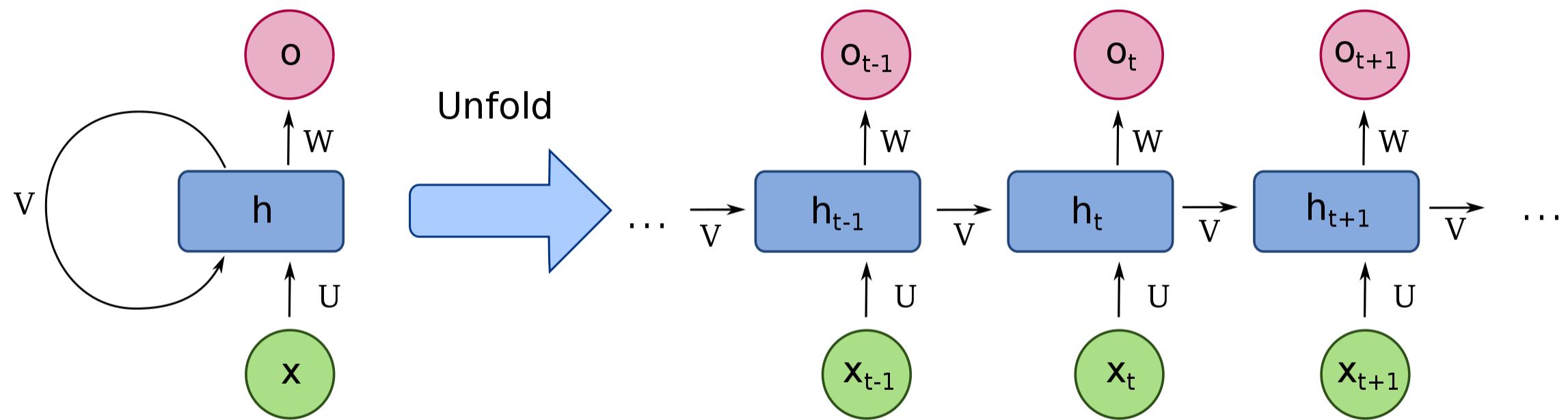


Recurrent Neural Networks

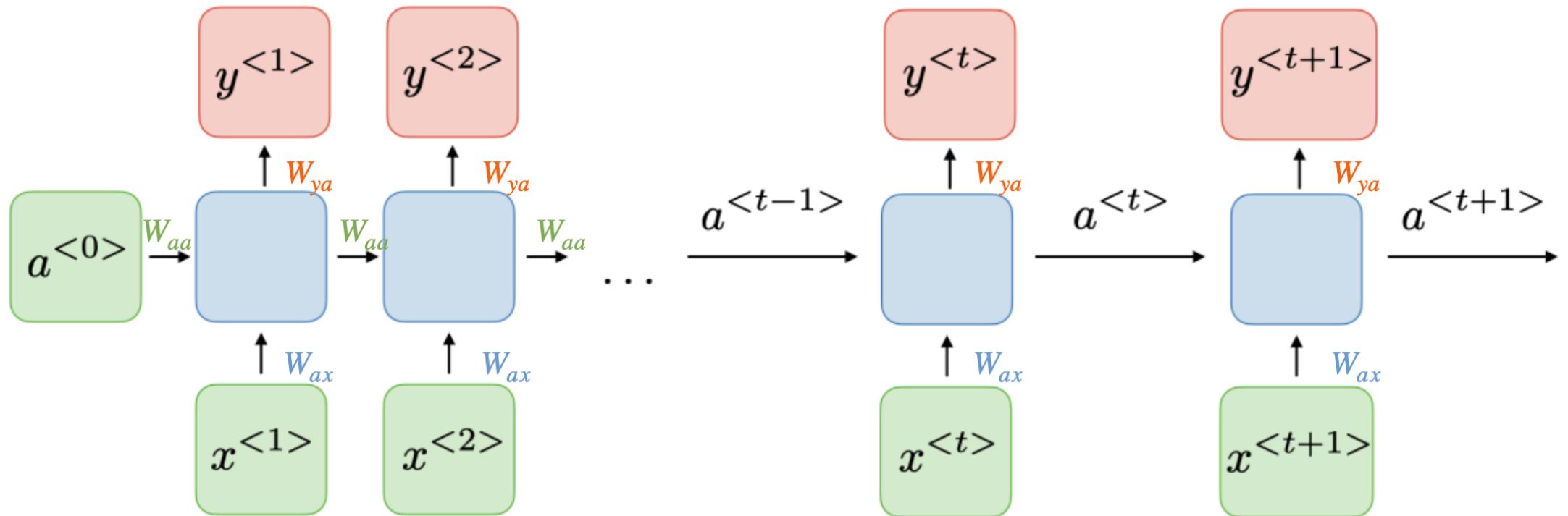
Géraldine Conti, Matthew Vowels, Mykhailo Vladymyrov

