Problem statement: Use an RNN to predict the sentiment of the following document

I love cats → 1 (positive)

Epoch 1

(S) Sequence Length = 4

(E) Embedding Size = 3

(H) Hidden State Dimensions = 2

(Y) Output Dimension = 1

1 **Compute Hidden State Contribution From Input**

$$\begin{bmatrix}0 & 2 & 3\end{bmatrix} \times \begin{bmatrix}3 & 1\\ -1 & -2\\ 0 & 1\end{bmatrix} = \begin{bmatrix}-2 & -1\end{bmatrix}$$

Input Token “I”
1 x 3
(1 token x E)

EH Weights
3 x 2
(E x H)

Proposed Hidden State From Input
1 x 2

Sequence Step 1/4

Epoch 1

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} 1 & 2 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

Previous Hidden State	HH Weights	Proposed Hidden State From Prev Hidden State
1 x 2	2 x 2 (H x H)	1 x 2

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} 0 & 2 & 3 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & -1 \end{bmatrix}$$

Input Token “I”	EH Weights	Proposed Hidden State From Input
1 x 3 (1 token x E)	3 x 2 (E x H)	1 x 2

Sequence Step 1/4

(S) Sequence Length = 4

(E) Embedding Size = 3

(H) Hidden State Dimensions = 2

(Y) Output Dimension = 1

Epoch 1

2 Compute Hidden State Contribution From Previous Hidden State

$$\underbrace{\begin{bmatrix} 1 & 2 \end{bmatrix}}_{\text{Previous Hidden State } 1 \times 2} \times \underbrace{\begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix}}_{\text{HH Weights } 2 \times 2 \ (H \times H)} = \underbrace{\begin{bmatrix} 1 & 2 \end{bmatrix}}_{\text{Proposed Hidden State From Prev Hidden State } 1 \times 2}$$

1 Compute Hidden State Contribution From Input

$$\underbrace{\begin{bmatrix} 0 & 2 & 3 \end{bmatrix}}_{\text{Input Token “I” } 1 \times 3 \ (1 \text{ token} \times E)} \times \underbrace{\begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix}}_{\text{EH Weights } 3 \times 2 \ (E \times H)} = \underbrace{\begin{bmatrix} -2 & -1 \end{bmatrix}}_{\text{Proposed Hidden State From Input } 1 \times 2}$$

3 Update Hidden State

$$\tanh\Big(\underbrace{\begin{bmatrix} -2 & -1 \end{bmatrix}}_{\text{Proposed Hidden State From Input}} + \underbrace{\begin{bmatrix} 1 & 2 \end{bmatrix}}_{\text{Proposed Hidden State From Prev Hidden State}}\Big) = \begin{bmatrix} -0.76 & 0.76 \end{bmatrix} \sim \underbrace{\begin{bmatrix} -1 & 1 \end{bmatrix}}_{\text{New Hidden State } 1 \times 2}$$

Sequence Step 1/4

Epoch 1

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} 0 & 2 & 3 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & -1 \end{bmatrix}$$

Input Token “I”
1 x 3
(1 token x E)

EH Weights
3 x 2
(E x H)

Proposed Hidden State From Input
1 x 2

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

Previous Hidden State
1 x 2

HH Weights
2 x 2
(H x H)

Proposed Hidden State From Prev Hidden State
1 x 2

3 Update Hidden State

$$\tanh\left(\begin{bmatrix} -2 & -1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \end{bmatrix} \right) = \begin{bmatrix} -0.76 & 0.76 \end{bmatrix} \sim \begin{bmatrix} -1 & 1 \end{bmatrix}$$

Proposed Hidden State From Input

Proposed Hidden State From Prev Hidden State

New Hidden State
1 x 2

4 Compute Output

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & -2 \end{bmatrix} = \begin{bmatrix} -4 \end{bmatrix}$$

New Hidden State
1 x 2

HY Weights
2 x 1
(H x Y)

