RNN and Backpropagation

TOTAL POINTS 4

1. Consider a task with input sequences of fixed length. Could RNN architecture still be useful for such task? 1 point

● Yes

○ No

---

2. Consider an RNN for a language generation task. $\hat{y}_t$ is an output of this RNN at each time step, L is a length of the input sequence, N is a number of words in the vocabulary. Choose correct statements about $\hat{y}_t$: 1 point

● $\hat{y}_t$ is a vector of length N.

☐ $\hat{y}_t$ is a vector of length $(L - t)$.

☐ $\hat{y}_t$ is a vector of length $L \times N$.

☐ Each element of $\hat{y}_t$ is either 0 or 1.

● Each element of $\hat{y}_t$ is a number from 0 to 1.

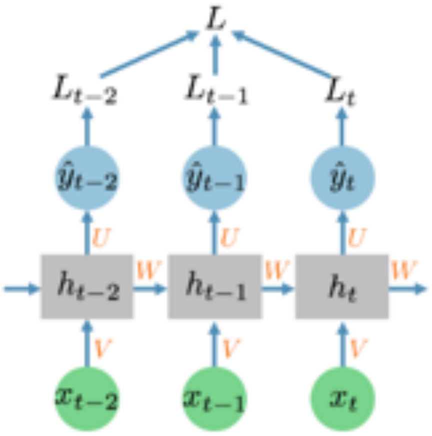
☐ Each element of $\hat{y}_t$ is a number from 0 to N.

---

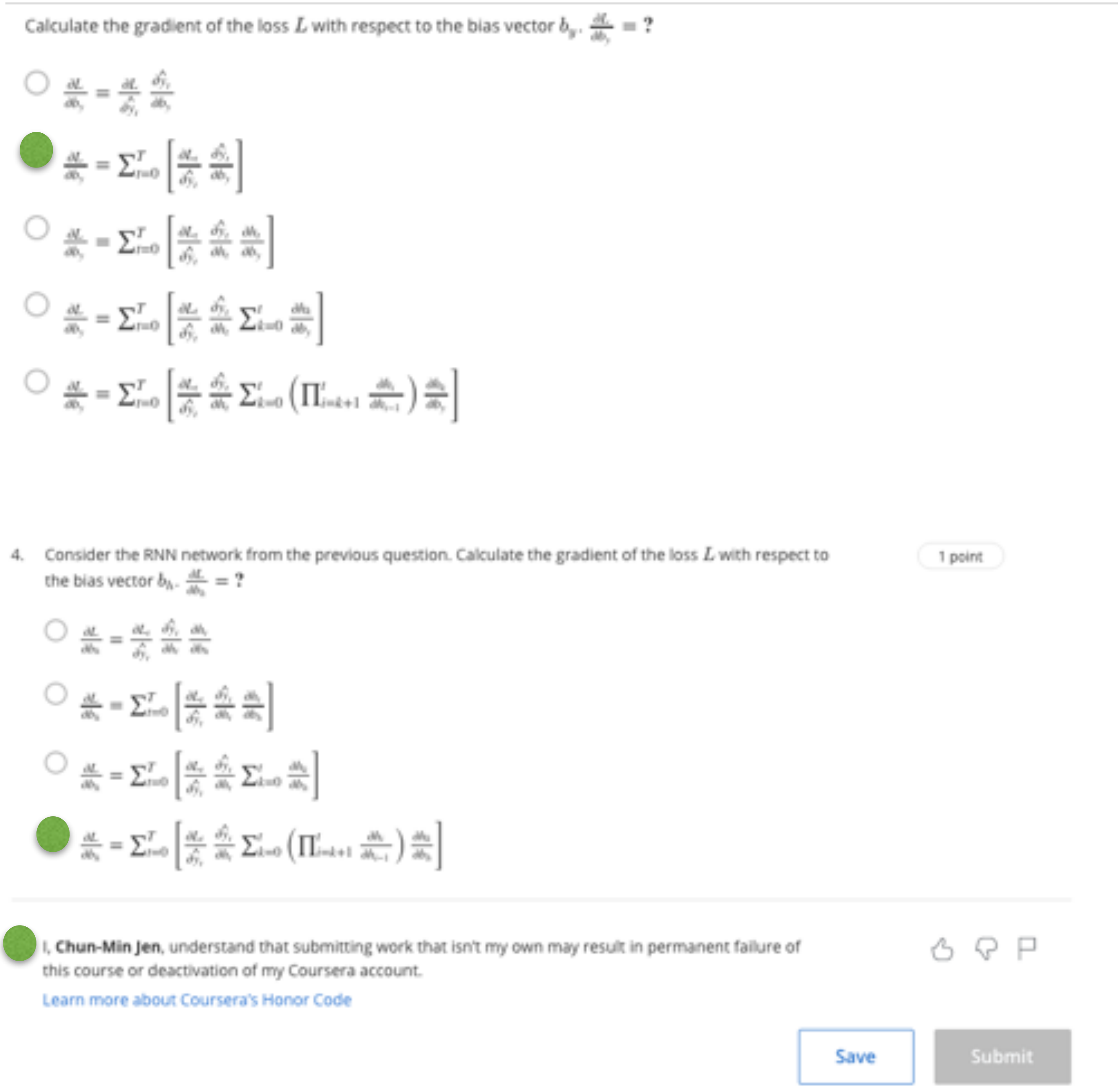
3. Consider the RNN from the lecture: 1 point