Bern Winter School –
Natural Language Processing,
Murren 2024

RNN model



RNN model

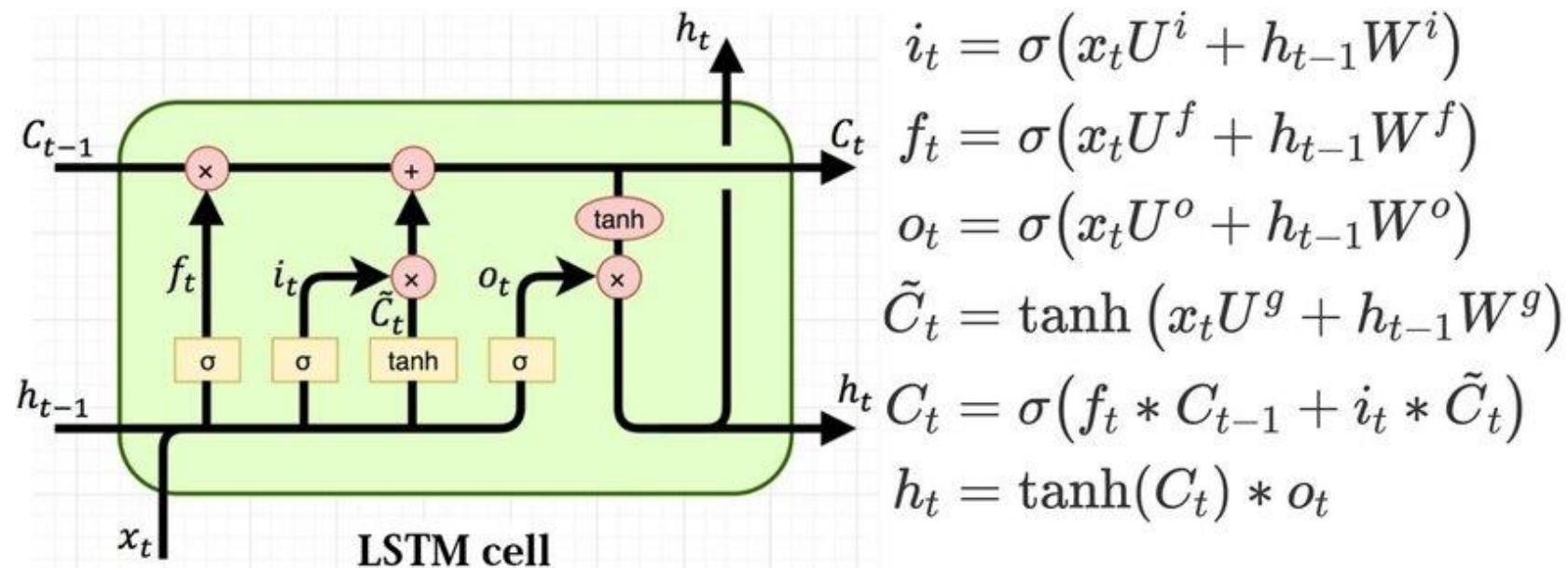


$$a^{<t>} = g_1(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a)$$

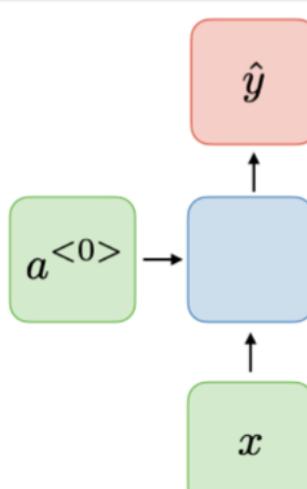
$$y^{<t>} = g_2(W_{ya}a^{<t>} + b_y)$$

$W_{ax}, W_{aa}, W_{ya}, b_a$ and b_y are weights that are shared temporally and g_1, g_2 activation functions

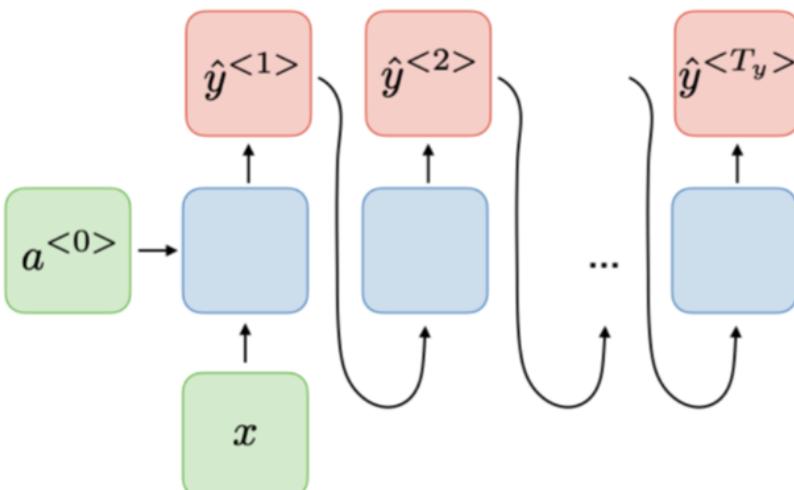
LSTM model