Sequence Step 1/4

Epoch 1

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} 0 & 2 & 3 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & -1 \end{bmatrix}$$

Input Token "I"
1 x 3
(1 token x E)

EH Weights
3 x 2
(E x H)

Proposed Hidden State From Input
1 x 2

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} 1 & 2 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

Previous Hidden State
1 x 2

HH Weights
2 x 2
(H x H)

Proposed Hidden State From Prev Hidden State
1 x 2

3 Update Hidden State

$$\tanh\left(\begin{bmatrix} -2 & -1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \end{bmatrix}\right) = \begin{bmatrix} -0.76 & 0.76 \end{bmatrix} \sim \begin{bmatrix} -1 & 1 \end{bmatrix}$$

Proposed Hidden State From Input

Proposed Hidden State From Prev Hidden State

New Hidden State
1 x 2

4 Compute Output

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & -2 \end{bmatrix} = \begin{bmatrix} -4 \end{bmatrix}$$

New Hidden State
1 x 2

HY Weights
2 x 1
(H x Y)

5 Compute Output

$$\mathrm{sigmoid}\left(\begin{bmatrix} -4 \end{bmatrix}\right) = \begin{bmatrix} 0.18 \end{bmatrix}$$

Our predicted sentiment at sequence step 1

Sequence Step 1/4

Epoch 1

(S) Sequence Length = 4

(E) Embedding Size = 3

(H) Hidden State Dimensions = 2

(Y) Output Dimension = 1

1 **Compute Hidden State Contribution From Input**

$$\begin{bmatrix} 2 & -1 & 2 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 6 \end{bmatrix}$$

Input Token
“Love”
1 x 3
(1 token x E)

EH Weights
3 x 2
(E x H)

Proposed Hidden State From Input
1 x 2

Sequence Step 2/4

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

Epoch 1

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} -4 & -3 \end{bmatrix}$$

Previous Hidden State
1 x 2

HH Weights
2 x 2
(H x H)

Proposed Hidden State From Prev Hidden State
1 x 2

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} 2 & -1 & 2 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 6 \end{bmatrix}$$

Input Token
“Love”
1 x 3
(1 token x E)

EH Weights
3 x 2
(E x H)

Proposed Hidden State From Input
1 x 2

Sequence Step 2/4

Epoch 1

(S) Sequence Length = 4

(E) Embedding Size = 3

(H) Hidden State Dimensions = 2

(Y) Output Dimension = 1

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} -4 & -3 \end{bmatrix}$$

Previous Hidden State 1 x 2

HH Weights 2 x 2 (H x H)

Proposed Hidden State From Prev Hidden State 1 x 2

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} 2 & -1 & 2 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 6 \end{bmatrix}$$

Input Token "Love" 1 x 3 (1 token x E)

EH Weights 3 x 2 (E x H)

Proposed Hidden State From Input 1 x 2

3 Update Hidden State

$$\tanh\left(\begin{bmatrix} 7 & 6 \end{bmatrix} + \begin{bmatrix} -4 & -3 \end{bmatrix}\right) = \begin{bmatrix} .99 & .99 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \end{bmatrix}$$

Proposed Hidden State From Input

Proposed Hidden State From Prev Hidden State

New Hidden State 1 x 2

Sequence Step 2/4

Epoch 1

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

1 Compute Hidden State Contribution From Input

2	-1	2

Input Token "Love" 1 x 3 (1 token x E)

X

3	1
-1	-2
0	1

EH Weights 3 x 2 (E x H)

=

7	6

Proposed Hidden State From Input 1 x 2

2 Compute Hidden State Contribution From Previous Hidden State

-1	1

Previous Hidden State 1 x 2

X

3	1
-1	-2

HH Weights 2 x 2 (H x H)

