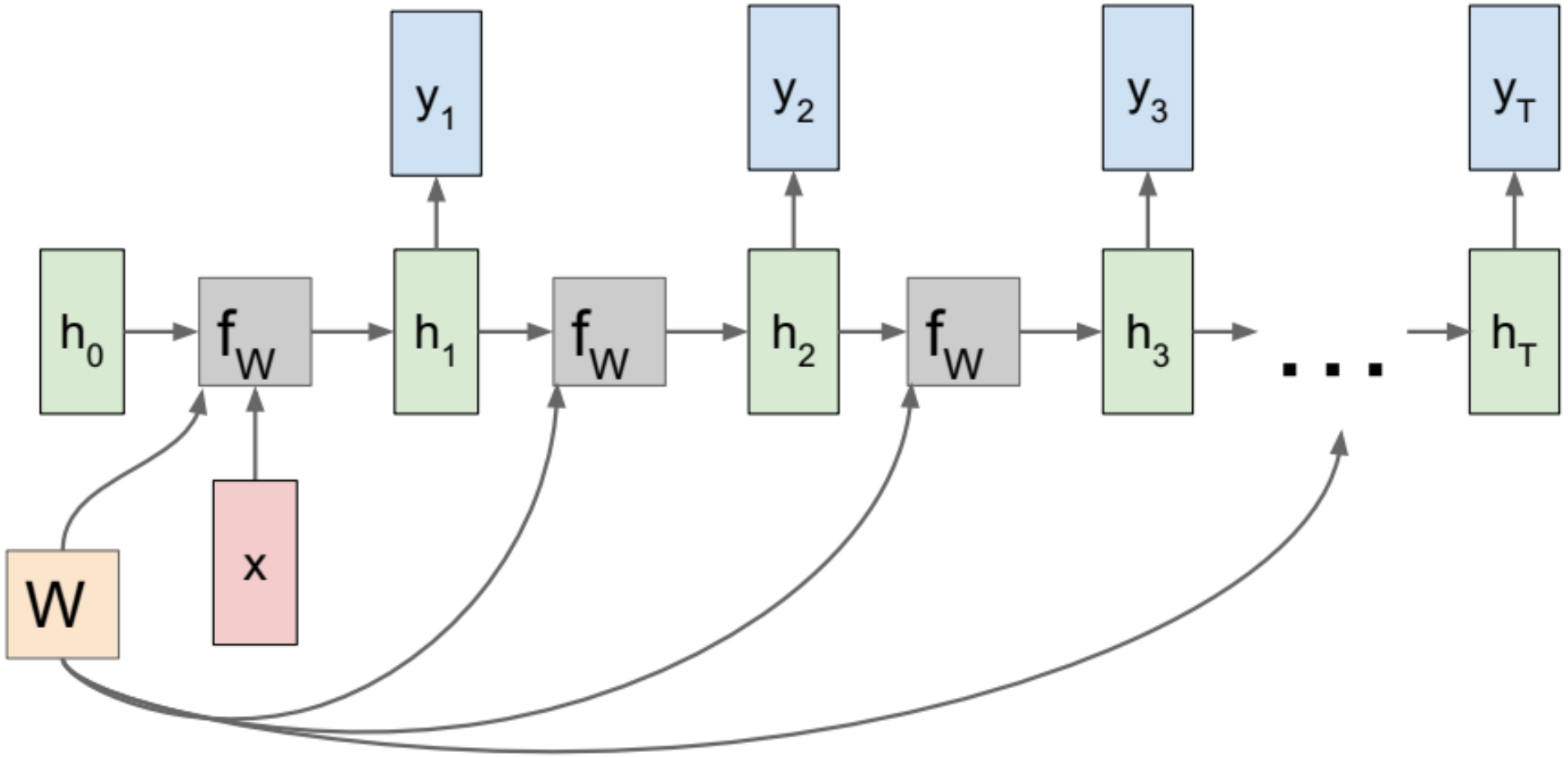

بسم الله الرحمن الرحيم

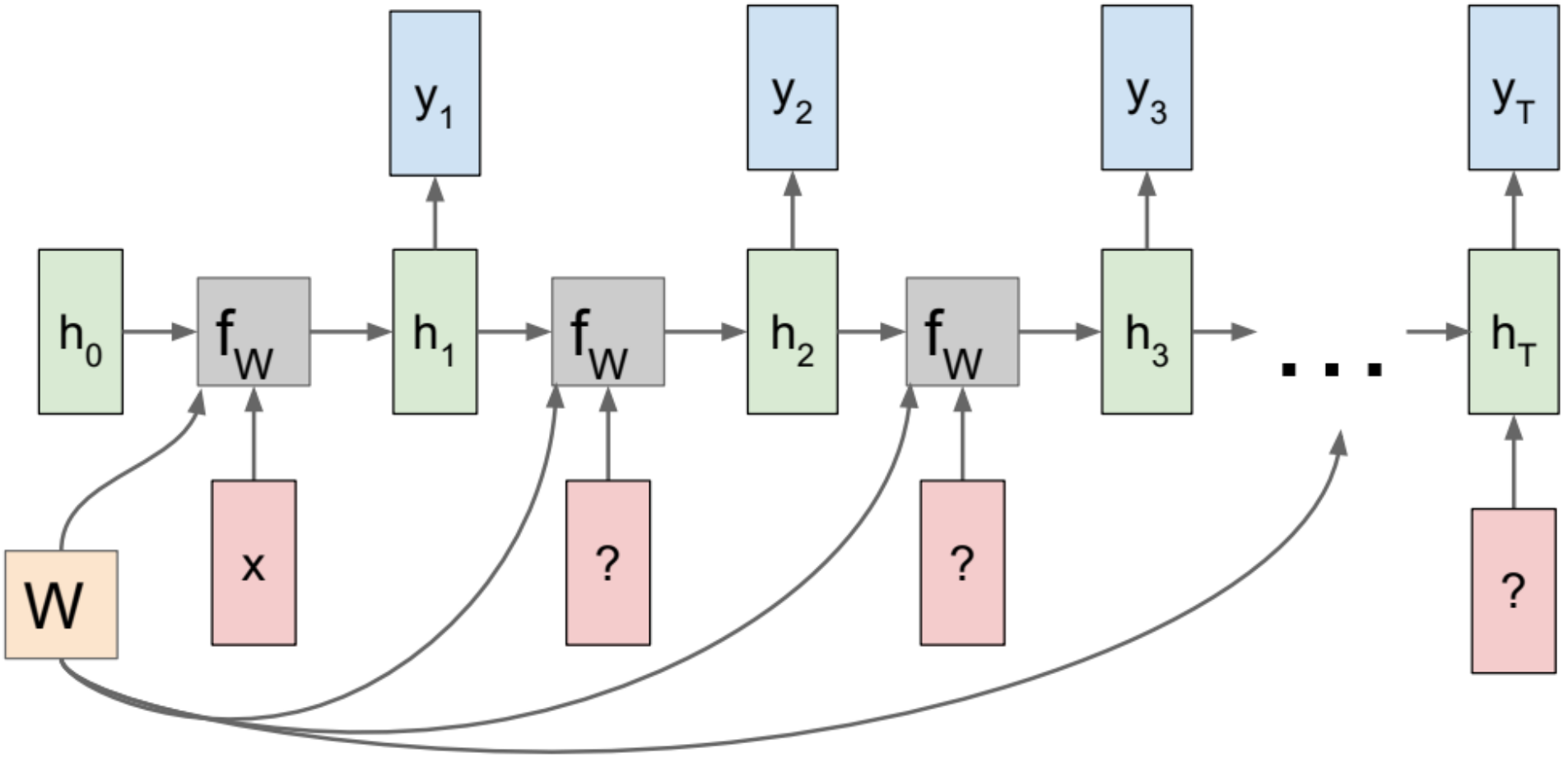
شبکه‌های عصبی بازگشتی

Recurrent Neural Networks