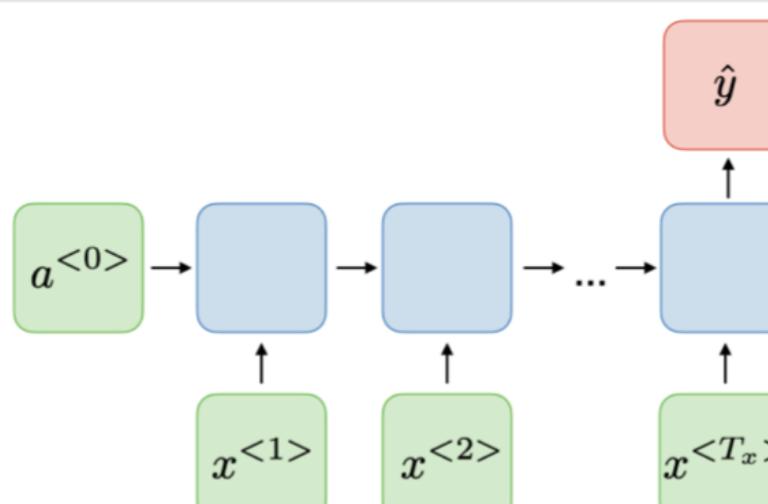
Applications of RNN

Type of RNN	Illustration	Example
One-to-one $T_x = T_y = 1$	 <p>The diagram illustrates a one-to-one Recurrent Neural Network (RNN). It consists of three nodes: an input node labeled x, a hidden state node labeled $a^{<0>}$, and an output node labeled \hat{y}. Arrows indicate the flow of information: an arrow from x to $a^{<0>}$, and another arrow from $a^{<0>}$ to \hat{y}.</p>	Traditional neural network