=

-4	-3

Proposed Hidden State From Prev Hidden State 1 x 2

3 Update Hidden State

$$\tanh(\begin{bmatrix}7 & 6\end{bmatrix} + \begin{bmatrix}-4 & -3\end{bmatrix}) = \begin{bmatrix}.99 & .99\end{bmatrix} \sim \begin{bmatrix}1 & 1\end{bmatrix}$$

Proposed Hidden State From Input

Proposed Hidden State From Prev Hidden State

New Hidden State 1 x 2

4 Compute Output

1	1

New Hidden State 1 x 2

X

2	-2

HY Weights 2 x 1 (H x Y)

=

0

Sequence Step 2/4

Epoch 1

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} 2 & -1 & 2 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 6 \end{bmatrix}$$

Input Token "Love" 1 x 3 (1 token x E)

EH Weights 3 x 2 (E x H)

Proposed Hidden State From Input 1 x 2

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} -4 & -3 \end{bmatrix}$$

Previous Hidden State 1 x 2

HH Weights 2 x 2 (H x H)

Proposed Hidden State From Prev Hidden State 1 x 2

3 Update Hidden State

$$\tanh\left(\begin{bmatrix} 7 & 6 \end{bmatrix} + \begin{bmatrix} -4 & -3 \end{bmatrix}\right) = \begin{bmatrix} .99 & .99 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \end{bmatrix}$$

Proposed Hidden State From Input

Proposed Hidden State From Prev Hidden State

New Hidden State 1 x 2

4 Compute Output

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & -2 \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

New Hidden State 1 x 2

HY Weights 2 x 1 (H x Y)

5 Compute Output

$$\text{sigmoid}\left(\begin{bmatrix} 0 \end{bmatrix}\right) = \begin{bmatrix} 0.5 \end{bmatrix}$$

Our predicted sentiment at sequence step 2

Sequence Step 2/4

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

Epoch 1

1 **Compute Hidden State Contribution From Input**

$$\begin{bmatrix} -1 & -1 & -2 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 3 \end{bmatrix}$$

Input Token
"Cats"
1 x 3
(1 token x E)

EH Weights
3 x 2
(E x H)

Proposed Hidden State From Input
1 x 2

Sequence Step 3/4

(S) Sequence Length = 4

(E) Embedding Size = 3

(H) Hidden State Dimensions = 2

(Y) Output Dimension = 1

Epoch 1

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \end{bmatrix}$$

Previous Hidden State 1 x 2

HH Weights 2 x 2 (H x H)

Proposed Hidden State From Prev Hidden State 1 x 2

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} -1 & -1 & -2 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 3 \end{bmatrix}$$

Input Token "Cats" 1 x 3 (1 token x E)

EH Weights 3 x 2 (E x H)

Proposed Hidden State From Input 1 x 2

Sequence Step 3/4

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

Epoch 1

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \end{bmatrix}$$

Previous Hidden State 1 x 2

HH Weights 2 x 2 (H x H)

Proposed Hidden State From Prev Hidden State 1 x 2

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} -1 & -1 & -2 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 3 \end{bmatrix}$$

Input Token "Cats" 1 x 3 (1 token x E)

EH Weights 3 x 2 (E x H)

Proposed Hidden State From Input 1 x 2

3 Update Hidden State

$$\tanh\left(\begin{bmatrix} -2 & 3 \end{bmatrix} + \begin{bmatrix} 2 & -1 \end{bmatrix} \right) = \begin{bmatrix} 0 & .96 \end{bmatrix} \sim \begin{bmatrix} 0 & 1 \end{bmatrix}$$

Proposed Hidden State From Input

Proposed Hidden State From Prev Hidden State

New Hidden State 1 x 2

Sequence Step 3/4

Epoch 1

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \end{bmatrix}$$

Previous Hidden State 1 x 2 — HH Weights 2 x 2 (H x H) — Proposed Hidden State From Prev Hidden State 1 x 2

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} -1 & -1 & -2 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 3 \end{bmatrix}$$

Input Token "Cats" 1 x 3 (1 token x E) — EH Weights 3 x 2 (E x H) — Proposed Hidden State From Input 1 x 2

