DATA ANALYSIS AND VISUALIZATION

INSTRUCTOR: UMME AMMARAH

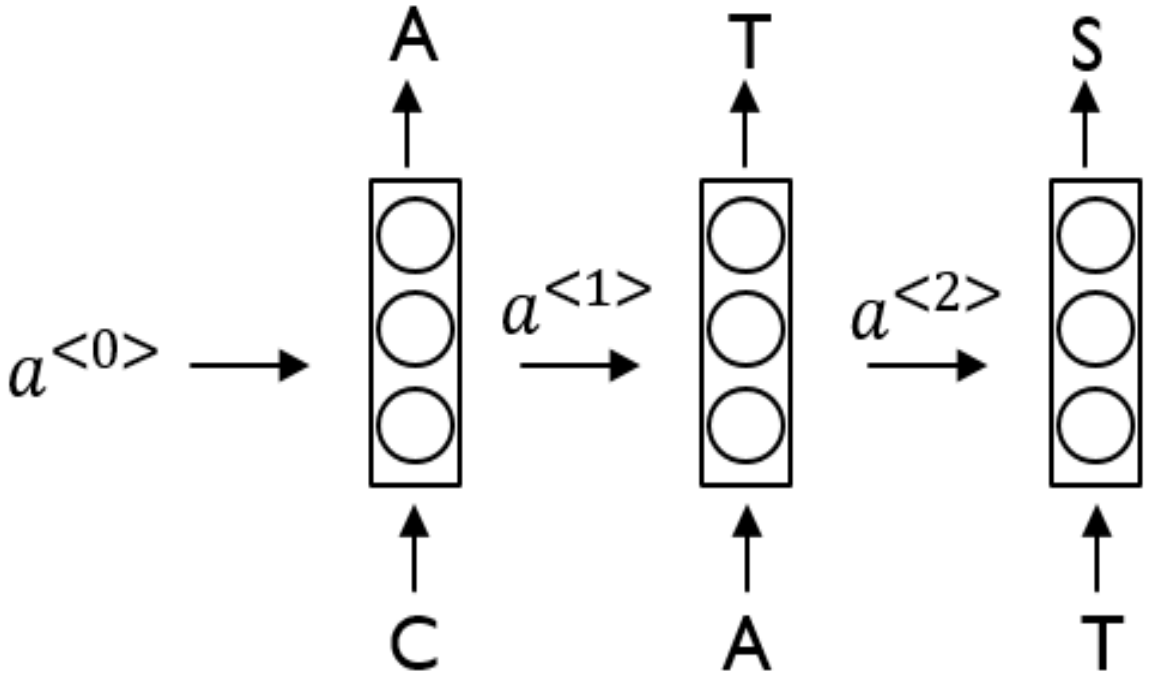
QUIZ 2

DEEP LEARNING

RECURRENT NEURAL NETWORK

DATA

- Input x = ‘C’-‘A’-‘T’
- Vocabulary = $\begin{bmatrix} A \\ C \\ S \\ T \end{bmatrix}$
- Predict letter ‘S’ as last output



WEIGHTS

- $W^0 = \begin{bmatrix} 0.6 & 0.8 & 0.4 & 0.8 \\ 0.2 & 0.2 & 0.8 & 0.7 \\ 0.9 & 0.8 & 0.1 & 0.2 \end{bmatrix}$ $W^{11} = \begin{bmatrix} 0.1 & 0.5 & 0.1 \\ 0.5 & 0.9 & 0.3 \\ 0.3 & 0.2 & 0.1 \end{bmatrix}$ $W^1 = \begin{bmatrix} 0.9 & 0.8 & 0.3 \\ 0.2 & 0.3 & 0.4 \\ 0.6 & 0.9 & 0.1 \\ 0.5 & 0.0 & 0.3 \end{bmatrix}$

- $b^0 = [0.1 \quad 0.3 \quad 0.2]$ $b^1 = [0.1 \quad 0.2 \quad 0.3 \quad 0.4]$

- $h^0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

STEP 1: INPUT EMBEDDING

- Input x = 'C'-'A'-'T'
- Vocabulary = $\begin{bmatrix} A \\ C \\ S \\ T \end{bmatrix}$
- X = $x^{<1>}\ x^{<2>} x^{<3>}$
- $x^{<1>} = C,\ x^{<2>} = A,\ x^{<3>} = T$

- $x^{<1>} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$
- $x^{<2>} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$
- $x^{<3>} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