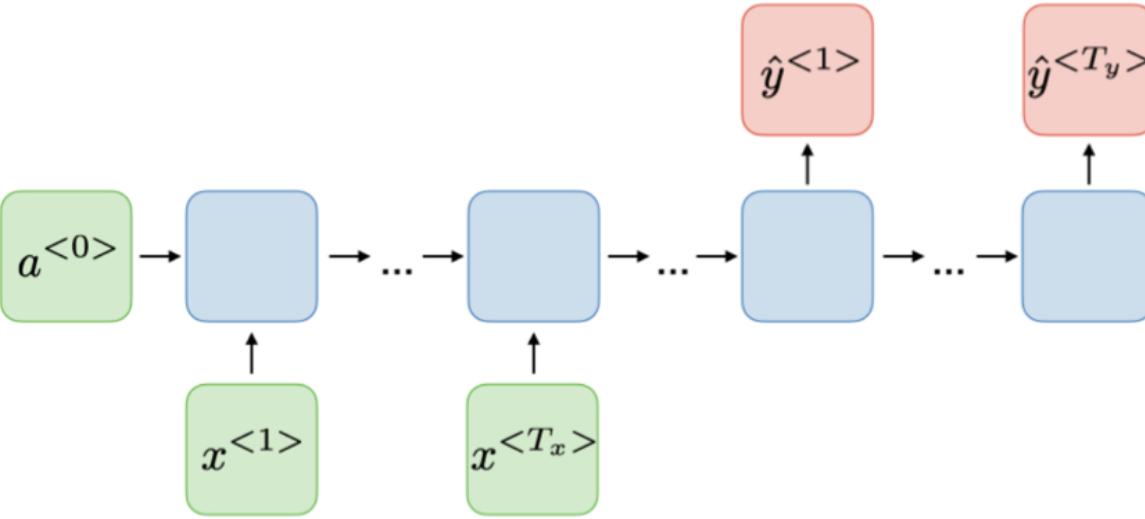
Applications of RNN

Type of RNN	Illustration	Example
One-to-many $T_x = 1, T_y > 1$		Music generation

Applications of RNN

Type of RNN	Illustration	Example
Many-to-one $T_x > 1, T_y = 1$	 <p>The diagram illustrates a Many-to-one Recurrent Neural Network (RNN) architecture. It starts with an initial hidden state $a^{<0>}$ (green box). This state is fed into the first RNN cell (blue box), which then produces the output $x^{<1>} \rightarrow a^{<1>}$. This process repeats for $T_x - 1$ steps, producing hidden states $a^{<2>} \rightarrow \dots \rightarrow a^{<T_x>}$ and corresponding inputs $x^{<2>} \rightarrow \dots \rightarrow x^{<T_x>}$. The final hidden state $a^{<T_x>}$ is passed through a final layer (red box) to produce the predicted output \hat{y}.</p>	Sentiment classification

Applications of RNN

Type of RNN	Illustration	Example
Many-to-many $T_x \neq T_y$	 <p>The diagram illustrates a Many-to-many Recurrent Neural Network (RNN). It consists of two sequences of hidden states. The input sequence (x) starts with a green box labeled $a^{<0>}$ followed by several blue boxes connected by arrows. An upward arrow from the first blue box points to a green box labeled $x^{<1>}$. Another upward arrow from the second blue box points to a green box labeled $x^{<T_x>}$. The output sequence (\hat{y}) starts with a red box labeled $\hat{y}^{<1>}$ followed by several blue boxes connected by arrows. An upward arrow from the first blue box points to a red box labeled $\hat{y}^{<T_y>}$.</p>	Machine translation