$$h_t = f_t(Vx_t + Wh_{t-1} + b_h)$$

$$\hat{y}_t = f_y(Uh_t + b_y)$$



Calculate the gradient of the loss L with respect to the bias vector b_y. $\frac{\partial L}{\partial b_y} = ?$



Calculate the gradient of the loss L with respect to the bias vector b_y. $\frac{\partial L}{\partial b_y} = ?$

○ $\frac{\partial L}{\partial b_y} = \frac{\partial L}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial b_y}$

● $\frac{\partial L}{\partial b_y} = \sum_{t=0}^{T} \left[\frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial b_y} \right]$

○ $\frac{\partial L}{\partial b_y} = \sum_{t=0}^{T} \left[\frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial h_t} \frac{\partial h_t}{\partial b_y} \right]$

○ $\frac{\partial L}{\partial b_y} = \sum_{t=0}^{T} \left[\frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial h_t} \sum_{k=0}^{t} \frac{\partial h_k}{\partial b_y} \right]$

○ $\frac{\partial L}{\partial b_y} = \sum_{t=0}^{T} \left[\frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial h_t} \sum_{k=0}^{t} \left(\prod_{i=k+1}^{t} \frac{\partial h_i}{\partial h_{i-1}} \right) \frac{\partial h_k}{\partial b_y} \right]$

4. Consider the RNN network from the previous question. Calculate the gradient of the loss L with respect to the bias vector b_h. $\frac{\partial L}{\partial b_h} = ?$

1 point

○ $\frac{\partial L}{\partial b_h} = \frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial h_t} \frac{\partial h_t}{\partial b_h}$

○ $\frac{\partial L}{\partial b_h} = \sum_{t=0}^{T} \left[\frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial h_t} \frac{\partial h_t}{\partial b_h} \right]$

○ $\frac{\partial L}{\partial b_h} = \sum_{t=0}^{T} \left[\frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial h_t} \sum_{k=0}^{t} \frac{\partial h_k}{\partial b_h} \right]$

● $\frac{\partial L}{\partial b_h} = \sum_{t=0}^{T} \left[\frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial h_t} \sum_{k=0}^{t} \left(\prod_{i=k+1}^{t} \frac{\partial h_i}{\partial h_{i-1}} \right) \frac{\partial h_k}{\partial b_h} \right]$

● I, **Chun-Min Jen**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Learn more about Coursera's Honor Code