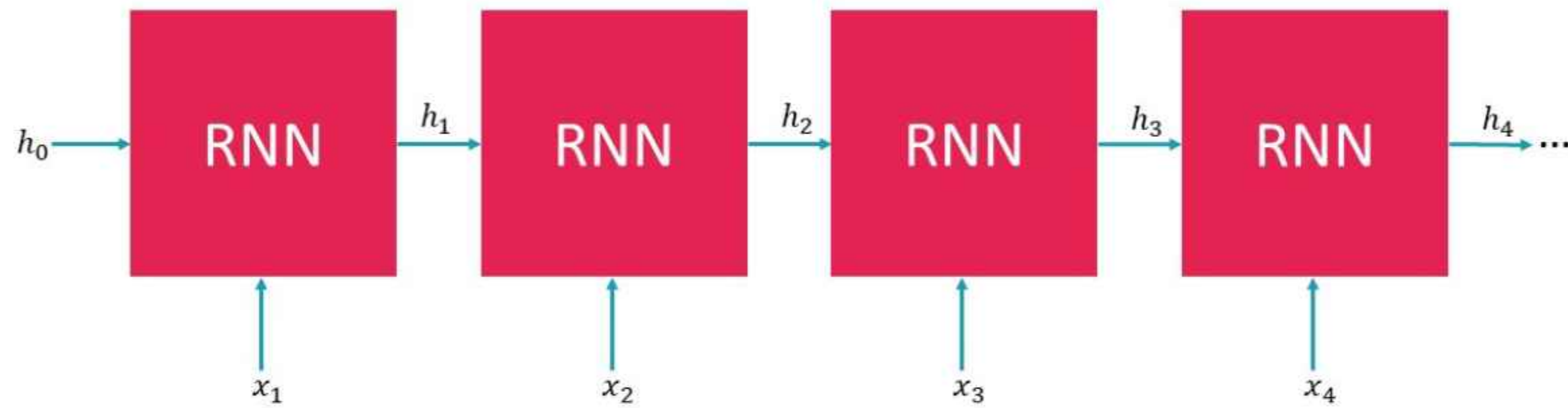


# RNN

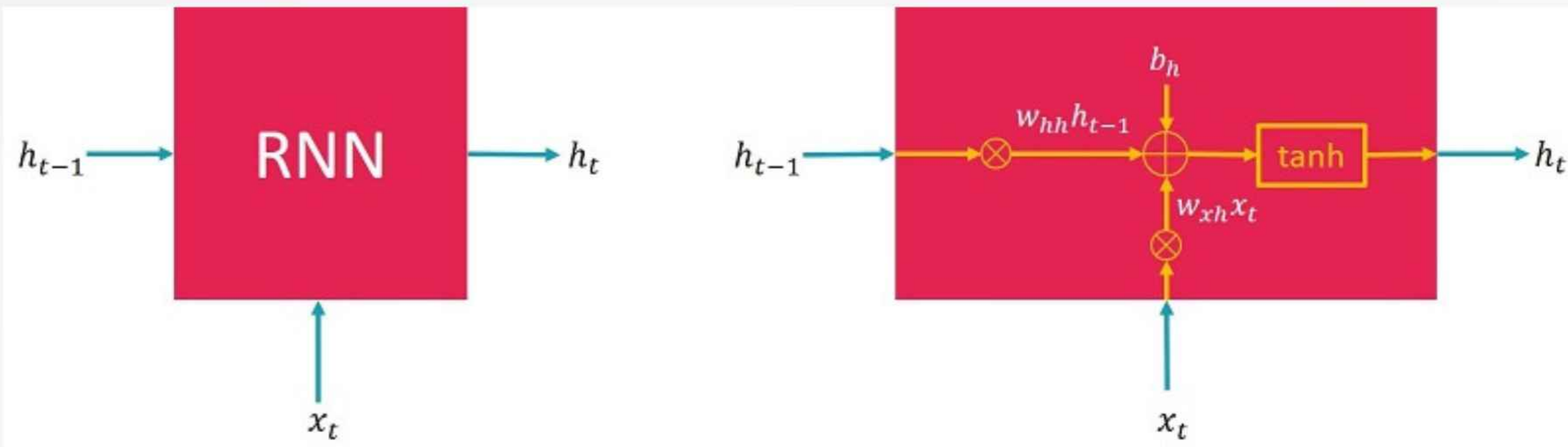




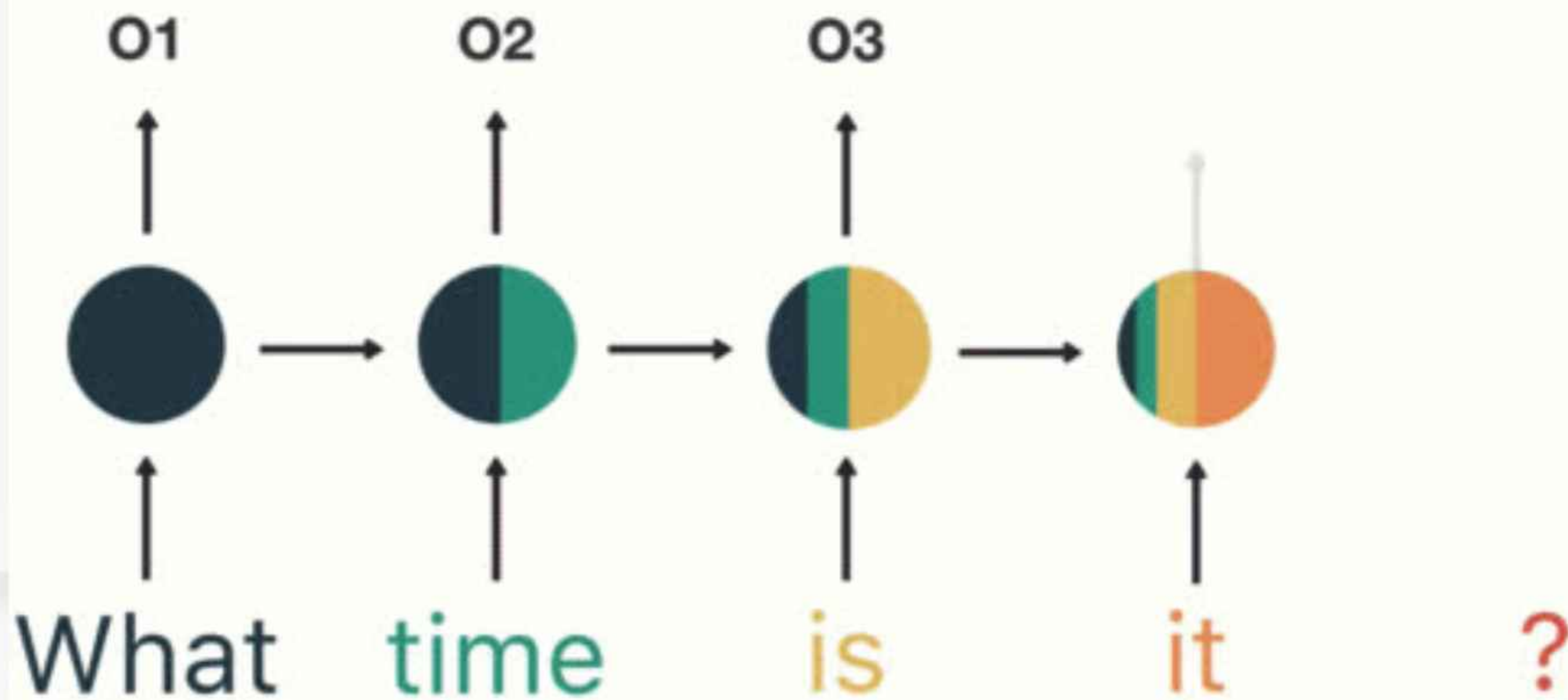