LSTM model

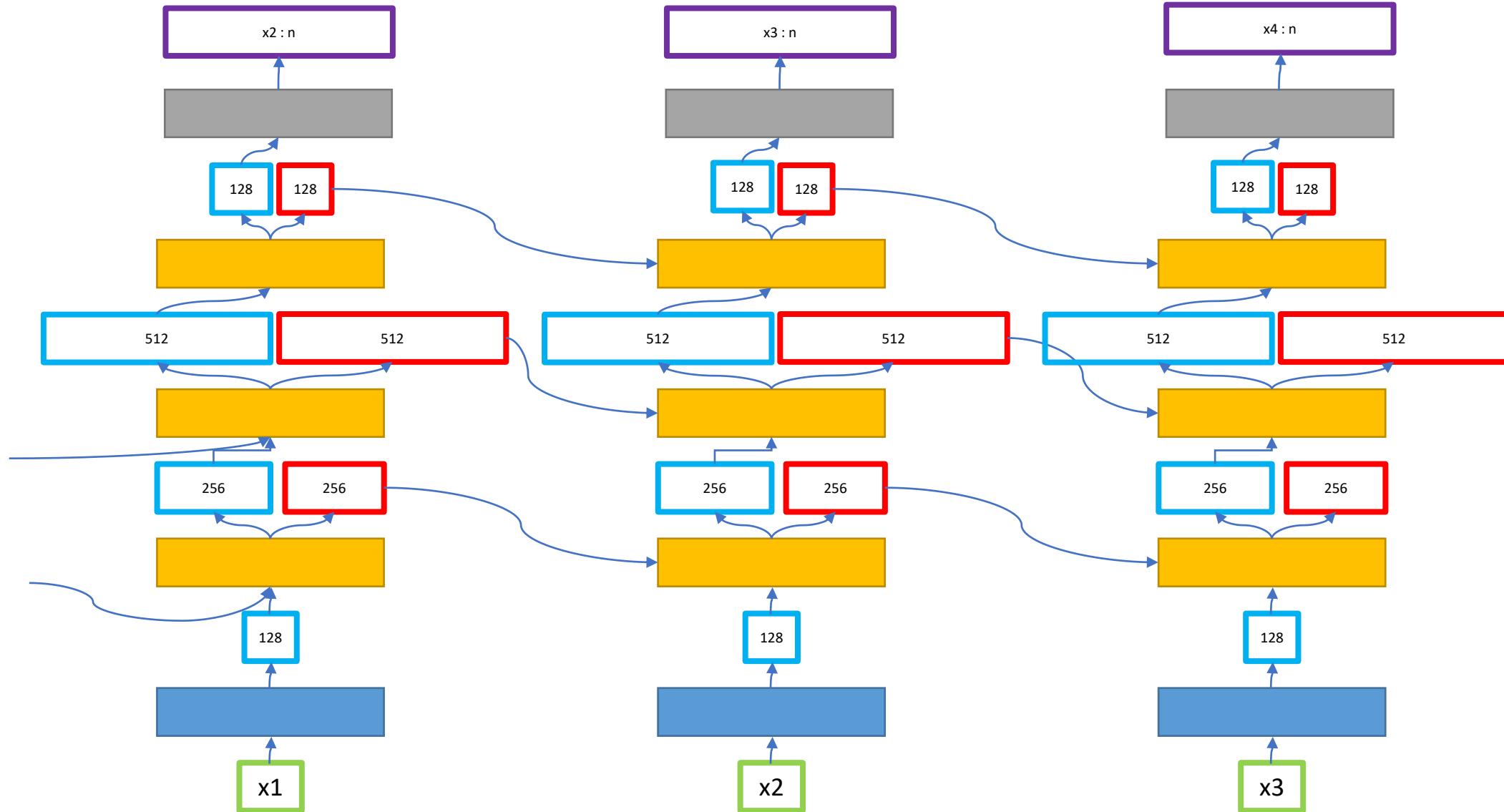
Embedding

LSTM

Dense

Data, size

LSTM state, size



Key elements of neural networks

- Fully connected (dense) layer
- Drop out
- Normalization
- Residual connection
- Attention mechanism