3 Update Hidden State

$$\tanh\left(\begin{bmatrix} -2 & 3 \end{bmatrix} + \begin{bmatrix} 2 & -1 \end{bmatrix} \right) = \begin{bmatrix} 0 & .96 \end{bmatrix} \sim \begin{bmatrix} 0 & 1 \end{bmatrix}$$

Proposed Hidden State From Input — Proposed Hidden State From Prev Hidden State — New Hidden State 1 x 2

4 Compute Output

$$\begin{bmatrix} 0 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & -2 \end{bmatrix} = \begin{bmatrix} -2 \end{bmatrix}$$

New Hidden State 1 x 2 — HY Weights 2 x 1 (H x Y)

Sequence Step 3/4

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

Epoch 1

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} -1 & -1 & -2 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 3 \end{bmatrix}$$

Input Token "Cats" 1 x 3 (1 token x E)

EH Weights 3 x 2 (E x H)

Proposed Hidden State From Input 1 x 2

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \end{bmatrix}$$

Previous Hidden State 1 x 2

HH Weights 2 x 2 (H x H)

Proposed Hidden State From Prev Hidden State 1 x 2

3 Update Hidden State

$$\tanh\left(\begin{bmatrix} -2 & 3 \end{bmatrix} + \begin{bmatrix} 2 & -1 \end{bmatrix}\right) = \begin{bmatrix} 0 & .96 \end{bmatrix} \sim \begin{bmatrix} 0 & 1 \end{bmatrix}$$

Proposed Hidden State From Input

Proposed Hidden State From Prev Hidden State

New Hidden State 1 x 2

4 Compute Output

$$\begin{bmatrix} 0 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & -2 \end{bmatrix} = \begin{bmatrix} -2 \end{bmatrix}$$

New Hidden State 1 x 2

HY Weights 2 x 1 (H x Y)

5 Compute Output

$$\mathrm{sigmoid}\left(\begin{bmatrix} -2 \end{bmatrix}\right) = \begin{bmatrix} 0.12 \end{bmatrix}$$

Our predicted sentiment at sequence step 3

Sequence Step 3/4

(S) Sequence Length = 4

(E) Embedding Size = 3

(H) Hidden State Dimensions = 2

(Y) Output Dimension = 1

Epoch 1

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

Input Token
“EMPTY_TOKEN”
1 x 3
(1 token x E)

EH Weights
3 x 2
(E x H)

Proposed Hidden State From Input
1 x 2

Sequence Step 4/4

(S) Sequence Length = 4

(E) Embedding Size = 3

(H) Hidden State Dimensions = 2

(Y) Output Dimension = 1

Epoch 1

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} 0 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} -1 & -2 \end{bmatrix}$$

Previous Hidden State 1 x 2

HH Weights 2 x 2 (H x H)

Proposed Hidden State From Prev Hidden State 1 x 2

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

Input Token "EMPTY_TOKEN" 1 x 3 (1 token x E)

EH Weights 3 x 2 (E x H)

Proposed Hidden State From Input 1 x 2

Sequence Step 4/4

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

Epoch 1

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} 0 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} -1 & -2 \end{bmatrix}$$

Previous Hidden State 1 x 2

HH Weights 2 x 2 (H x H)

Proposed Hidden State From Prev Hidden State 1 x 2

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

Input Token “EMPTY_TOKEN” 1 x 3 (1 token x E)

EH Weights 3 x 2 (E x H)

Proposed Hidden State From Input 1 x 2

3 Update Hidden State

$$\tanh\left(\begin{bmatrix} 0 & 0 \end{bmatrix} + \begin{bmatrix} -1 & -2 \end{bmatrix} \right) = \begin{bmatrix} -.76 & -.96 \end{bmatrix} \sim \begin{bmatrix} -1 & -1 \end{bmatrix}$$

