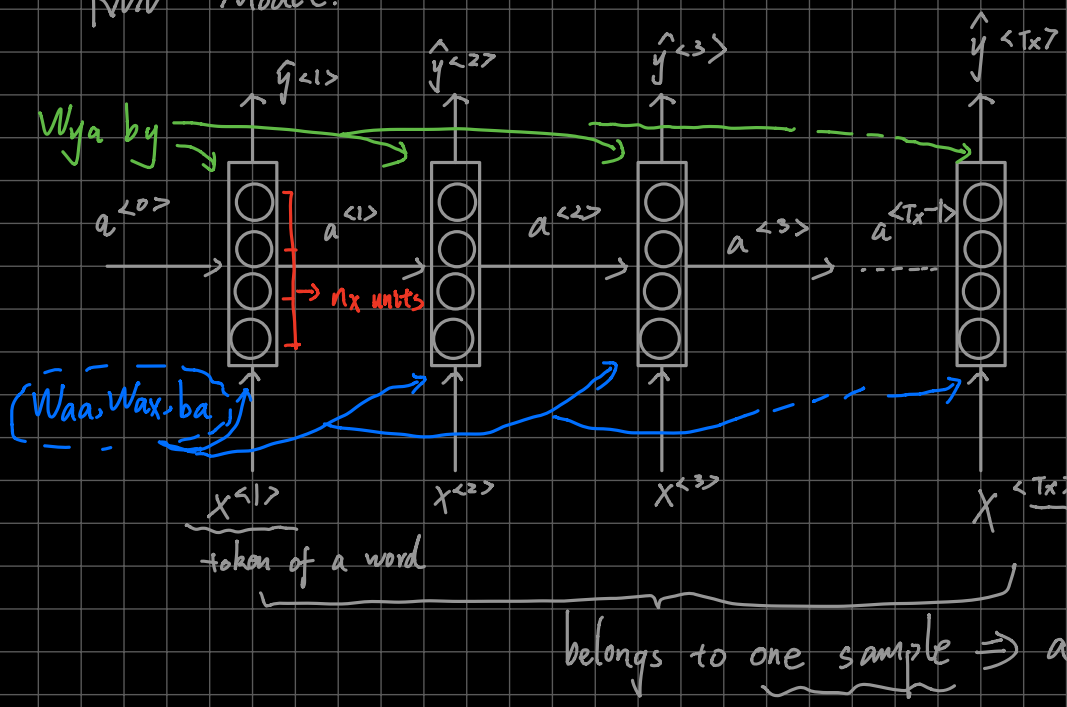
Sequence Model

RNN Model:



belongs to one sample ⇒ a sentence.

Notation:

$(2)[3]\langle 4\rangle$

a_5

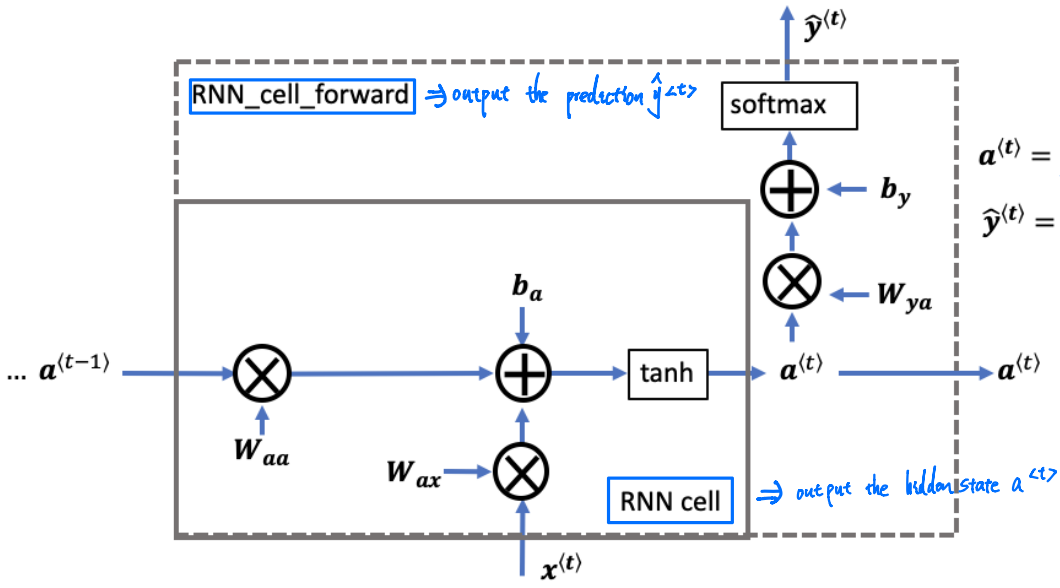
sample 2, RNN layer 3, timestep 4, unit 5.