STEP 2: COMPUTE HIDDEN LAYER ACTIVATION

- Take $x^{<1>}$ vector. Pass it to hidden layer, i.e. multiply with weight matrix w^0

$$\mathbf{h}^{(t)} = \tanh(W^0\mathbf{x}^{(t)} + W^{11}\mathbf{h}^{(t-1)} + \mathbf{b}_0)$$

$$\tanh\left(\begin{bmatrix}0.6 & 0.8 & 0.4 & 0.8\\ 0.2 & 0.2 & 0.8 & 0.7\\ 0.9 & 0.8 & 0.1 & 0.2\end{bmatrix}\begin{bmatrix}0\\1\\0\\0\end{bmatrix} + \begin{bmatrix}0.1 & 0.5 & 0.1\\ 0.5 & 0.9 & 0.3\\ 0.3 & 0.2 & 0.1\end{bmatrix}\begin{bmatrix}0\\0\\0\end{bmatrix} + \begin{bmatrix}0.1\\0.3\\0.2\end{bmatrix}\right)$$

CONT

$$\tanh\left(\begin{bmatrix}0.6 & 0.8 & 0.4 & 0.8\\ 0.2 & 0.2 & 0.8 & 0.7\\ 0.9 & 0.8 & 0.1 & 0.2\end{bmatrix}\begin{bmatrix}0\\1\\0\\0\end{bmatrix} + \begin{bmatrix}0.1 & 0.5 & 0.1\\ 0.5 & 0.9 & 0.3\\ 0.3 & 0.2 & 0.1\end{bmatrix}\begin{bmatrix}0\\0\\0\end{bmatrix} + \begin{bmatrix}0.1\\0.3\\0.2\end{bmatrix}\right)$$

$$\tanh\left(\begin{bmatrix}0.8\\0.2\\0.8\end{bmatrix} + \begin{bmatrix}0\\0\\0\end{bmatrix} + \begin{bmatrix}0.1\\0.3\\0.2\end{bmatrix}\right)$$

$$\tanh\left(\begin{bmatrix}0.9\\0.5\\1\end{bmatrix}\right) = \begin{bmatrix}0.71\\0.46\\0.76\end{bmatrix} = h^1$$

STEP 3: COMPUTE Y FOR TIME 1

- $y^t = softmax(W^1 h^t + b^1)$

$$softmax(\begin{bmatrix} 0.9 & 0.8 & 0.3 \\ 0.2 & 0.3 & 0.4 \\ 0.6 & 0.9 & 0.1 \\ 0.5 & 0.0 & 0.3 \end{bmatrix} \begin{bmatrix} 0.71 \\ 0.46 \\ 0.76 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{bmatrix})$$

$$softmax(\begin{bmatrix} 1.23 \\ 0.58 \\ 0.91 \\ 0.58 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{bmatrix}) = softmax(\begin{bmatrix} 1.33 \\ 0.78 \\ 1.21 \\ 0.98 \end{bmatrix})$$

STEP 4: APPLY SOFTMAX AND PREDICT NEXT OUTPUT CHARACTER

$$softmax(\begin{bmatrix}1.33\\0.78\\1.21\\0.98\end{bmatrix})$$

$$S(y)_i = \frac{\exp(y_i)}{\sum_{j=1}^{n} \exp(y_j)}$$

$$\begin{bmatrix}\frac{e^{1.33}}{11.97}\\ \frac{e^{0.78}}{11.97}\\ \frac{e^{1.21}}{11.97}\\ \frac{e^{0.98}}{11.97}\end{bmatrix}$$

Denominator = exp(1.33) + exp(0.78) + exp(1.21) + exp(0.98)

= 3.78 + 2.18 + 3.35 + 2.66

= 11.97

$$= \begin{bmatrix}0.31\\0.18\\0.27\\0.22\end{bmatrix}$$

Index zero has the highest probability which corresponds to letter A, so output at time 1 is A

STEP 5: REPEAT FROM STEP 2 TO 4 FOR EACH INPUT IN TIME.

- $h^2 = \begin{bmatrix} 0.79 \\ 0.90 \\ 0.90 \end{bmatrix}$
- $h^3 = \begin{bmatrix} 0.90 \\ 0.98 \\ 0.72 \end{bmatrix}$
- $y^2 = \begin{bmatrix} 0.36 \\ 0.16 \\ 0.31 \\ 0.17 \end{bmatrix}$ = A
- $y^3 = \begin{bmatrix} 0.37 \\ 0.14 \\ 0.33 \\ 0.16 \end{bmatrix}$ = A