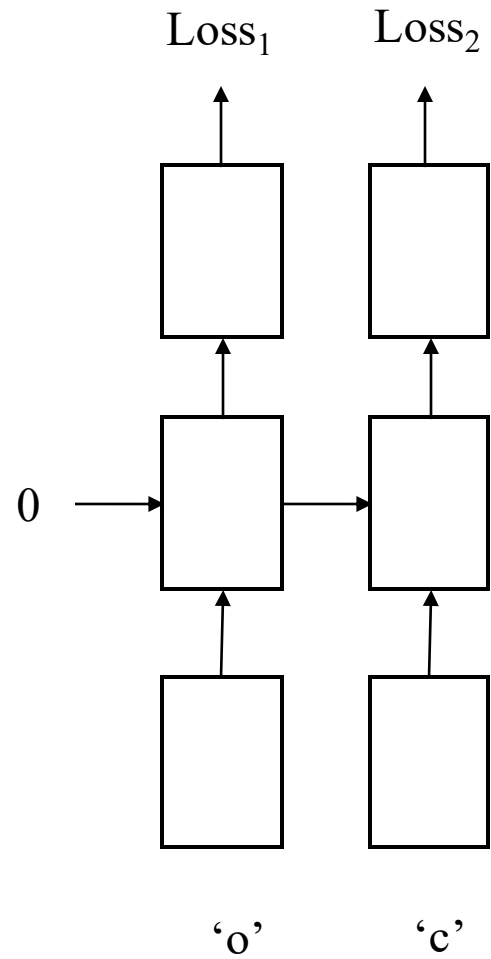


RNN and their computational graph

$$\mathbf{Loss} = \text{Loss}_1 + \text{Loss}_2$$



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

$W_x = 1$

$W_y = -1$

Elman RNN

Activation is $f(x) = x$

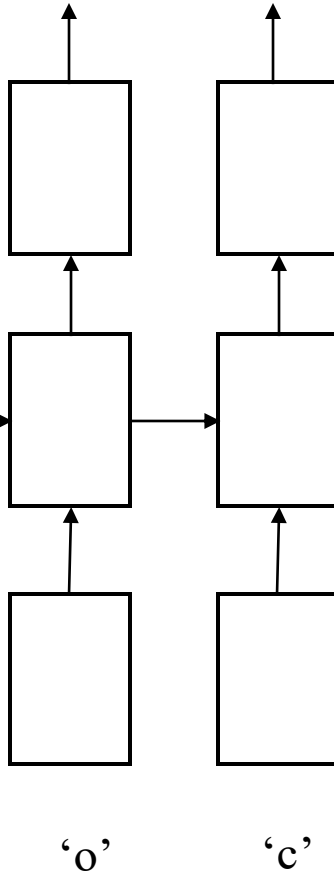
No bias

$$\mathbf{Loss} = \text{Loss}_1 + \text{Loss}_2$$

Loss₁ Loss₂

$\frac{\partial \text{Loss}}{\partial h_0}$

0



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

$W_x = 1$

$W_y = -1$

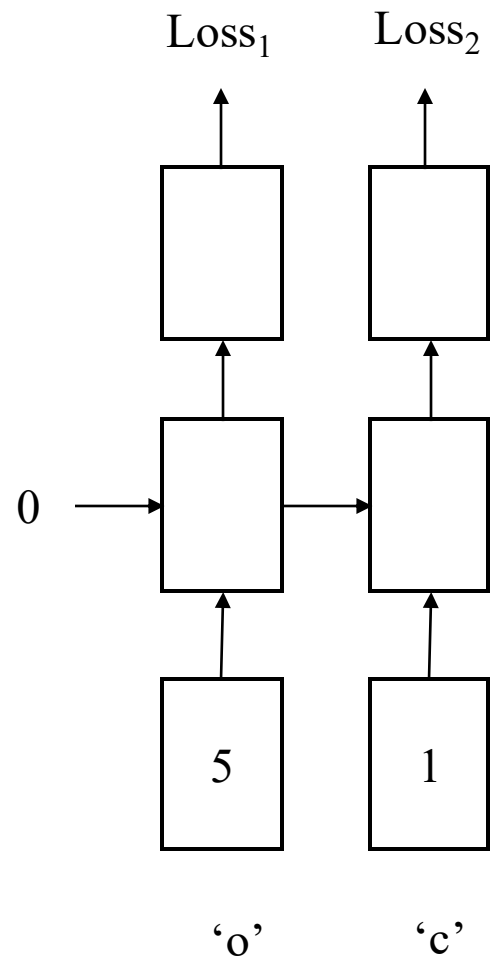
Elman RNN

Activation is $f(x) = x$

No bias

What is the gradient of the loss on h_0 ?