Proposed Hidden State From Input

Proposed Hidden State From Prev Hidden State

New Hidden State 1 x 2

Sequence Step 4/4

Epoch 1

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} 0 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} -1 & -2 \end{bmatrix}$$

Previous Hidden State 1 x 2 — HH Weights 2 x 2 (H x H) — Proposed Hidden State From Prev Hidden State 1 x 2

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

Input Token “EMPTY_TOKEN” 1 x 3 (1 token x E) — EH Weights 3 x 2 (E x H) — Proposed Hidden State From Input 1 x 2

3 Update Hidden State

$$\tanh\left(\begin{bmatrix} 0 & 0 \end{bmatrix} + \begin{bmatrix} -1 & -2 \end{bmatrix} \right) = \begin{bmatrix} -.76 & -.96 \end{bmatrix} \sim \begin{bmatrix} -1 & -1 \end{bmatrix}$$

Proposed Hidden State From Input — Proposed Hidden State From Prev Hidden State — New Hidden State 1 x 2

4 Compute Output

$$\begin{bmatrix} -1 & -1 \end{bmatrix} \times \begin{bmatrix} 2 & -2 \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

New Hidden State 1 x 2 — HY Weights 2 x 1 (H x Y)

Sequence Step 4/4

Epoch 1

(S) Sequence Length = 4
(E) Embedding Size = 3
(H) Hidden State Dimensions = 2
(Y) Output Dimension = 1

1 Compute Hidden State Contribution From Input

$$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

Input Token "EMPTY_TOKEN" 1 x 3 (1 token x E)

EH Weights 3 x 2 (E x H)

Proposed Hidden State From Input 1 x 2

2 Compute Hidden State Contribution From Previous Hidden State

$$\begin{bmatrix} 0 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} -1 & -2 \end{bmatrix}$$

Previous Hidden State 1 x 2

HH Weights 2 x 2 (H x H)

Proposed Hidden State From Prev Hidden State 1 x 2

3 Update Hidden State

$$\tanh\left(\begin{bmatrix} 0 & 0 \end{bmatrix} + \begin{bmatrix} -1 & -2 \end{bmatrix}\right) = \begin{bmatrix} -.76 & -.96 \end{bmatrix} \sim \begin{bmatrix} -1 & -1 \end{bmatrix}$$

Proposed Hidden State From Input

Proposed Hidden State From Prev Hidden State

New Hidden State 1 x 2

4 Compute Output

$$\begin{bmatrix} -1 & -1 \end{bmatrix} \times \begin{bmatrix} 2 & -2 \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

New Hidden State 1 x 2

HY Weights 2 x 1 (H x Y)

5 Compute Output

$$\text{sigmoid}\left(\begin{bmatrix} 0 \end{bmatrix}\right) = \begin{bmatrix} 0.5 \end{bmatrix}$$

Our predicted sentiment at sequence step 4

Sequence Step 4/4

Epoch 1

predicted y = 0.5
true y = 1

Binary Cross-Entropy (Log Loss) Function

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Loss is -log(0.5) → 0.3

Epoch 1

predicted y = 0.5
true y = 1

Binary Cross-Entropy (Log Loss) Function

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Loss is -log(0.5) → 0.3

If this was the only sample in our training dataset, we would be done with 1 epoch. We then use our log loss function to calculate the partial derivatives needed to update our weights

3	1
-1	-2
0	1

EH Weights
3 x 2
(E x H)

3	1
-1	-2

HH Weights
2 x 2
(H x H)

2	-2

HY Weights
2 x 1
(H x Y)

Epoch 2

predicted y = 0.78
true y = 1

Binary Cross-Entropy (Log Loss) Function

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Loss is -log(0.78) → 0.108

Pretend we update our weights via backpropagation, and make a new prediction for "I love cats" of 0.78.

