RNN and Backpropagation

测验, 4 个问题

4/4 分 (100%)

恭喜！您通过了！

下一项

1 / 1 分

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

- (•) Yes

正确
RNN could be useful since it may need much less number of parameters.

- ( ) No

---

1 / 1 分

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 $\hat{y}_t$:

- [x] $\hat{y}_t$ is a vector of length N .

正确
The output at each time step is a distribution over a vocabulary, therefore the length of $\hat{y}_t$ is equal to the vocabulary size.

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

未选择的是正确的

RNN and Backpropagation

4/4 分 (100%)

测验, 4 个问题

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

未选择的是正确的

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

未选择的是正确的

- [x] Each element of $\hat{y}_t$ is a number from 0 to 1.

正确

Elements of $\hat{y}_t$ are probabilities so they are numbers from 0 to 1 and the sum of them equal to 1.

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

未选择的是正确的

---

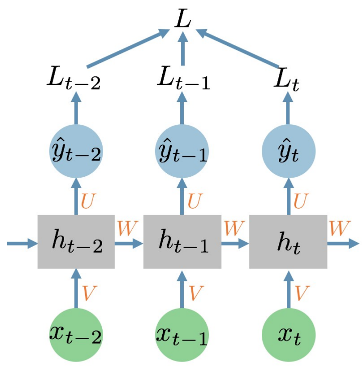
1 / 1 分

3。

Consider the RNN from the lecture:

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

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



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

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

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

正确

It is correct since b_y influence each L_t only once trough $\hat{y}_t$.

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

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

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

---

1 / 1 分

4。

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

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

正确

It is correct. Hidden units depend on b_h at each time step, therefore we need to backpropagate through time here.