$$\mathbf{Loss} = \text{Loss}_1 + \text{Loss}_2$$



Vocab = ['i' , 'o' , 'c' , 'a']

Embeddings = [-2 , 5 , 1 , 9]

$W_h = -5$

$W_x = 1$

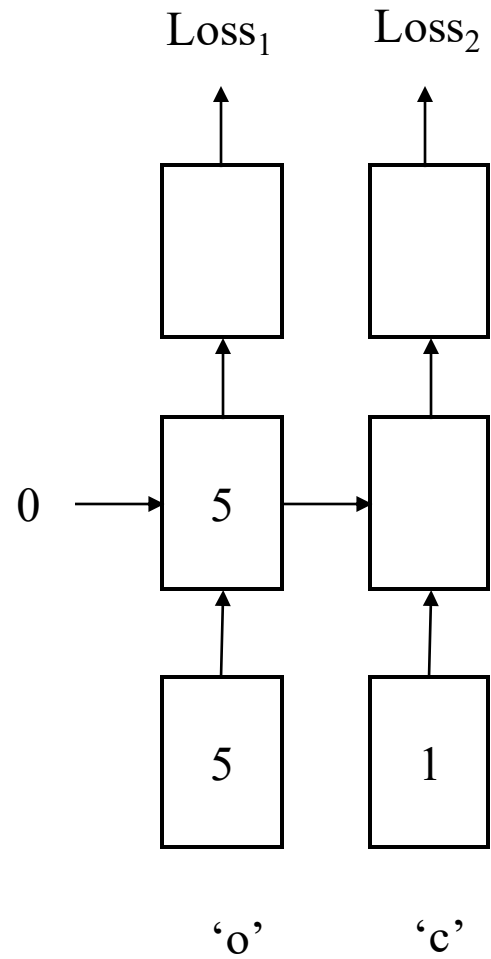
$W_y = -1$

Elman RNN

Activation is $f(x) = x$

No bias

$$\mathbf{Loss} = \text{Loss}_1 + \text{Loss}_2$$



$$\delta (W_h \cdot h_{t-1} + W_x \cdot x)$$

$$\begin{pmatrix} W_h & W_x \end{pmatrix} \begin{pmatrix} h_{t-1} \\ x \end{pmatrix}$$

Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

$W_x = 1$

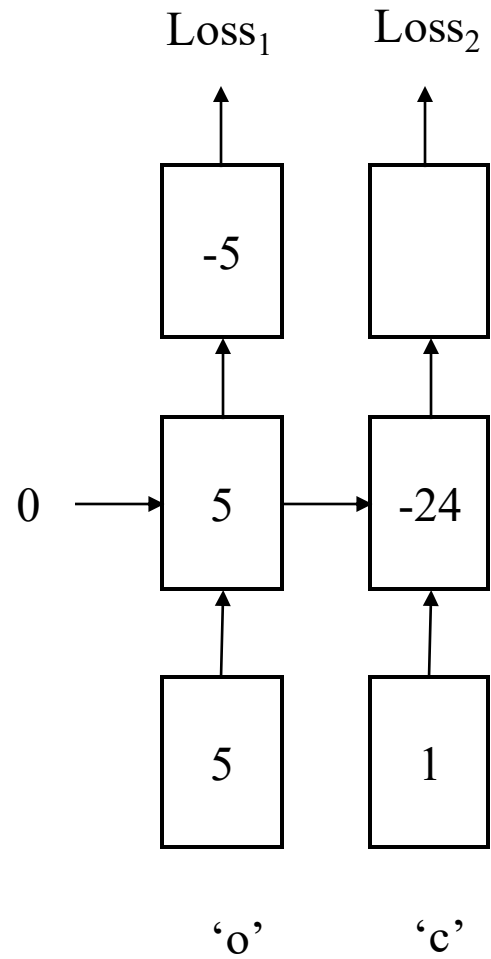
$W_y = -1$

Elman RNN

Activation is $f(x) = x$

No bias

$$\mathbf{Loss} = \text{Loss}_1 + \text{Loss}_2$$



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

$W_x = 1$

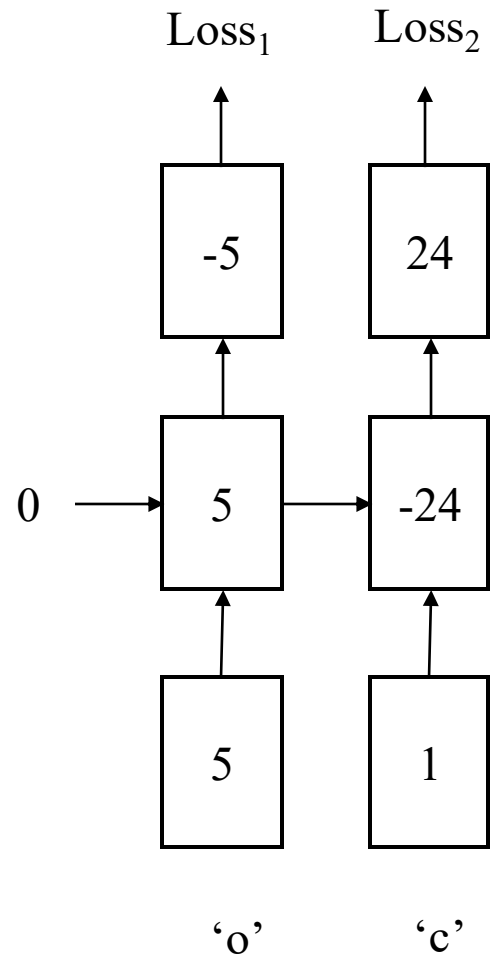
$W_y = -1$

Elman RNN

Activation is $f(x) = x$

No bias

$$\mathbf{Loss} = \text{Loss}_1 + \text{Loss}_2$$



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

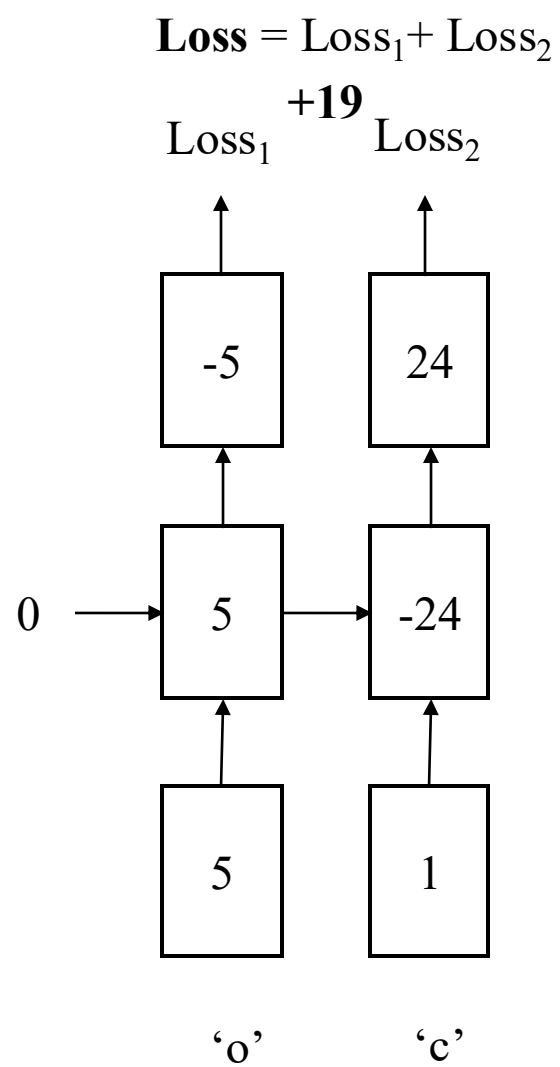