# RNN



# RNN



# RNN





# RNN

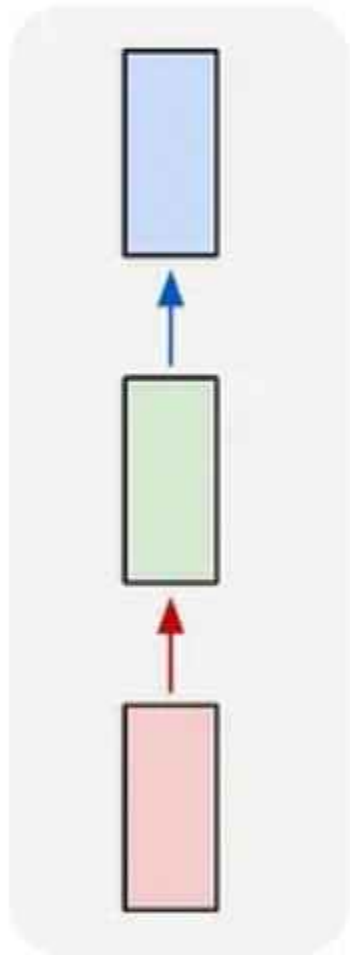
تبدیل متن به صوت

طبقه‌بندی احساس (مثبت / منفی)

