RNN and Backpropagation
TOTAL POINTS 4

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

(**X**) Yes

No

1 point

2.Question 2
Consider an RNN for a language generation task. y^t is an output of this RNN at each time step, LL is a length of the input sequence, NN is a number of words in the vocabulary. Choose correct statements about y^t:

(**X**) y^t is a vector of length N.

() y^t is a vector of length (L-t).

() y^t is a vector of length L×N.

() Each element of y^t is either 0 or 1.

(**X**) Each element of y^t is a number from 0 to 1.

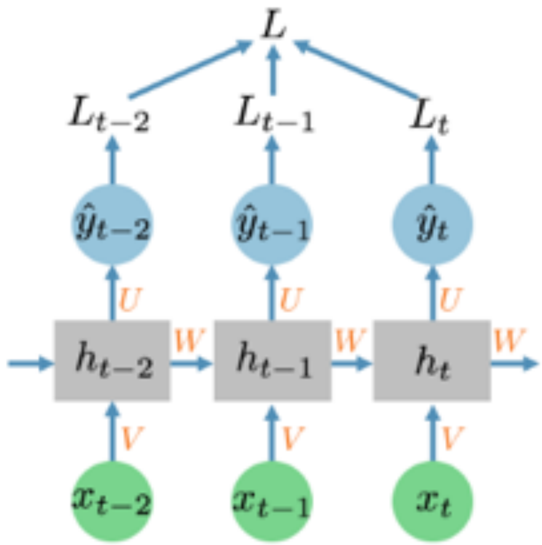
() Each element of y^t is a number from 0 to N.

1 point

3. Consider the RNN from the lecture:

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

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



() ∂L∂by=∂L∂y^t∂y^t∂by

(**X**) ∂L∂by=∑Tt=0[∂Lt∂y^t∂y^t∂by]

() ∂L∂by=∑Tt=0[∂Lt∂y^t∂y^t∂ht∂ht∂by]

() ∂L∂by=∑Tt=0[∂Lt∂y^t∂y^t∂ht∑tk=0∂hk∂by]

() ∂L∂by=∑Tt=0[∂Lt∂y^t∂y^t∂ht∑tk=0(∏ti=k+1∂hi∂hi−1)∂hk∂by]

1 point

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

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

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

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

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

○ $\frac{\partial L}{\partial b_y} = \sum_{t=0}^{T} \left[\frac{\partial L_t}{\partial \hat{y}_t} \frac{d\hat{y}_t}{dh_t} \sum_{k=0}^{t} \left(\prod_{i=k+1}^{t} \frac{dh_i}{dh_{i-1}} \right) \frac{dh_k}{db_y} \right]$

4.Question 4

() ∂L∂bh=∂Lt∂y^t∂y^t∂ht∂ht∂bh

() ∂L∂bh=∑Tt=0[∂Lt∂y^t∂y^t∂ht∂ht∂bh]

() ∂L∂bh=∑Tt=0[∂Lt∂y^t∂y^t∂ht∑tk=0∂hk∂bh]

(**X**) ∂L∂bh=∑Tt=0[∂Lt∂y^t∂y^t∂ht∑tk=0(∏ti=k+1∂hi∂hi−1)∂hk∂bh]

1 point

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

Save

Submit