$W_x = 1$

$W_y = -1$

Elman RNN

Activation is $f(x) = x$

No bias



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

$W_x = 1$

$W_y = -1$

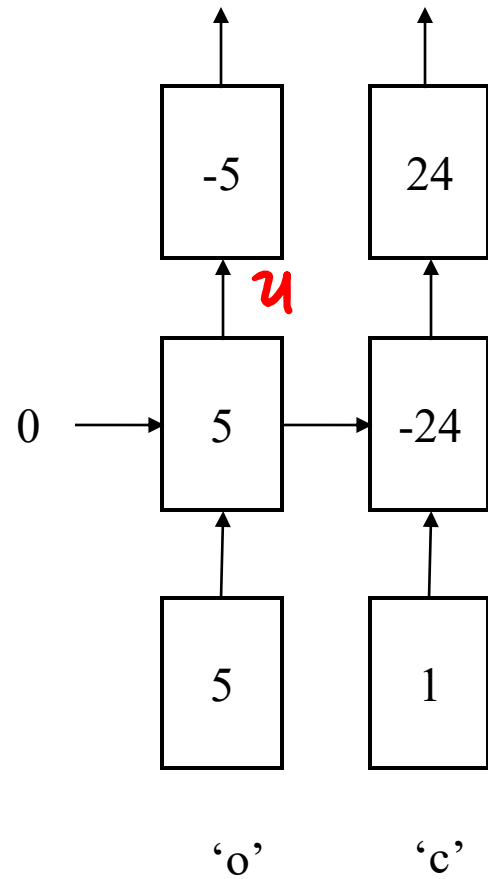
Elman RNN

Activation is $f(x) = x$

No bias

$$\text{Loss} = \text{Loss}_1 + \text{Loss}_2$$

$$\text{Loss}_1 + 19 + \text{Loss}_2$$



$$y = W_y \cdot u$$

$$-1 \cdot 5$$

$$\frac{24}{24} = -1$$

Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

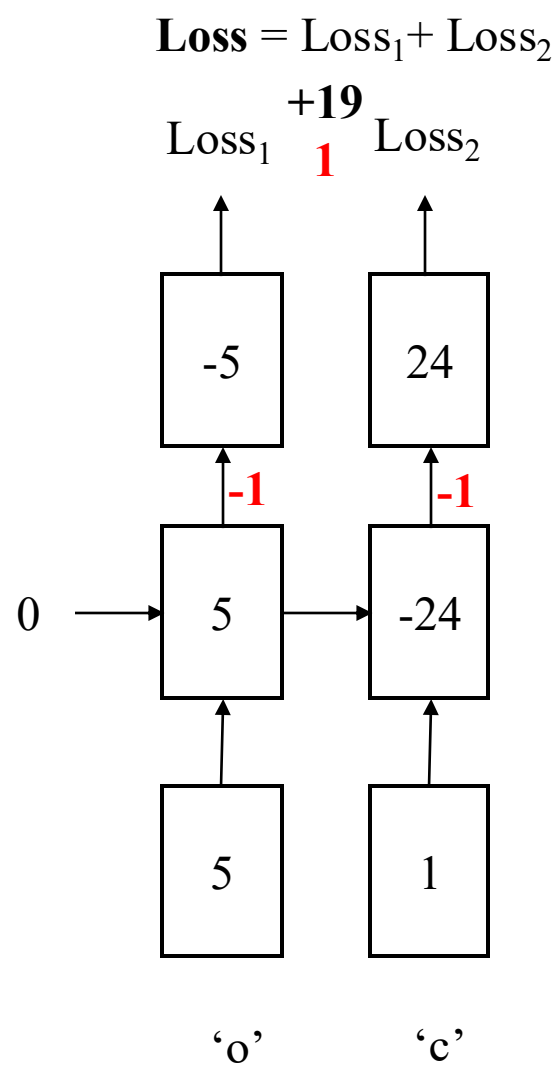
$W_x = 1$

$W_y = -1$

Elman RNN

Activation is $f(x) = x$