length of a sequences

△ There is only one $W_{aa}, W_{ax}, b_a, W_{ya}, b_y$.

△ n_x = vocabulary size = $x^{(i)[l]\langle j\rangle}$'s size

Forward propagation:



or relu

$$a^{\langle t\rangle} = \tanh(W_{ax}\, x^{\langle t\rangle} + W_{aa}\, a^{\langle t-1\rangle} + b_a)$$

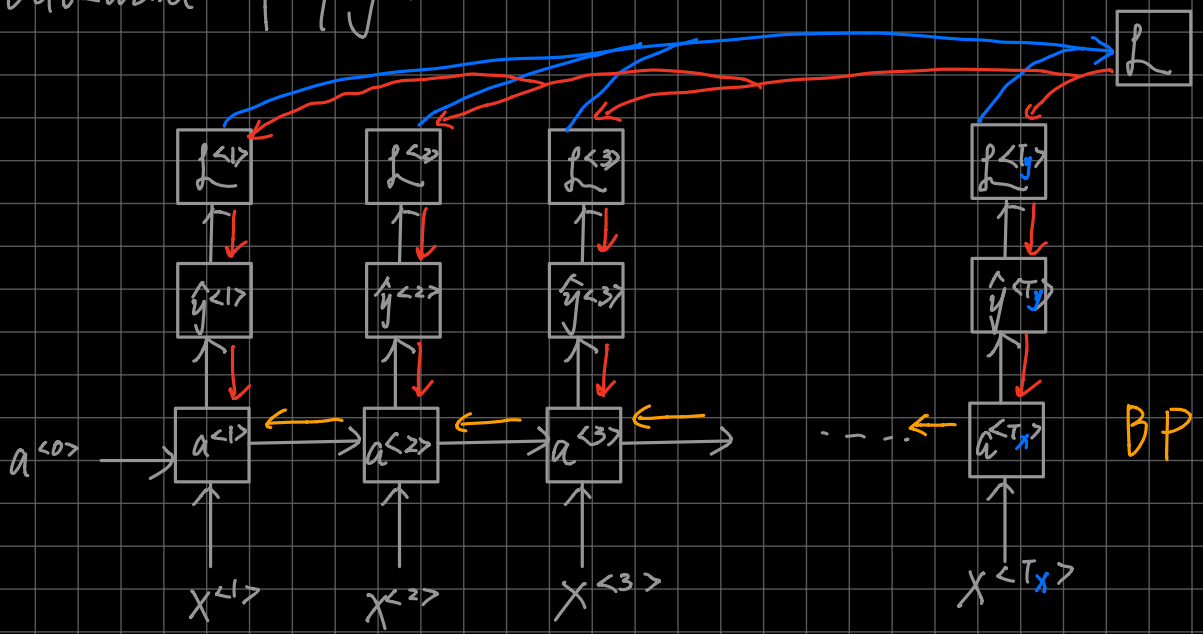
$$\hat{y}^{\langle t\rangle} = softmax(W_{ya} a^{\langle t\rangle} + b_y)$$

or sigmoid

$$\begin{cases} a^{<t>} = g_1\left(W_a \cdot \begin{bmatrix} a^{<t-1>} \\ x^{t} \end{bmatrix} + b_a\right) \\ \hat{y}^{<t>} = g_2\left(W_{ya} \cdot a^{<t>} + b_y\right) \end{cases}$$

$[W_{aa}, W_{ax}]$

Backward propagation:



BP through time.

