

Homework 3

Part 2

1)

$$\begin{aligned}\frac{\partial L}{\partial V} &= \sum_{t=1}^T \frac{\partial L^{(t)}}{\partial V} \\ &= \sum_{t=1}^T -\frac{1}{\hat{y}_i^t} * \frac{e^{O_i}(\sum e^{O_i} - e^{O_i})}{(\sum e^{O_i})^2} * y^t (h^t)^T \\ &= \sum_{t=1}^T ((\hat{y}^t)^T y^t - 1) * y^t (h^t)^T\end{aligned}$$

假设一共T次循环

$$\begin{aligned}\frac{\partial L}{\partial h^T} &= V^T (\hat{y}^T - y^T) \\ \frac{\partial L}{\partial h^t} &= V^T (\hat{y}^t - y^t) + W^T \text{diag}(1 - (h^{(t+1)})^2) \frac{\partial L}{\partial h^{(t+1)}} \\ \frac{\partial L}{\partial W} &= \sum_{t=1}^T \text{diag}(1 - (h^t)^2) * \frac{\partial L}{\partial h^t} * (h^t)^T \\ \frac{\partial L}{\partial U} &= \sum_{t=1}^T \text{diag}(1 - (h^t)^2) * \frac{\partial L}{\partial h^t} * (x^t)^T\end{aligned}$$

2)

$$\frac{\partial L^t}{\partial W} = \frac{\partial L^t}{\partial h^t} \frac{\partial h^t}{\partial W} + \frac{\partial L^t}{\partial h^t} \frac{\partial h^t}{\partial h^{(t-1)}} \frac{\partial h^{(t-1)}}{\partial W} + \dots + \frac{\partial L^t}{\partial h^t} \left(\prod_{i=1}^t \frac{\partial h^{i+1}}{\partial h^i} \right) \frac{\partial h^1}{\partial W}$$

由于函数求导结果特征，梯度连乘 $\prod_{i=1}^t \frac{\partial h^{i+1}}{\partial h^i}$ 会导致在sigmoid 和 tanh函数上出现梯度爆炸或梯度消失