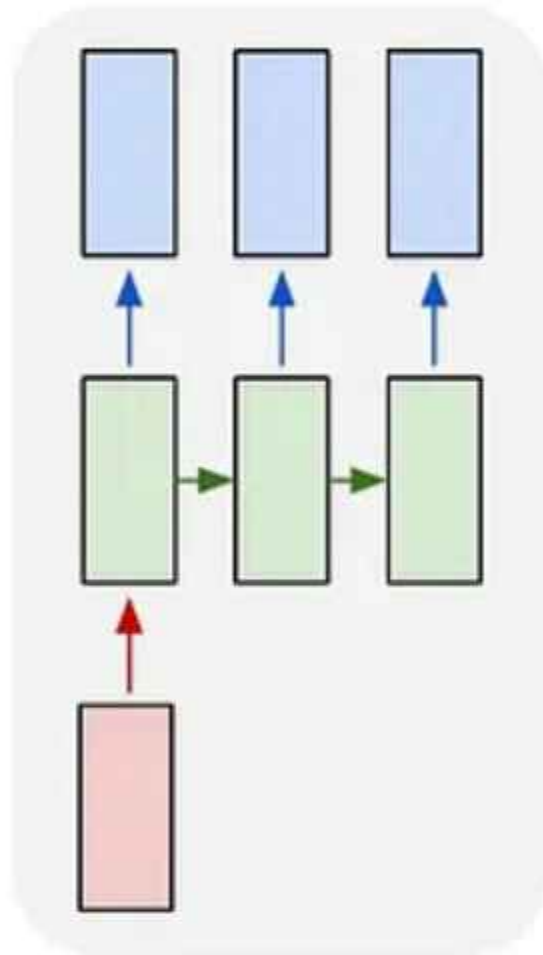
تصویر

عدد

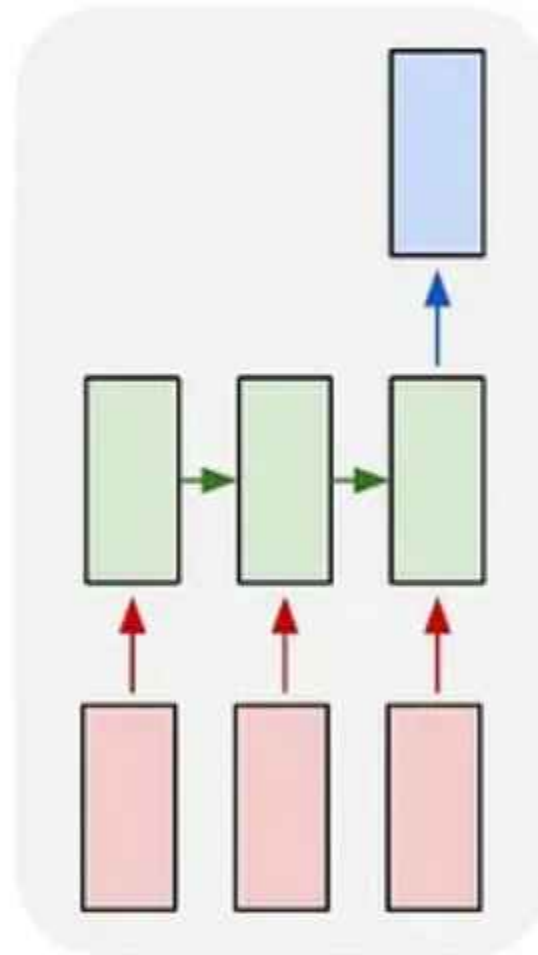
one to one



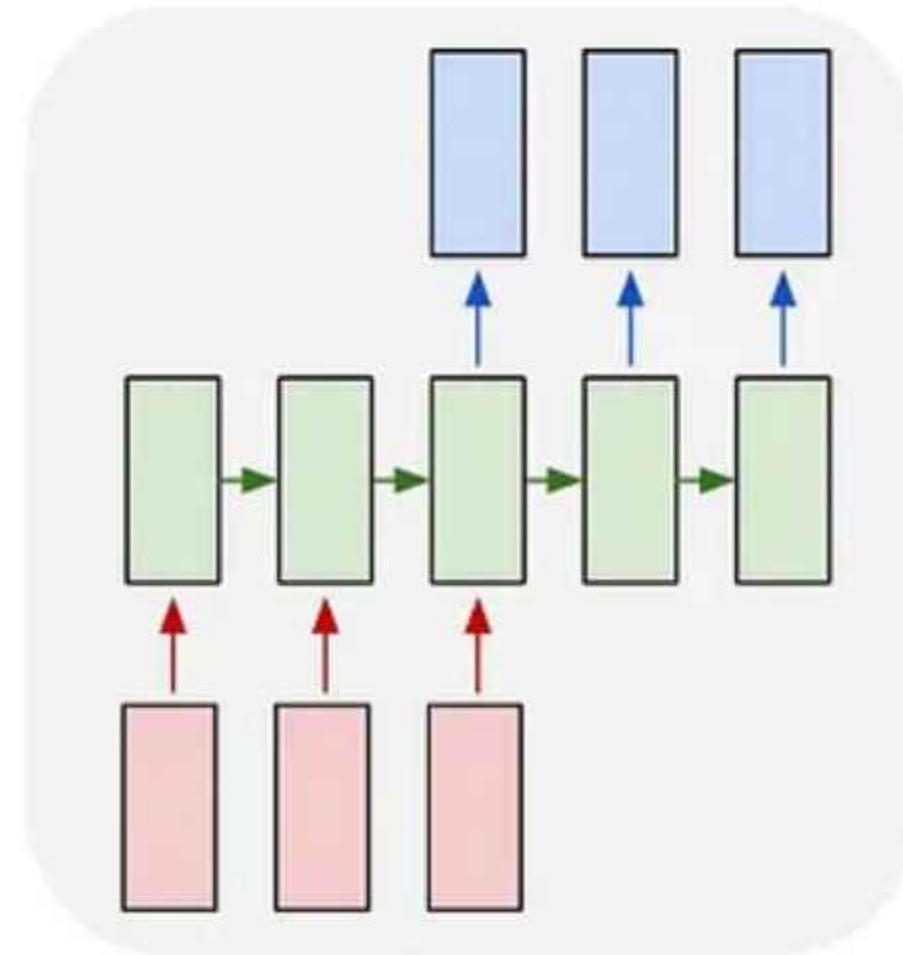
one to many



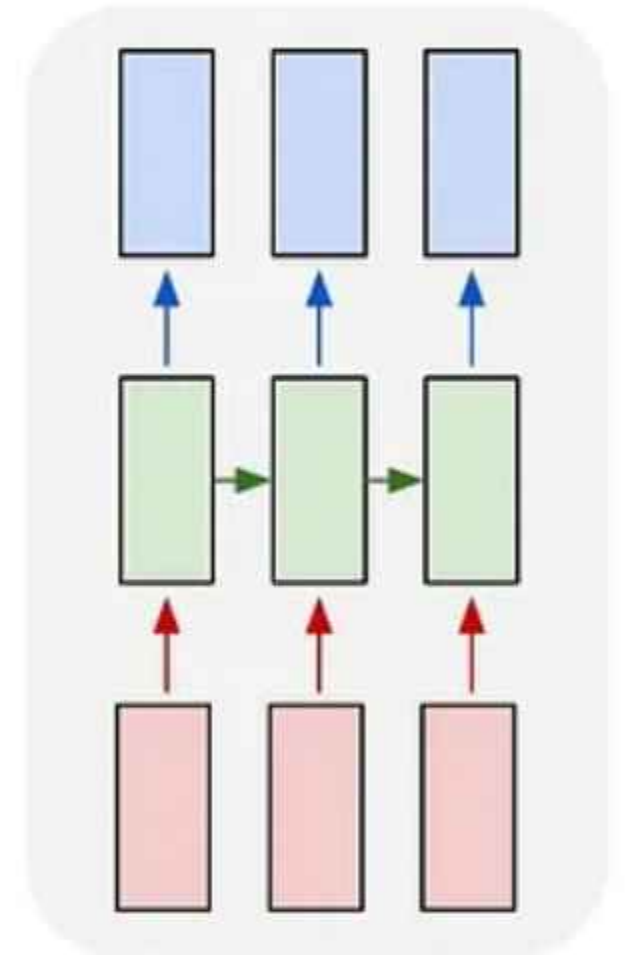
many to one



many to many



many to many



توصیف تصویر

ترجمه

# RNN

نوع RNN	ورودی‌ها	خروجی‌ها	مثال کاربردی
One-to-One	یک داده ایستا	یک برچسب	تشخیص تصویر
One-to-Many	یک داده ایستا	دنباله‌ای از داده‌ها	تولید متن از تصویر
Many-to-One	دنباله‌ای از داده‌ها	یک مقدار خروجی	تحلیل احساسات