2	2
1	-1
-2	2

EH Weights
3 x 2
(E x H)

-1	2
-0	2

HH Weights
2 x 2
(H x H)

1	-3

HY Weights
2 x 1
(H x Y)

Epoch 3

predicted y	= 0.98
true y	= 1

Binary Cross-Entropy (Log Loss) Function

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Loss is -log(0.98) → 0.008

Pretend we update our weights via backpropagation, and make a new prediction for "I love cats" of 0.98.

3	2
2	-1
-2	2

EH Weights
3 x 2
(E x H)

-2	2
-0	2

HH Weights
2 x 2
(H x H)

1	-4

HY Weights
2 x 1
(H x Y)

Backpropagation Through Time

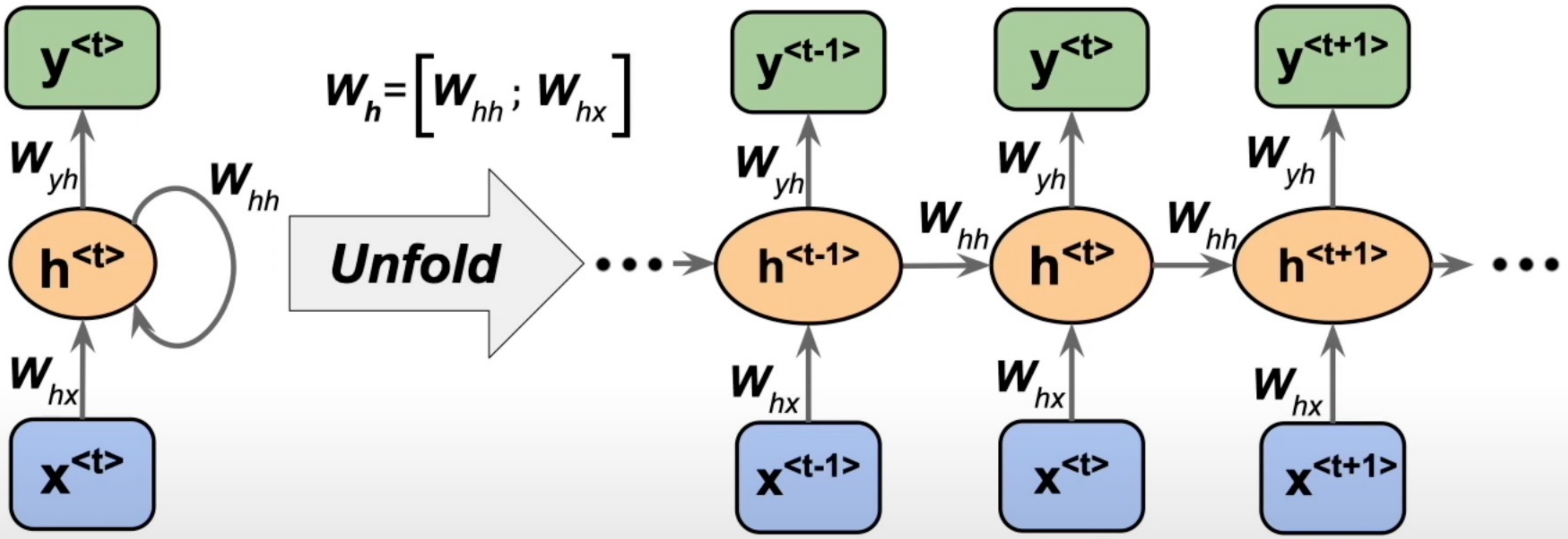

Image source: Sebastian Raschka, Vahid Mirjalili. *Python Machine Learning. 3rd Edition.* Packt, 2019

