

22-51-Q2

(a)(i) Q $h_1 = ?$

Solution

$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t)$$

$$y_t = W_{hy} h_t$$

$$h_1 = \tanh(W_{hh} h_0 + W_{xh} x_1)$$

$$= \tanh\left(\begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix} \begin{bmatrix} 0 \\ 6 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 0.1 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix}\right)$$

$$= \tanh\left(\begin{bmatrix} 2.3 \\ 1 \end{bmatrix}\right)$$

$$= \begin{bmatrix} 0.9801 \\ 0.7616 \end{bmatrix}$$

$$1.5 + 0.8 = 2.3$$

$$0.6 + 0.4 = 1$$

(ii) $y_1 = ?$

$$\begin{aligned} \text{Solution } y_1 &= W_{hy} h_1 = \begin{bmatrix} 0.1 & 0.4 \end{bmatrix} \begin{bmatrix} 0.9801 \\ 0.7616 \end{bmatrix} \\ &= \underline{0.40265} \quad \text{老师 } 0.4026 \end{aligned}$$

(iii) $y_2 = ?$

Solution

$$y_2 = W_{hy} h_2$$

$$h_2 = \tanh(W_{hh} h_1 + W_{xh} x_2)$$

$$= \tanh\left(\begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix} \begin{bmatrix} 0.9801 \\ 0.7616 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 0.1 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \end{bmatrix}\right)$$

$$= \tanh \left(\begin{bmatrix} 1.95033 \\ 1.39867 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 0.9603 \\ 0.8851 \end{bmatrix}$$

$$y_2 = [0.1 \ 0.4] \begin{bmatrix} 0.9603 \\ 0.8851 \end{bmatrix} = 0.4501$$

(iv) ① Adjust input Dimensions: a feature vector of size 3

② Sequence length adjust to 5

③ many to one configuration
remove unnecessary outputs

④ single output neuron

⑤ linear activation function

(ii) Vision Transformer (ViT)

① ViT is based on Transformer architecture that uses attention mechanism and can achieve very good accuracy

② VGG is a CNN that does not leverage on global attention. It uses convolutional layers to progressively extract higher level abstraction features

③ LSTM is a model that uses memory to analyse sequential data.

Hence it is not suitable for image classification application