Many-to-Many	دنباله‌ای از داده‌ها	دنباله‌ای از داده‌ها	ترجمه زبان، شناسایی اشیاء در ویدئو

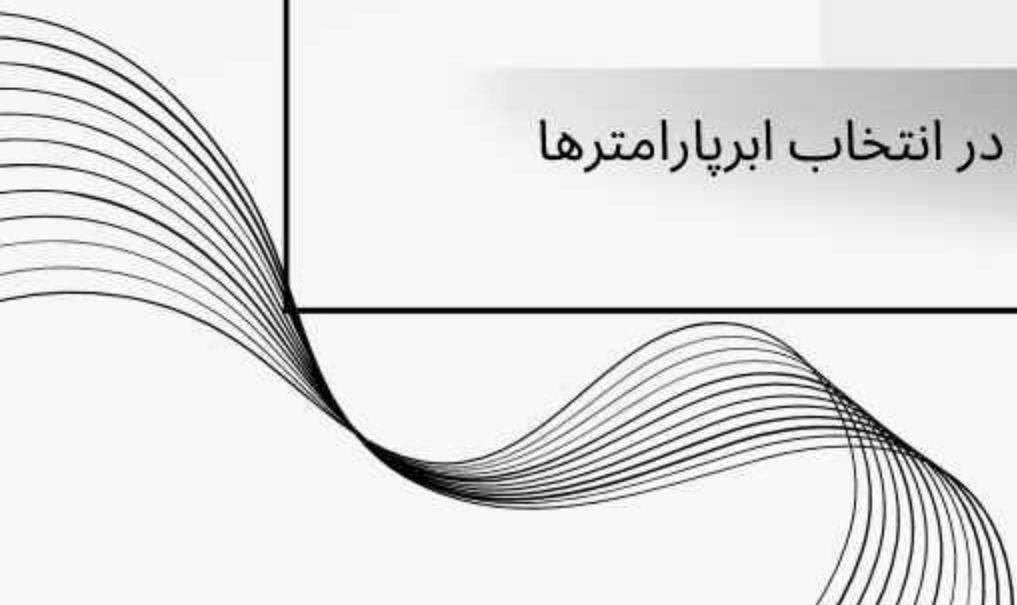
# RNN

علت اصلی	توضیح	مشکل
کاهش گرادیان و ناتوانی در حفظ اطلاعات از توالی‌های طولانی.	نمی‌تواند وابستگی‌های طولانی‌مدت را RNN به درستی یاد بگیرد و اطلاعات مراحل اولیه را حفظ کند.	مشکل یادگیری وابستگی‌های طولانی‌مدت
یا $\tanh$ استفاده از توابع فعال‌سازی مانند $\text{sigmoids}$ که گرادیان‌های کوچک تولید می‌کنند.	گرادیان‌ها در طول فرآیند آموزش به قدری کوچک می‌شوند که به روزرسانی‌های وزن‌ها انجام نمی‌شود.	(Vanishing Gradient) کاهش گرادیان
عدم تنظیم مناسب نرخ یادگیری یا استفاده از توابع فعال‌سازی که باعث افزایش شدید گرادیان‌ها می‌شود.	گرادیان‌ها در طول آموزش به قدری بزرگ می‌شوند که به روزرسانی‌ها ناپایدار و غیرقابل کنترل می‌شود.	(Exploding Gradient) افزایش گرادیان
وابستگی‌های توالی‌ای که نیاز به پردازش خط به خط دارند	پردازش هر گام زمانی به صورت سریالی است، که باعث کندی در آموزش می‌شود.	مشکل زمان‌بر بودن آموزش