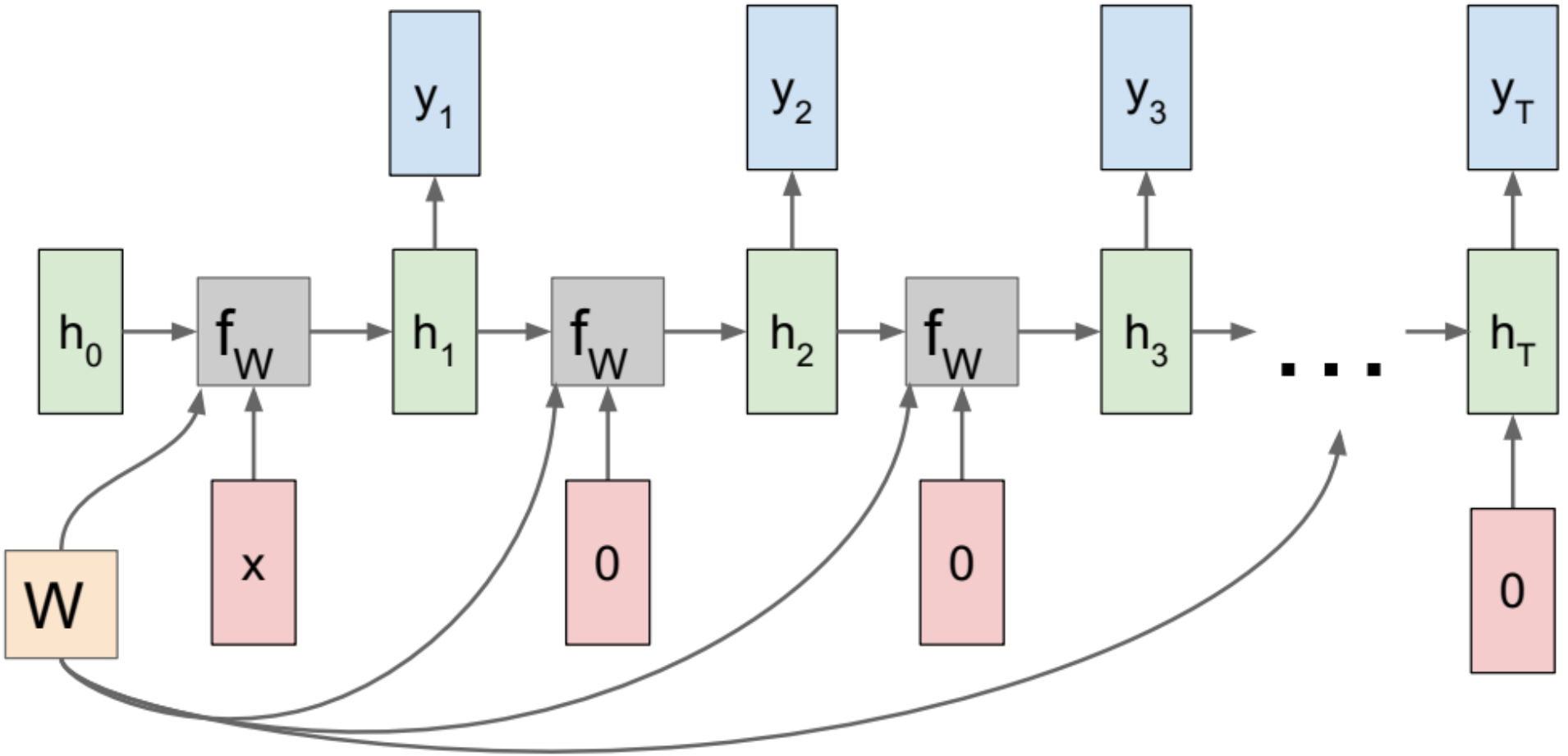
One to Many



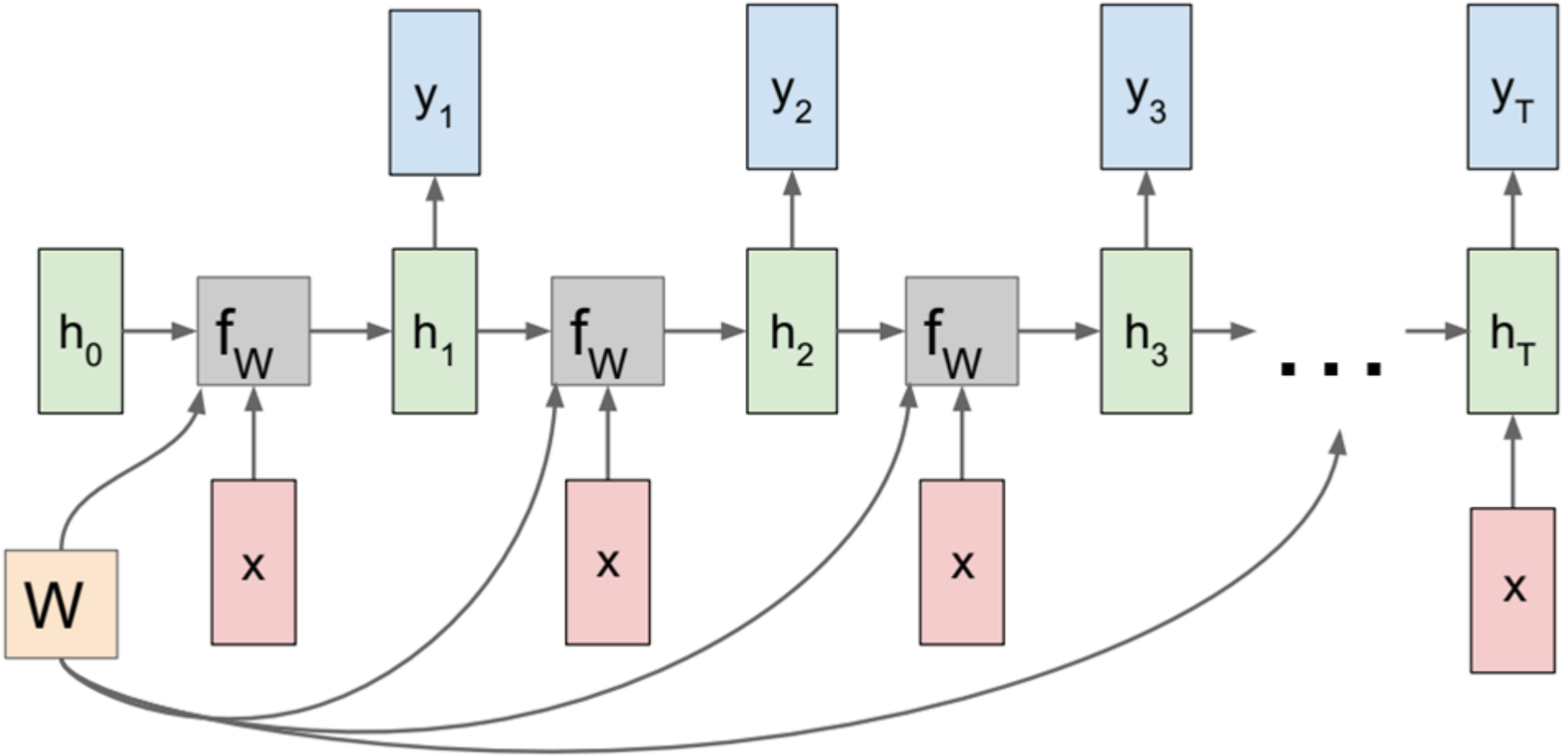
One to Many



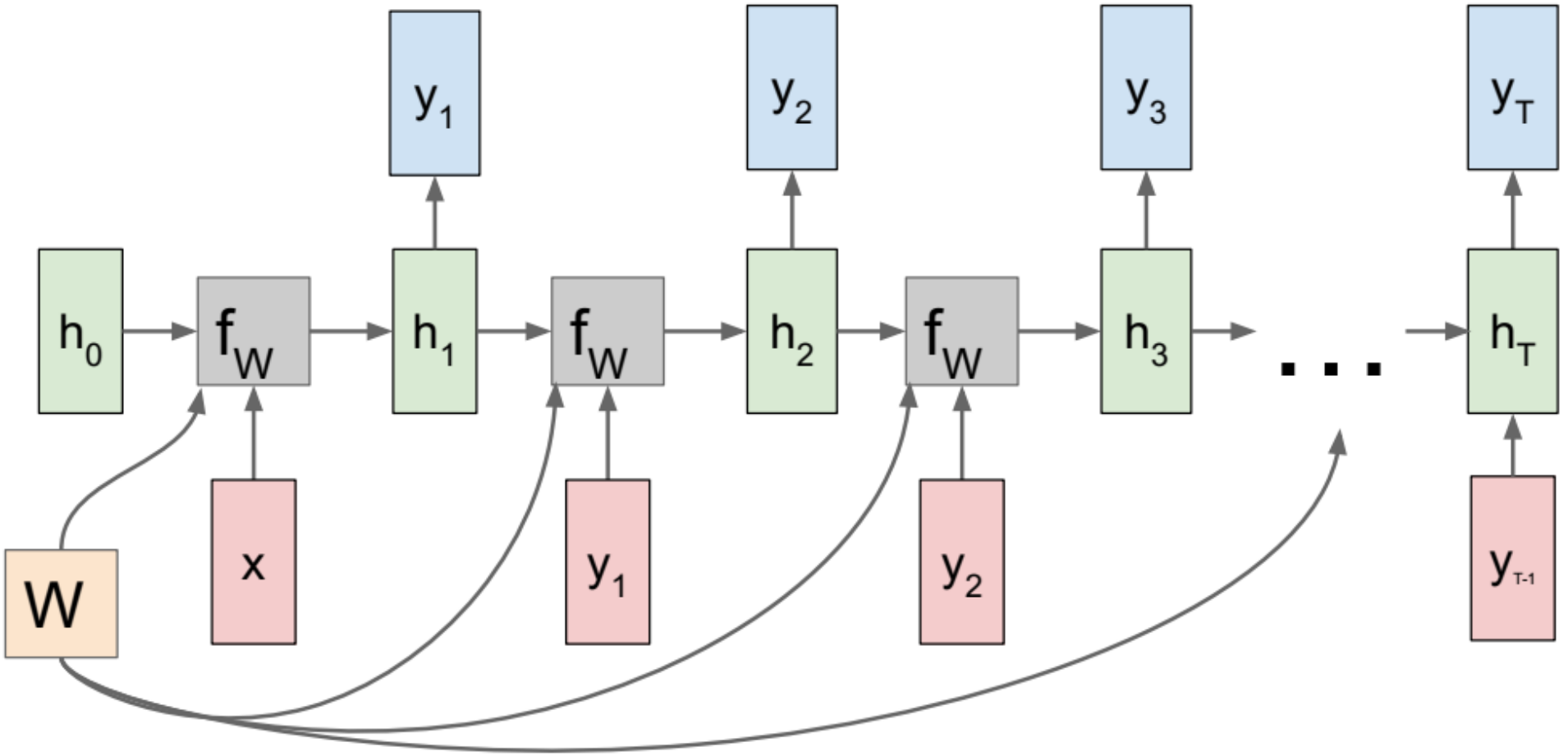
