Exam 3 review 2022 nov 18

Friday, November 18, 2022 12:03

$\boxed{5b}$

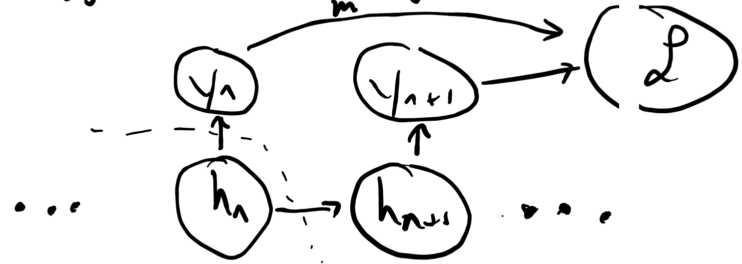
$$\begin{bmatrix} h_n = \tanh(U x_n + V h_{n-1}) = [h_{n,1}, \dots, h_{n,d}]^T \\ y_n = \mathrm{softmax}(W h_n) = [y_{n,1}, \dots, y_{n,c}]^T \end{bmatrix}$$

GIVEN $\frac{\partial \mathcal{L}}{\partial y_{n,k}}$ and $\frac{\partial \mathcal{L}}{\partial h_{n+1,j}}$, FIND $\frac{\partial \mathcal{L}}{\partial h_{n,i}}$

$$\Rightarrow h_{n,i} = \tanh\left(\sum_m U_{i,m} x_{n,m} + \sum_\ell V_{i,\ell} h_{n-1,\ell}\right)$$

$$y_{n,k} = \frac{\exp\left(\sum_i W_{k,i} h_{n,i}\right)}{\sum_{k'} \exp\left(\sum_i W_{k',i} h_{n,i}\right)}$$

$$h_{n+1,j} = \tanh\left(\sum_m U_{j,m} x_{n+1,m} + \sum_i V_{j,i} h_{n,i}\right)$$



$$\frac{\partial \mathcal{L}}{\partial h_n} = \frac{\partial \mathcal{L}}{\partial y_n}\frac{\partial y_n}{\partial h_n} + \frac{\partial \mathcal{L}}{\partial h_{n+1}}\frac{\partial h_{n+1}}{\partial h_n}$$

$$\frac{\partial \mathcal{L}}{} = \sum^{c} \frac{\partial \mathcal{L}}{} \frac{\partial y_{n,k}}{} + \sum^{d} \frac{\partial \mathcal{L}}{\partial} \frac{\partial h_{n+1,j}}{}$$

$\partial h_{n,i} \quad \sum_{k=1} \partial y_{n,k} \quad \partial h_{n,i} \qquad j=1 \ \partial h_{n+1,j} \quad \partial h_{n,i}$

$$\frac{\partial y_{n,k}}{\partial h_{n,i}} = \frac{\partial NUM / \partial h_{n,i}}{DEN} - \frac{\partial DEN}{\partial h_{n,i}} \frac{NUM}{DEN^2}$$

$$= \frac{\exp\left(\sum_{i'} W_{k,i'} h_{n,i'}\right) W_{k,i}}{\sum_{k'} \exp\left(\sum_{i'} W_{k',i'} h_{n,i'}\right)} - \frac{\exp\left(\sum_{i'} W_{k,i'} h_{n,i'}\right)}{\left(\sum_{k'} \exp\left(\sum_{i'} W_{k',i'} h_{n,i'}\right)\right)^2} \times \sum_{k'} \exp\left(\sum_{i'} W_{k',i'} h_{n,i'}\right) W_{k',i}$$

$$= y_{n,k} W_{k,i} - y_{n,k} \left(\frac{\sum_{l} \exp\left(\sum_m W_{l,m} h_{n,m}\right) W_{l,i}}{\sum_{l} \exp\left(\sum_m W_{l,m} h_{n,m}\right)} \right)$$

$$\boxed{\frac{\partial y_{n,k}}{\partial h_{n,i}} = y_{n,k} W_{k,i} - y_{n,k} \sum_{l} y_{n,l} W_{l,i}}$$

$$\frac{\partial h_{n+1,j}}{\partial h_{n,i}} = \tanh'(\sim) V_{j,i}$$

$$= \left(1 - \tanh^2(\sim)\right) V_{j,i}$$

$$= (1 - h_{n+1,j}^2) V_{j,i}$$

$$\boxed{\frac{\partial \mathcal{L}}{\partial h_{n,i}} = \sum_k \frac{\partial \mathcal{L}}{\partial y_{n,k}} \left(y_{n,k} W_{k,i} - y_{n,k} \sum_l y_{n,l} W_{l,i} \right) + \sum_j \frac{\partial \mathcal{L}}{\partial h_{n+1,j}} \left((1 - h_{n+1,j}^2) V_{j,i} \right)}$$