No bias



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

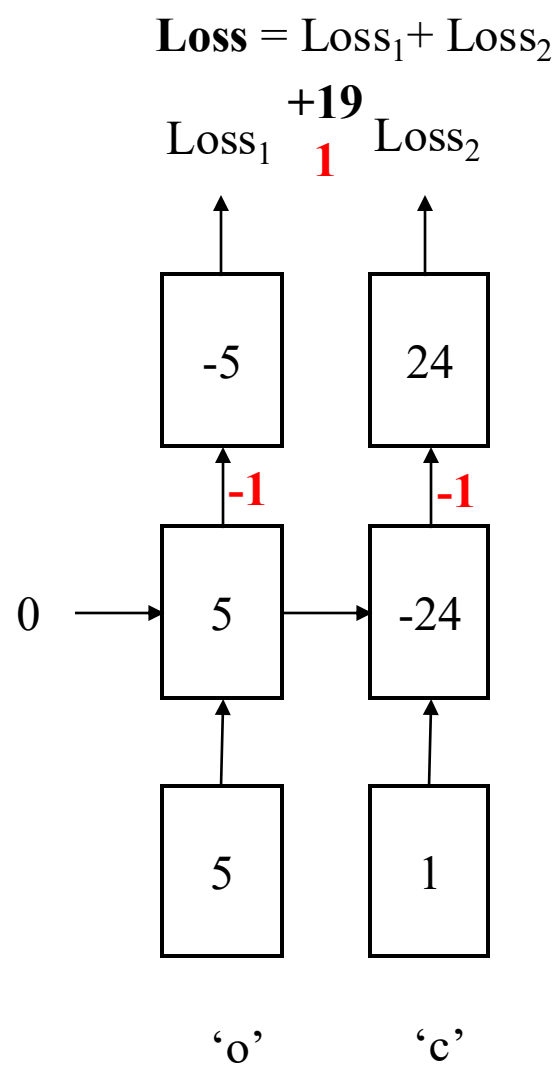
$W_x = 1$

$W_y = -1$

Elman RNN

Activation is $f(x) = x$

No bias



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

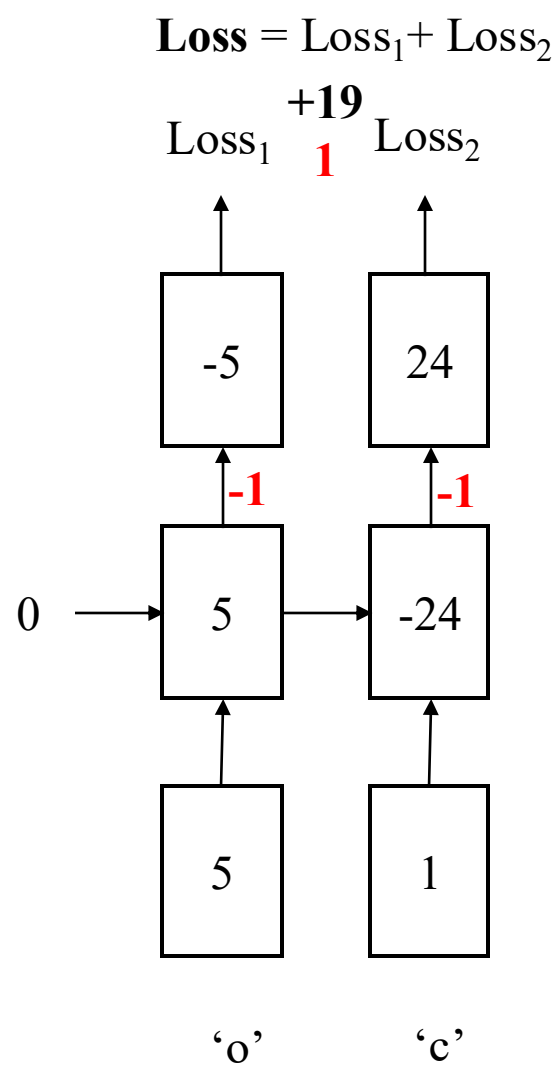
$W_x = 1$

$W_y = -1$

Elman RNN
 Activation is $f(x) = x$
 No bias

$$f(h_{t-1}, \dots) = W_h \cdot h_{t-1} + W_x \cdot x_t$$

$$\frac{\partial f(h_{t-1})}{\partial h_{t-1}} = W_h$$



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

$W_x = 1$