One to Many



One to Many

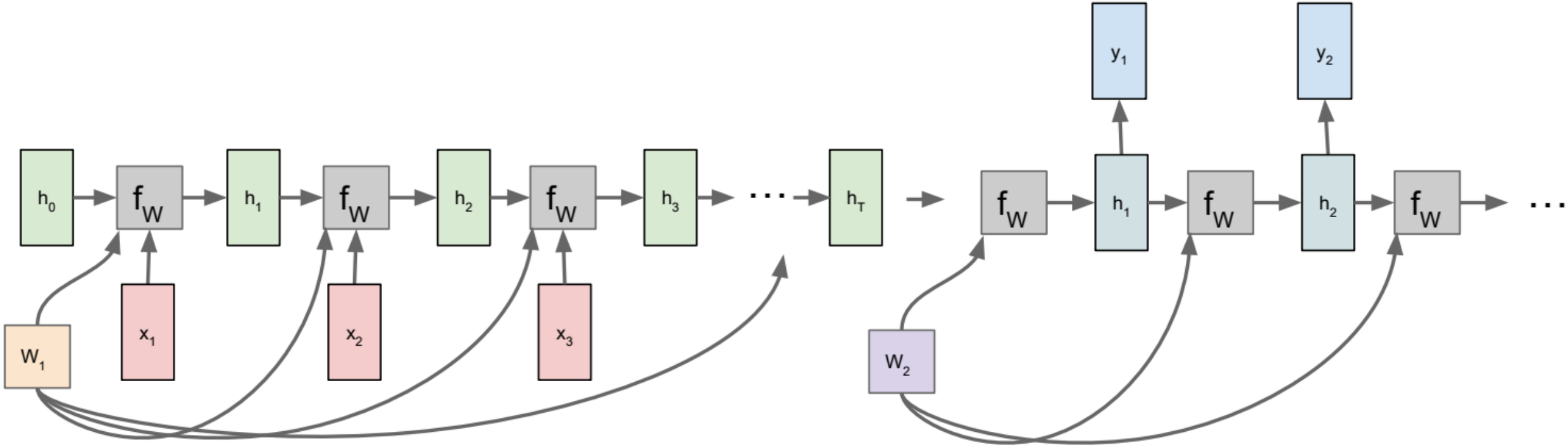


One to Many



Sequence to Sequence

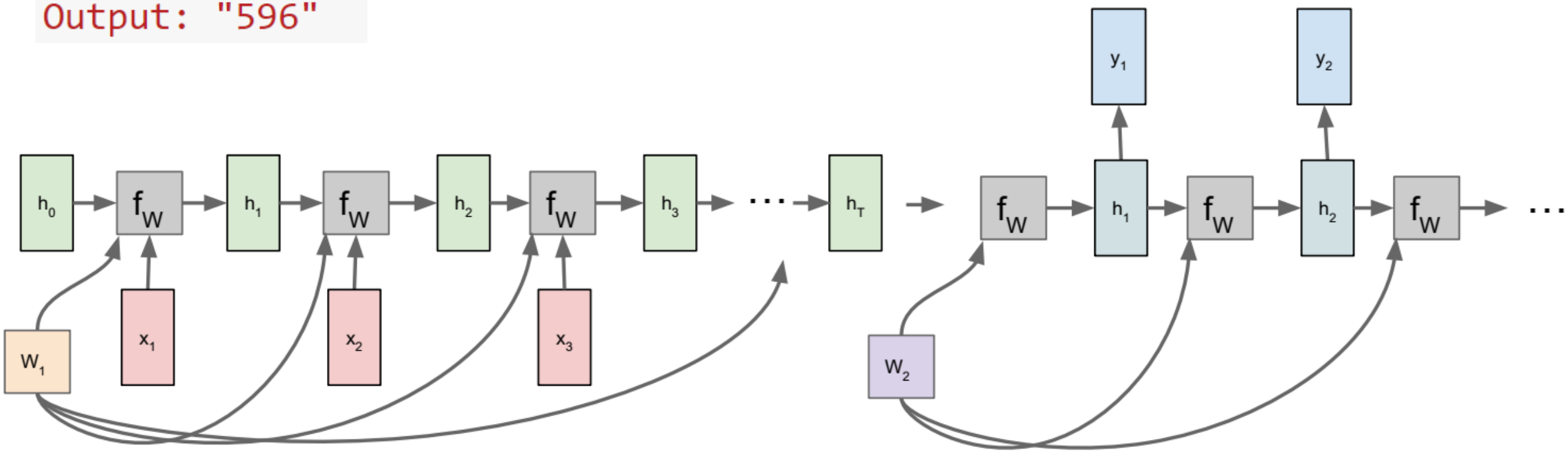
- Many to One + One to Many
  - Many to one: Encode input sequence in a single vector
  - One to many: Produce output sequence from single input vector



تبدیل دنباله به دنباله

- مثال: آموزش یک مدل برای یادگیری جمع کردن دو عدد، که ورودی و خروجی آن رشته‌هایی از کاراکترها هستند