WHERE $W_{k,i}$ IS THE $(k,i)^{\text{TH}}$ ELEMENT OF W

AND $V_{j,i}$ IS THE $(j,i)^{\text{TH}}$ ELEMENT OF V

16) $Y = \text{softmax}(QK^T) V$

$Y_{i,j}$ etc. are the $(i,j)^{\text{TH}}$ ELEMENTS

$$\frac{\partial Y_{i,j}}{\partial K_{k,l}}$$

j^{TH} COLUMN OF K^T IS THE j^{TH} ROW OF K

$$\alpha = \text{softmax}(QK^T)$$

$$\alpha_{i,j} = \frac{\exp\left(\sum_m Q_{i,m} K_{j,m}\right)}{\sum \exp\left(\sum Q \; K \; \right)}$$

$$\sum_n \exp\Big(\sum_m Q_{i,m} K_{n,m}\Big)$$

$$Y_{i,j} = \sum_q \alpha_{i,\underline{q}} V_{q,j} \quad ; \text{ so } \quad \sum_q \alpha_{i,q} = 1$$

$$\frac{\partial Y_{i,j}}{\partial K_{k,l}} = \sum_q \frac{\partial \alpha_{i,q}}{\partial K_{k,l}} V_{q,j} \qquad [p] = \begin{cases} 1 & \text{IF } p \text{ TRUE} \\ 0 & \text{IF } p \text{ FALSE} \end{cases}$$

$$\frac{\partial \alpha_{i,q}}{\partial K_{k,l}} = \frac{\partial \mathrm{NUM}/\partial K_{k,l}}{\mathrm{DEN}} - \frac{\mathrm{NUM}}{\mathrm{DEN}^2} \frac{\partial \mathrm{DEN}}{\partial K_{k,l}}$$

$$= \frac{\exp\big(\sum_m Q_{i,m} K_{q,m}\big) Q_{i,l} [q = k]}{\sum_n \exp\big(\sum_m Q_{i,m} K_{n,m}\big)}$$

$$- \frac{\exp\big(\sum_m Q_{i,m} K_{q,m}\big)}{\Big(\sum_n \exp\big(\sum_m Q_{i,m} K_{n,m}\big)\Big)^2} \exp\Big(\sum_m Q_{i,m} K_{k,m}\Big) Q_{i,l}$$

$$\frac{\partial Y_{i,j}}{\partial K_{k,l}} = \sum_q V_{q,j} \left(\frac{\exp\big(\sum_m Q_{i,m} K_{q,m}\big)}{\sum_n \exp\big(\sum_m Q_{i,m} K_{n,m}\big)} Q_{i,l} [q = k] \right.$$

$$- \frac{\exp\left(\sum_m Q_{i,m} K_{q,m}\right) \exp\left(\sum_m Q_{i,m} K_{k,m}\right) Q_{i,l}}{\left(\sum_n \exp\left(\sum_m Q_{i,m} K_{n,m}\right)\right)^2} \Bigg)$$

$$\frac{\partial Y_{i,j}}{\partial K_{k,l}} = \sum_q V_{q,j} \Big(\alpha_{i,q} Q_{i,l} [q=k] - \alpha_{i,q} \alpha_{i,l} Q_{i,l} \Big)$$

$$\begin{cases} \sum_q V_{q,j} \alpha_{i,q} Q_{i,l} (1 - \alpha_{i,l}) & q=k \\ \\ -\sum_q V_{q,j} \alpha_{i,q} \alpha_{i,l} Q_{i,l} & q \neq k \end{cases}$$

$$\frac{\text{SOFTMAX}}{p(i|f)} = \frac{\exp(f_i)}{\sum_j \exp(f_j)}$$

$$\Rightarrow \frac{\partial(-\ln p(i|f))}{\partial f_k} = \begin{cases} p(i|f) - 1 & k=i \\ p(k|f) & k \neq i \end{cases}$$

$$\partial \ln p(i|f) \qquad \Big(1 - p(i|f) \qquad k=i$$

$$\frac{\qquad}{\partial f_k} = \begin{cases} \\ -p(k|f) & k \neq i \end{cases}$$

$$\frac{\partial p(i|f)}{\partial f_k} = \begin{cases} p(i|f)\,\big(1 - p(i|f)\big) & k = i \\ -p(i|f)\,p(k|f) & k \neq i \end{cases}$$