$W_y = -1$

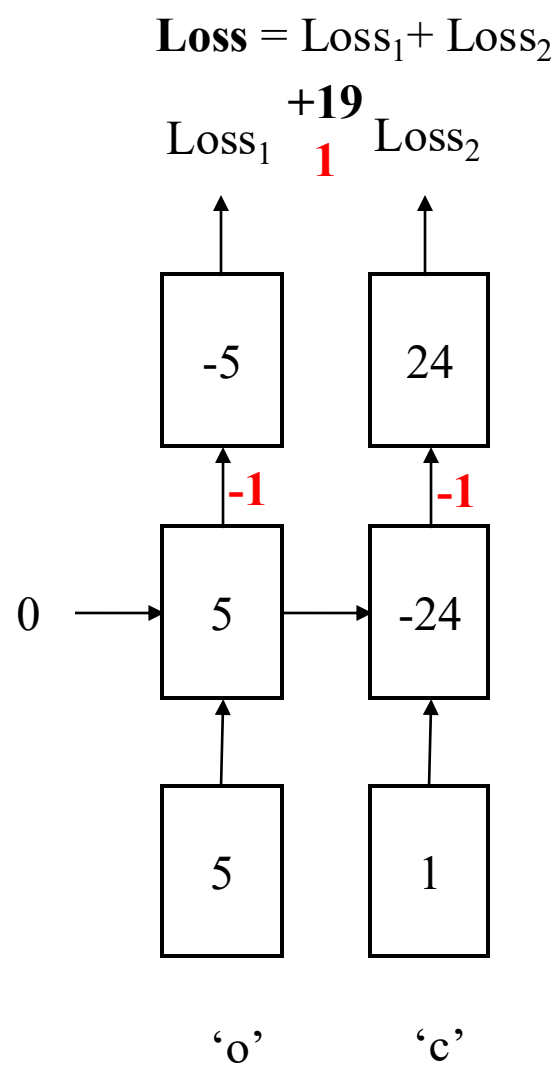
Elman RNN

Activation is $f(x) = x$

No bias

$$f(h_{t-1}, --) = W_h \cdot h_{t-1} + W_x \cdot x_t$$

$$\frac{\partial f(h_{t-1})}{\partial h_{t-1}} = W_h - 5$$



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

$W_x = 1$

$W_y = -1$

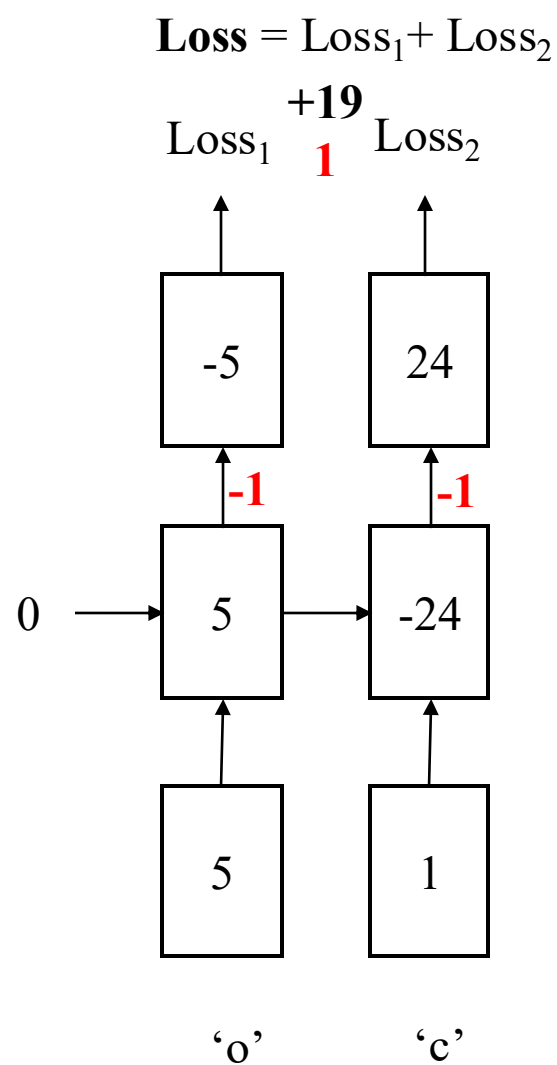
Elman RNN

Activation is $f(x) = x$

No bias

$$f(h_{t-1}, --) = W_h \cdot h_{t-1} + W_y \cdot x_t$$

$$\frac{\partial f(h_{t-1})}{\partial h_{t-1}} = W_h - 5 \cdot -1$$



Vocab = [‘i’, ‘o’, ‘c’, ‘a’]