# RNN

محاسبات به ترتیب زمانی انجام می‌شوند و قابل موازی‌سازی نیستند.	به دلیل وابستگی گام‌های زمانی به هم، پردازش موازی داده‌ها امکان‌پذیر نیست.	عدم توانایی پردازش موازی
ناپایداری در به‌روزرسانی وزن‌ها و انتخاب نادرست ابرپارامترها	ممکن است آموزش مدل به راحتی همگرا نشود و نیاز به تنظیمات خاصی داشته باشد.	مشکل همگرایی
ظرفیت محدود حافظه و عدم توانایی در ذخیره اطلاعات طولانی‌مدت	تنها می‌تواند اطلاعات مربوط به گام‌های RNN قبلی را حفظ کند و از آن‌ها استفاده کند.	حافظه محدود
وابستگی به داده‌های قبلی و ناتوانی در تشخیص نویز	حساسیت مدل به نویز و اختلالات در داده‌های ورودی.	آسیب‌پذیری به نویز
پیچیدگی در تنظیم دقیق پارامترهای مدل و عدم وجود روش‌های ساده برای انتخاب بهترین ترکیب	تنظیم ابرپارامترها (مانند نرخ یادگیری، تعداد لایه‌ها و طول توالی‌ها) دشوار است.	سختی در انتخاب ابرپارامترها



# RNN OR MLP

ویژگی	RNN (شبکه عصبی بازگشتی)	MLP (پرپرسترون چند لایه)
ساختار	شامل لایه‌های بازگشتی که داده‌ها را در طول دنباله‌ها پردازش می‌کنند.	که داده‌ها را از ورودی به (dense) شامل لایه‌های کاملاً متصل خروجی نقشه‌برداری می‌کنند.
نوع داده‌ها	مناسب برای داده‌های ترتیبی (مانند سری‌های زمانی و دنباله‌های متنی)	مناسب برای داده‌های ایستا و مستقل از زمان (غیر ترتیبی)
نحوه پردازش داده‌ها	داده‌ها به ترتیب گام‌های زمانی پردازش می‌شوند. مدل به ترتیب ورودی‌ها را دریافت کرده و از حالت‌های قبلی برای پردازش داده‌های جدید استفاده می‌کند.	مدل به‌طور همزمان همه ویژگی‌ها را در هر نمونه پردازش می‌کند و ترتیب داده‌ها اهمیت ندارد.
ساختار شبکه	که دارای حالت داخلی هستند و می‌توانند وابستگی‌های طولانی‌مدت را یاد بگیرند RNN لایه‌های	که به ورودی‌ها به‌طور مستقل (Dense) لایه‌های کامل متصل پردازش می‌کنند.
مناسب برای	مسائل با داده‌های دنباله‌ای (مانند پیش‌بینی سری زمانی، پردازش زبان طبیعی، تحلیل ویدیو، ...)	مسائل با داده‌های ایستا (مانند طبقه‌بندی، پیش‌بینی‌های رگرسیونی، ...)
استفاده از حافظه	از حافظه داخلی برای نگه‌داشتن اطلاعات گام‌های زمانی قبلی استفاده می‌کند RNN	هیچ‌گونه حافظه داخلی ندارد و به داده‌های ورودی مستقل MLP از زمان پردازش می‌کند.
توانایی یادگیری وابستگی‌های طولانی‌مدت	این محدودیت GRU یا LSTM اما با استفاده از (RNN به‌خصوص در مدل‌های ساده) محدودیت‌هایی دارد کاهش می‌یابد.	توانایی یادگیری وابستگی‌های طولانی‌مدت ندارد، به دلیل ساختار غیر ترتیبی.
کاربرد	تحلیل ویدیو و صوت (NLP)، پیش‌بینی سری‌های زمانی، پردازش زبان طبیعی	طبقه‌بندی تصاویر، تشخیص تقلب، پیش‌بینی رگرسیونی
پیچیدگی محاسباتی	پیچیده‌تر به دلیل پردازش داده‌ها در طول دنباله‌ها و نیاز به حافظه برای حالت‌های قبلی	ساده‌تر به دلیل ساختار غیر ترتیبی و عدم نیاز به حافظه

**THE END**