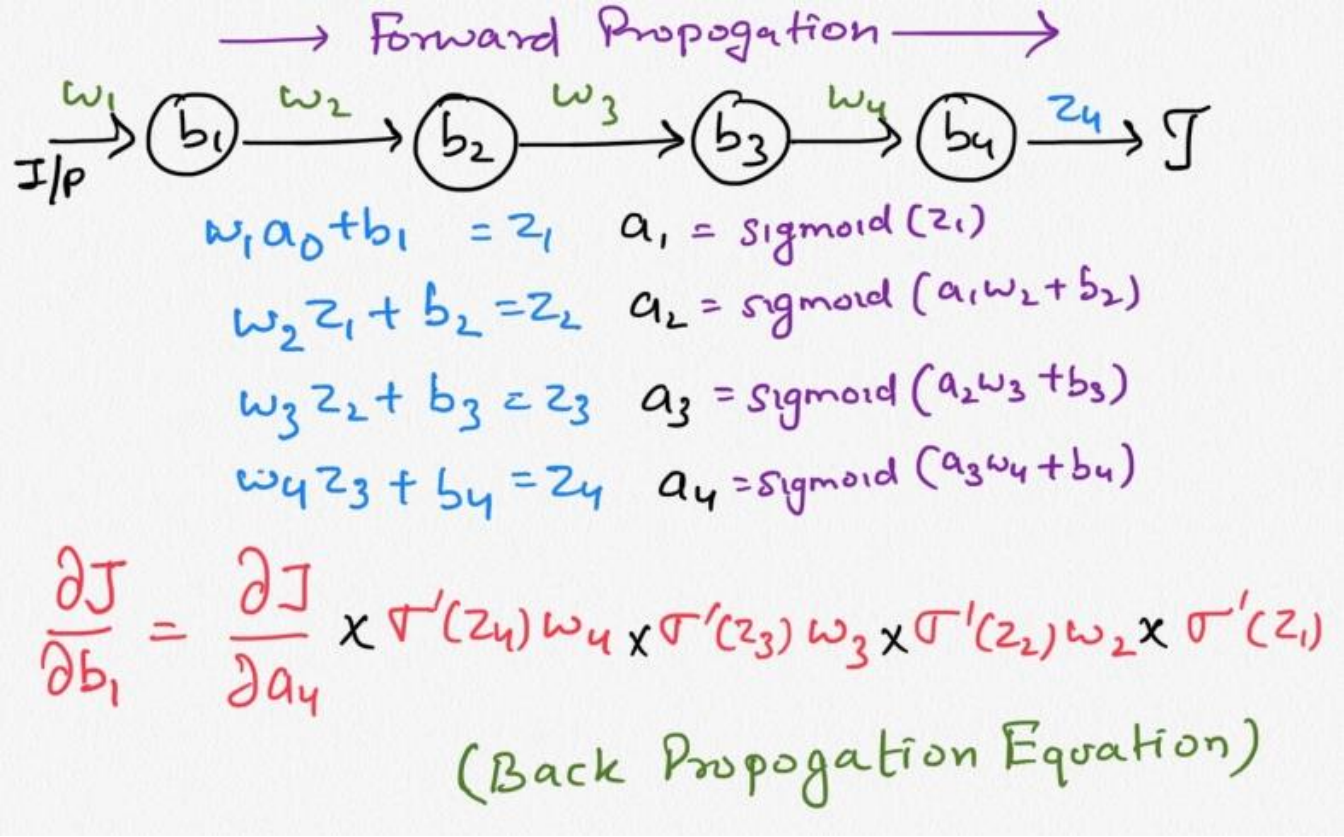
```
Input: "535+61"
Output: "596"
```

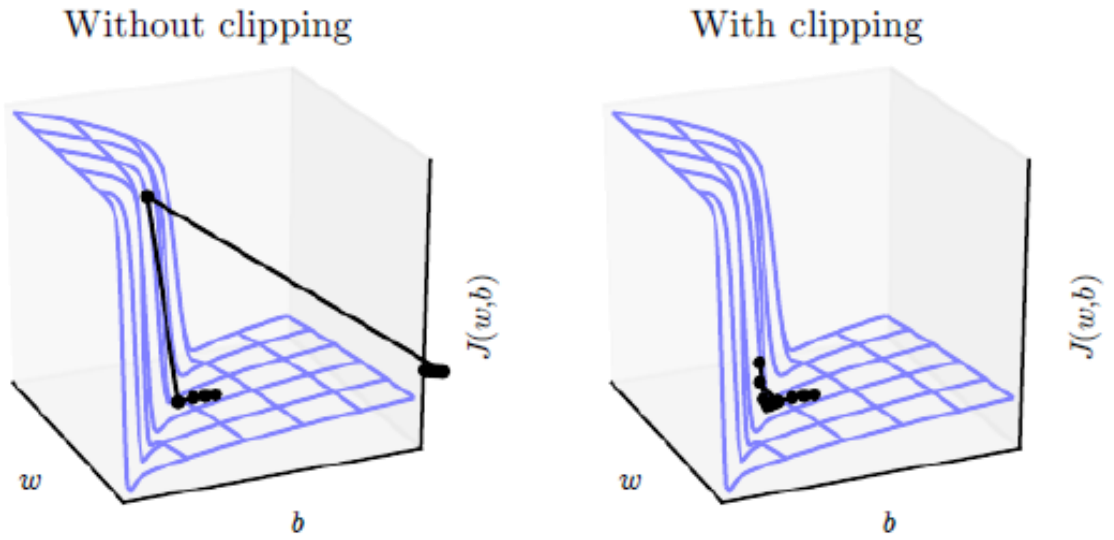


https://colab.research.google.com/drive/11qFxWP5ukCOKXXcP2jTBYEKRl3ThBYH6?usp=sharing

محوشدگی و انفجار گرادیان

- در شبکه‌های بازگشتی برای دنباله‌های طولانی این مشکل بسیار جدی است
- می‌توان با برش گرادیان از انفجار گرادیان جلوگیری کرد