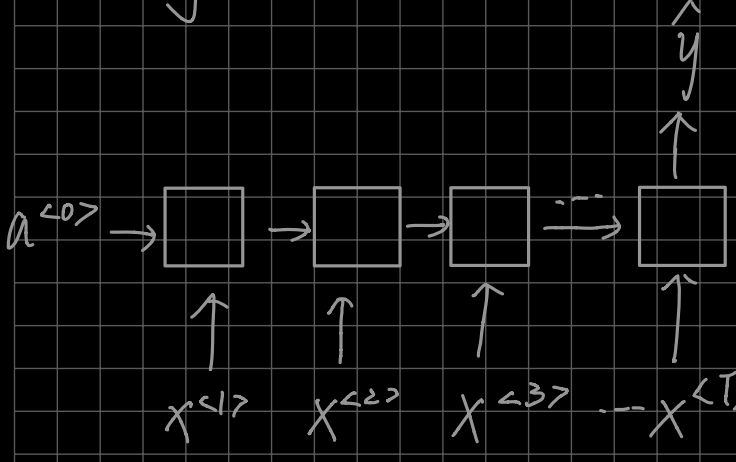
$$\mathcal{L}^{\langle t \rangle} = -\left(y^{\langle t \rangle} \cdot \log \hat{y}^{\langle t \rangle} + (1 - y^{\langle t \rangle}) \cdot \log (1 - \hat{y}^{\langle t \rangle}) \right)$$

$$\mathcal{L} = \sum_{t=1}^{T_x} \mathcal{L}^{\langle t \rangle} \left(\hat{y}^{\langle t \rangle}, y^{\langle t \rangle} \right)$$

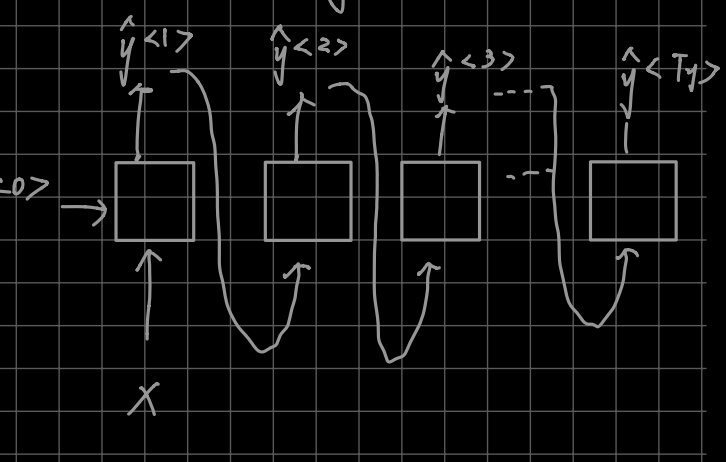
Types of RNNs

- Many-one: Sentiment classification, Video activity recognition
- one-Many: Music generation, Language generation
- Many-Many
  - Precise Matching: DNA sequence analysis, Name entity recognition
  - Non-Precise: Machine translation, Speech recognition

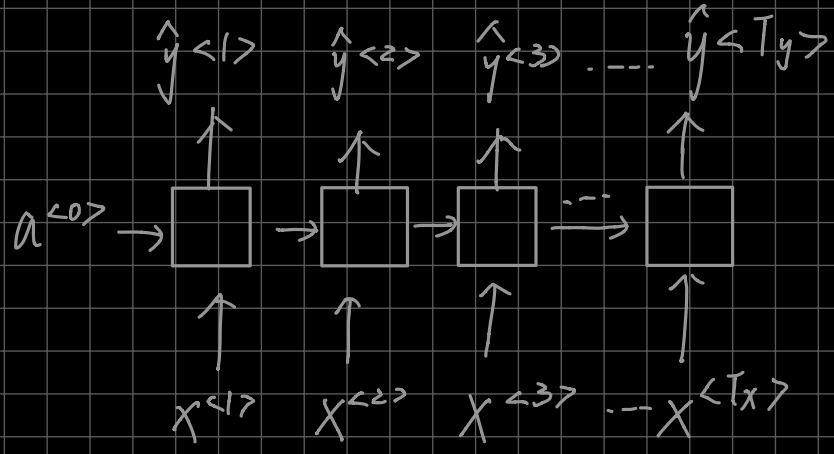
Many to one:

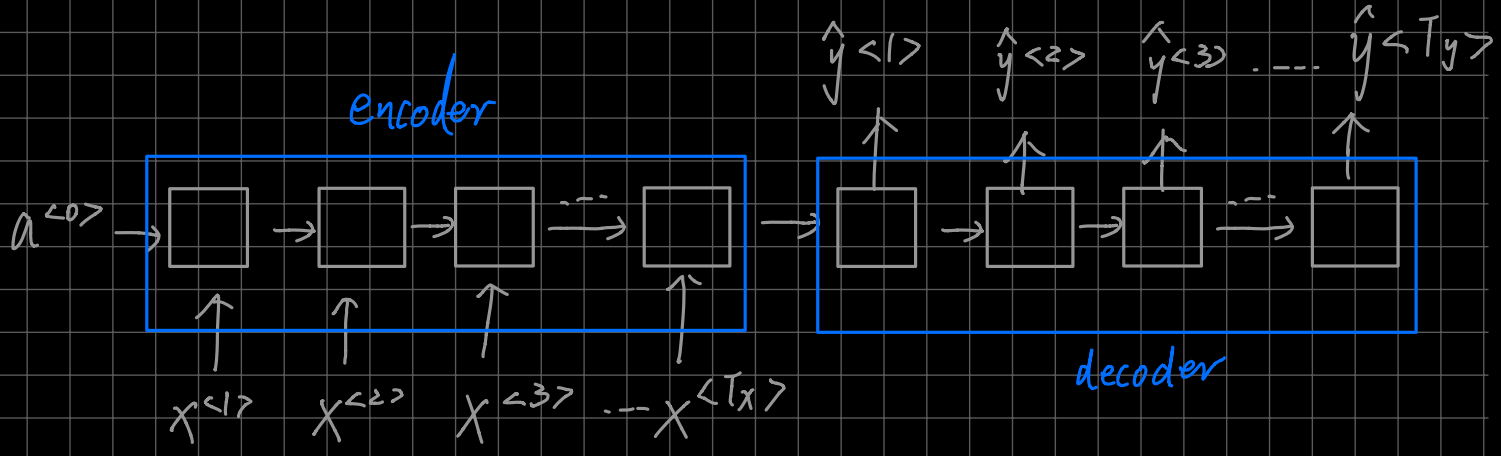


One to Many:

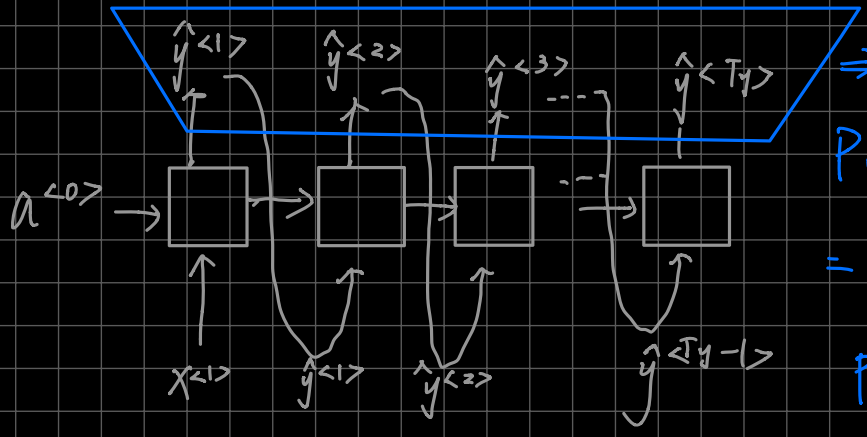


Many to many:





Sampling from a trained RNN:



⇒ Sampling sequence

$P(y^{<1>}, y^{<2>} \dots y^{<T_y>})$

$= P(y^{<1>}) \cdot P(y^{<2>} | y^{<1>}) \cdot$

$P(y^{<3>} | y^{<1>} y^{<2>}) \dots$

条件概率公式

RNN input's shape: $X = (n_x, m, T_x)$

n_x: number of units; m: batch size; T_x: time step

TimeStep input's shape: $X^{<t>} = (n_x, m)$

each time step, we input a batch of data

Hidden state's shape: $a^{<t>} = (n_a, m)$

Prediction: $\hat{y} = (n_y, m, T_y)$

$\hat{y}^{<t>} = (n_y, m)$