Backpropagation Through Time

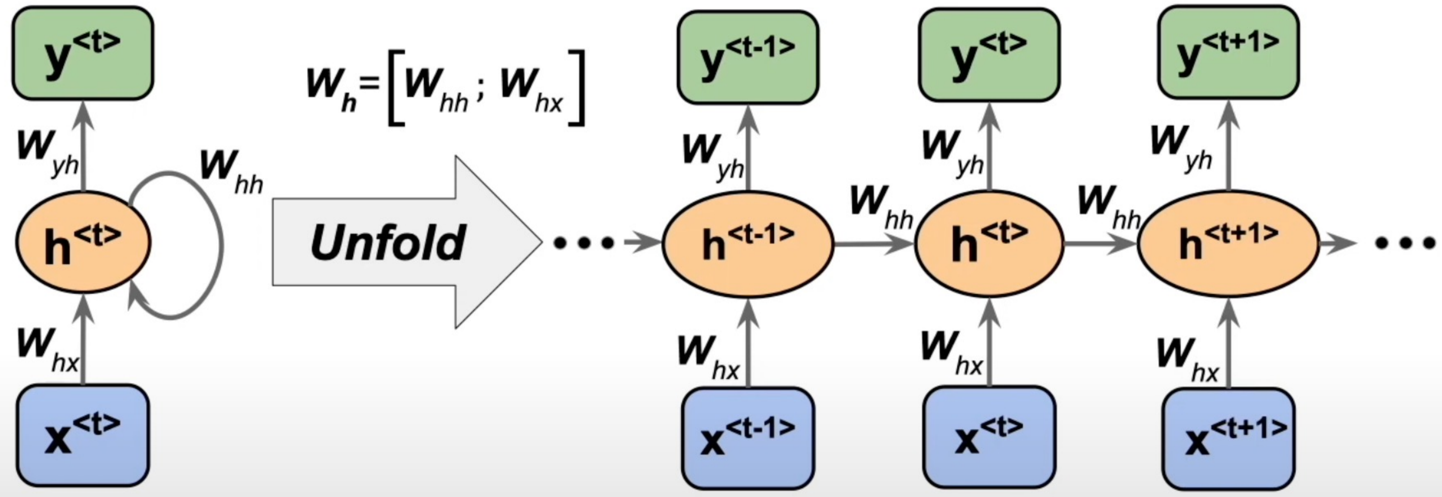

Image source: Sebastian Raschka, Vahid Mirjalili. *Python Machine Learning. 3rd Edition.* Packt, 2019

$$L = \sum_{t=1}^{T} L^{(t)} \qquad \frac{\partial L^{(t)}}{\partial \mathbf{W}_{hh}} = \frac{\partial L^{(t)}}{\partial y^{(t)}} \cdot \frac{\partial y^{(t)}}{\partial \mathbf{h}^{(t)}} \cdot \left(\sum_{k=1}^{t} \boxed{\frac{\partial \mathbf{h}^{(t)}}{\partial \mathbf{h}^{(k)}}} \cdot \frac{\partial \mathbf{h}^{(k)}}{\partial \mathbf{W}_{hh}} \right)$$

computed as a multiplication of adjacent time steps:

This is very problematic:
Vanishing/Exploding gradient problem!

$$\frac{\partial \mathbf{h}^{(t)}}{\partial \mathbf{h}^{(k)}} = \prod_{i=k+1}^{t} \frac{\partial \mathbf{h}^{(i)}}{\partial \mathbf{h}^{(i-1)}}$$

Sebastian, Raschka, [Backpropagation Through Time Overview]()

Backpropagation Through Time

$$h_t = \sigma(wh_{t-1}).$$

Hidden state at sequence step t

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Definition of the sigmoid activation function

$$\sigma'(x) = \frac{d}{dx}\sigma(x) = \sigma(x)(1-\sigma(x))$$

Derivative of the sigmoid activation function

$$\frac{\partial h_{t'}}{\partial h_t} = \prod_{k=1}^{t'-t} w\sigma'(wh_{t'-k})$$

$$= \underbrace{w^{t'-t}}_{!!!} \prod_{k=1}^{t'-t} \sigma'(wh_{t'-k})$$

Derivative of the hidden state

https://stats.stackexchange.com/questions/140537/why-do-rnns-have-a-tendency-to-suffer-from-vanishing-exploding-gradient