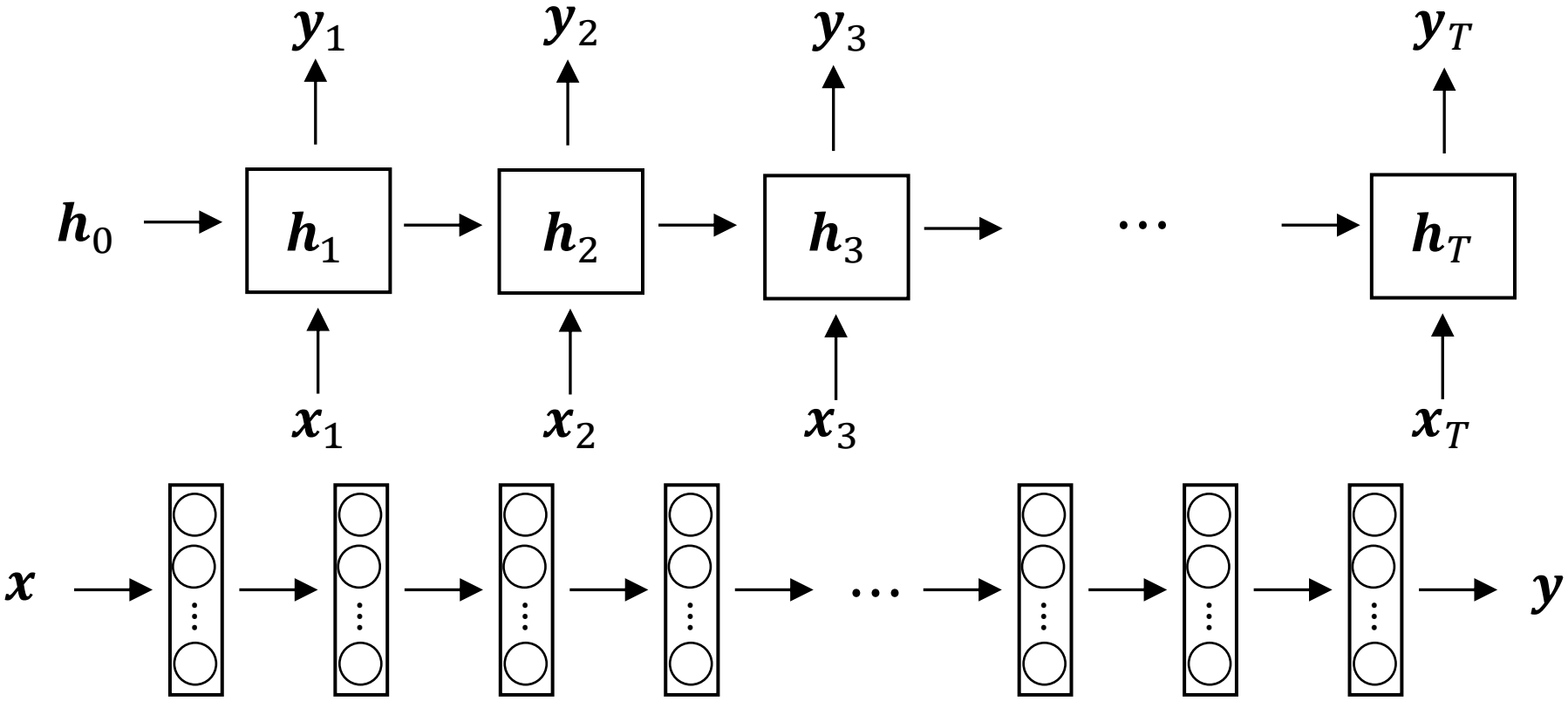


https://realelectricwizard.wordpress.com/2019/03/30/the-mystery-of-vanishing-and-exploding-gradients/

محوشدگی گرادیان در RNNها

The cat, which already ate …, was full.

The cats, which already ate …, were full.



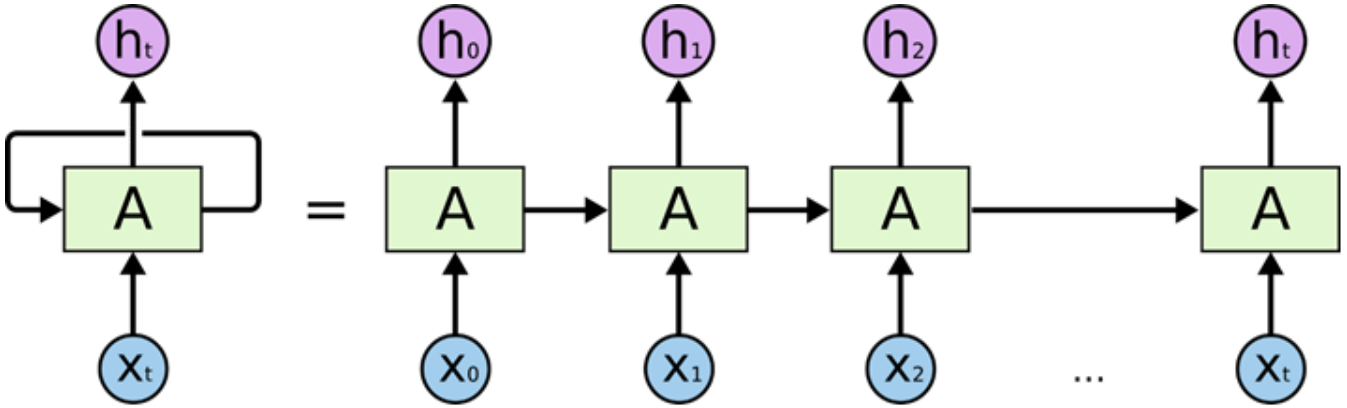
https://www.coursera.org/lecture/nlp-sequence-models/recurrent-neural-network-model-ftkzt

Gated RNNs

- شبکه‌های عصبی بازگشتی از حافظه کوتاه مدت رنج می‌برند
  - RNN ممکن است اطلاعات مهم ابتدایی را نادیده بگیرد
- Gated RNNs مبتنی بر ایده ایجاد مسیرهایی در طول زمان هستند که از محوشدگی یا انفجار گرادیان جلوگیری می‌کنند

The cat, which already ate …, was full.