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

$W_x = 1$

$W_y = -1$

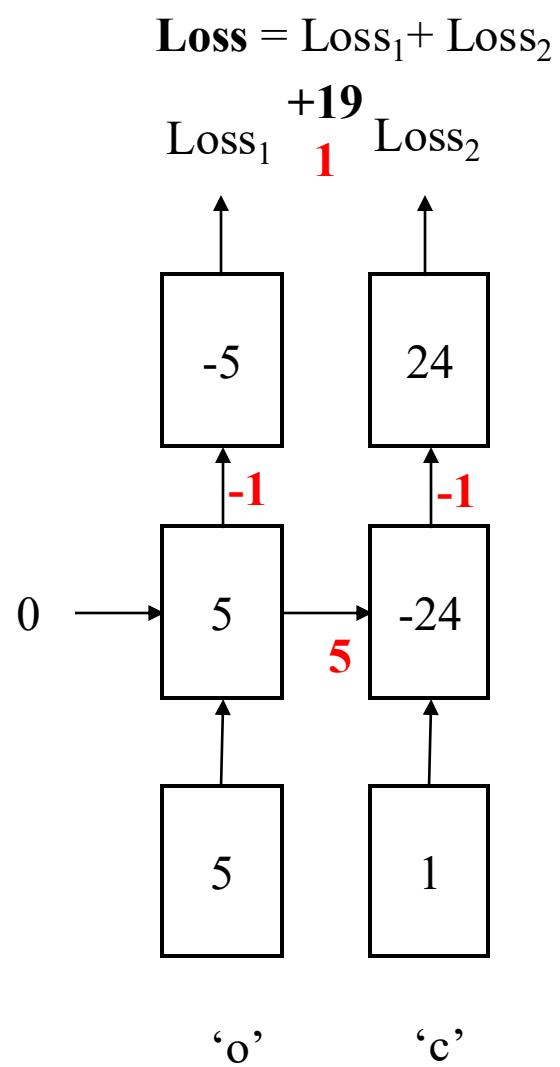
Elman RNN

Activation is $f(x) = x$

No bias

$$f(h_{t-1}, --) = W_h \cdot h_{t-1} + W_y \cdot x_t$$

$$\frac{\partial f(h_{t-1})}{\partial h_{t-1}} = W_h \underbrace{-5}_{\text{local}} \cdot \underbrace{-1}_{\text{upstream}}$$