The cats, which already ate …, were full.



Gated Recurrent Units

GRU (simplified)

$$\tilde{\boldsymbol{h}}^{(t)} = \tanh\left(\boldsymbol{W}_{hh}\,\boldsymbol{h}^{(t-1)} + \boldsymbol{W}_{xh}\,\boldsymbol{x}^{(t)} + \boldsymbol{b}_h\right)$$

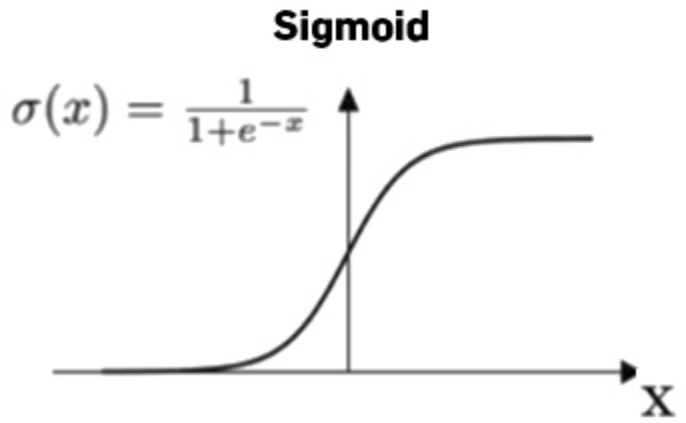
$$\boldsymbol{h}^{(t)} = \boldsymbol{u}^{(t)} \,.\, \tilde{\boldsymbol{h}}^{(t)} + \left(1 - \boldsymbol{u}^{(t)}\right).\, \boldsymbol{h}^{(t-1)}$$

$$\boldsymbol{u}^{(t)} = \sigma\left(\boldsymbol{W}_{hu}\,\boldsymbol{h}^{(t-1)} + \boldsymbol{W}_{xu}\,\boldsymbol{x}^{(t)} + \boldsymbol{b}_u\right)$$

The cat, which already ate …, was full.

Simple RNN

$$\boldsymbol{h}^{(t)} = \tanh\left(\boldsymbol{W}_{hh}\,\boldsymbol{h}^{(t-1)} + \boldsymbol{W}_{xh}\,\boldsymbol{x}^{(t)} + \boldsymbol{b}_h\right)$$



Gated Recurrent Units

GRU (simplified)

$$\tilde{\boldsymbol{h}}^{(t)} = \tanh\left(\boldsymbol{W}_{hh}\, \boldsymbol{h}^{(t-1)} + \boldsymbol{W}_{xh}\, \boldsymbol{x}^{(t)} + \boldsymbol{b}_h\right)$$

$$\boldsymbol{h}^{(t)} = \boldsymbol{u}^{(t)} . \tilde{\boldsymbol{h}}^{(t)} + \left(1 - \boldsymbol{u}^{(t)}\right). \boldsymbol{h}^{(t-1)}$$

$$\boldsymbol{u}^{(t)} = \sigma\left(\boldsymbol{W}_{hu}\, \boldsymbol{h}^{(t-1)} + \boldsymbol{W}_{xu}\, \boldsymbol{x}^{(t)} + \boldsymbol{b}_u\right)$$

GRU

$$\tilde{\boldsymbol{h}}^{(t)} = \tanh\left(\boldsymbol{W}_{hh}\left(\boldsymbol{r}^{(t)} . \boldsymbol{h}^{(t-1)}\right) + \boldsymbol{W}_{xh}\, \boldsymbol{x}^{(t)} + \boldsymbol{b}_h\right)$$

$$\boldsymbol{h}^{(t)} = \boldsymbol{u}^{(t)} . \tilde{\boldsymbol{h}}^{(t)} + \left(1 - \boldsymbol{u}^{(t)}\right). \boldsymbol{h}^{(t-1)}$$

$$\boldsymbol{u}^{(t)} = \sigma\left(\boldsymbol{W}_{hu}\, \boldsymbol{h}^{(t-1)} + \boldsymbol{W}_{xu}\, \boldsymbol{x}^{(t)} + \boldsymbol{b}_u\right)$$

$$\boldsymbol{r}^{(t)} = \sigma\left(\boldsymbol{W}_{hr}\, \boldsymbol{h}^{(t-1)} + \boldsymbol{W}_{xr}\, \boldsymbol{x}^{(t)} + \boldsymbol{b}_r\right)$$