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

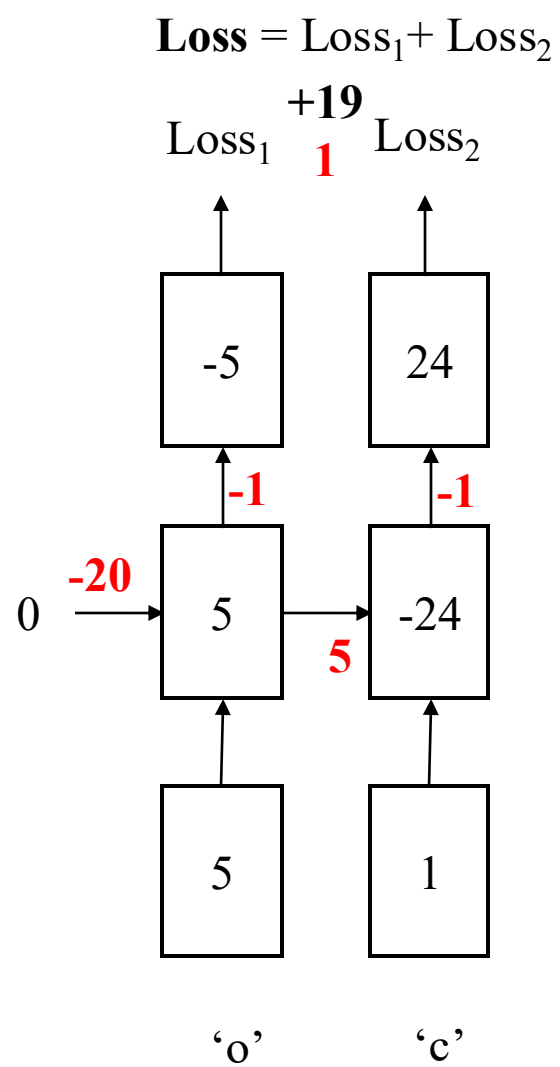
$W_x = 1$

$W_y = -1$

Elman RNN

Activation is $f(x) = x$

No bias



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

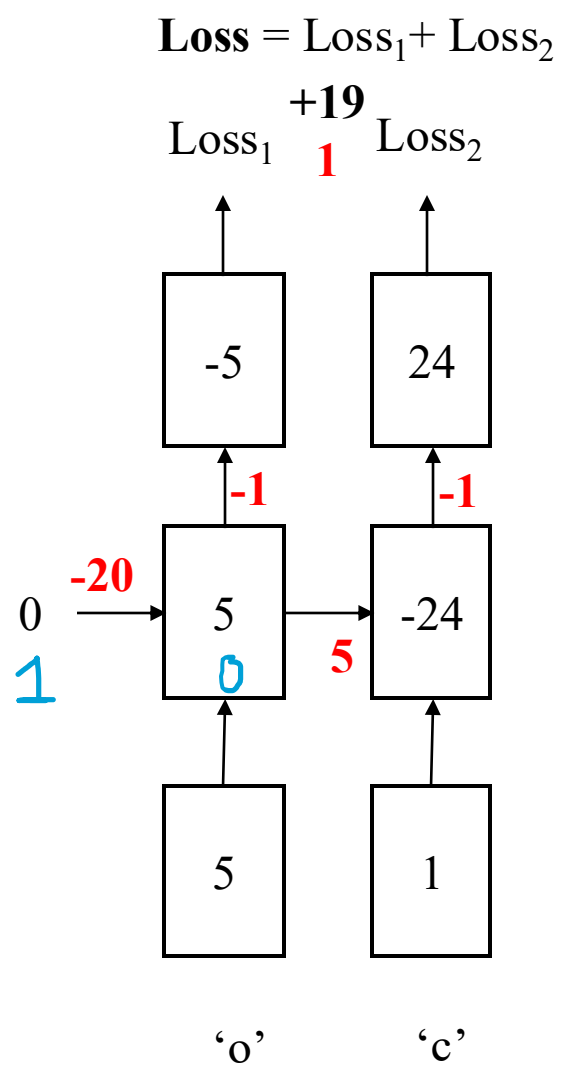
$W_x = 1$

$W_y = -1$

Elman RNN

Activation is $f(x) = x$

No bias



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

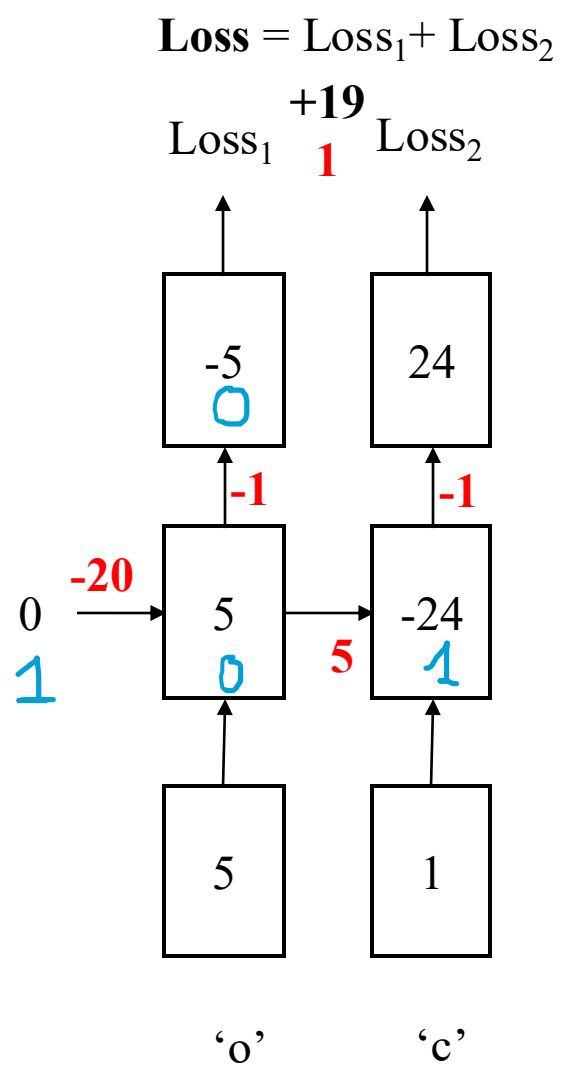
$W_x = 1$

$W_y = -1$

Elman RNN

Activation is $f(x) = x$

No bias



Vocab = ['i', 'o', 'c', 'a']

Embeddings = [-2, 5, 1, 9]

$W_h = -5$

$W_x = 1$

$W_y = -1$

Elman RNN

Activation is $f(x) = x$

No bias



Elman RNN
Activation is $f(x)=x$
No bias