Homework4

于翠翠

2019 年 11 月 16 日星期六

Part1

（1）由于：$J^{(t)}(\theta)=CE(y^{(t)},\hat{y}^{(t)})=-\sum_{j=1}^{|V|}y_j^{(t)}\log\hat{y}_j^{(t)}=-\log\hat{y}_j^{(t)}$

$$PP^{(t)}(y^{(t)},\hat{y}^{(t)})=\frac{1}{\overline{P}(x_{pred}^{(t+1)}=x^{(t+1)}\mid x^{(t)},\cdots,x^{(1)})}=\frac{1}{\sum_{j=1}^{|V|}y_j^{(t)}\cdot\hat{y}_j^{(t)}}$$

$$\overline{P}(x_{pred}^{(t+1)}=x^{(t+1)}\mid x^{(t)},\cdots,x^{(1)})=\sum_{j=1}^{|V|}y_j^{(t)}\cdot\hat{y}_j^{(t)}=\hat{y}_j^{(t)}$$

且 $y_j^{(t)}$ 是 One-hot 向量，所以

$$PP^{(t)}(y^{(t)},\hat{y}^{(t)})=\frac{1}{\hat{y}_j^{(t)}}$$

因为：

$$\log(PP^{(t)}(y^{(t)},\hat{y}^{(t)}))=-\log(\hat{y}_j^{(t)})$$

$$J^{(t)}(\theta)=CE(y^{(t)},\hat{y}^{(t)})=-\sum_{j=1}^{|V|}y_j^{(t)}\log\hat{y}_j^{(t)}=-\log\hat{y}_j^{(t)}$$

所以它们的增减性相同，当 $J^{(t)}(\theta)$ 减小时， $PP^{(t)}(y^{(t)},\hat{y}^{(t)})$ 也变小。

因为：

$$E(PP^{(t)}(y^{(t)},\hat{y}^{(t)}))=E(\frac{1}{\hat{y}_j^{(t)}})=|V|$$

所以：

$$E(J^{(t)}(\theta))=\log(|V|)=\log(10000)=\lg(10000)=4$$

（2）由于：

$$e^{(t)} = x^{(t)} L$$
$$h^{(t)} = sigmoid(h^{(t-1)} H + e^{(t)} I + b_1)$$
$$o^{(t)} = h^{(t-1)} H + e^{(t)} I + b_1$$
$$\hat{y}^{(t)} = soft\max(h^{(t)} U + b_2)$$
$$z^{(t)} = h^{(t)} U + b_2$$

$$J^{(t)}(\theta) = CE(y^{(t)}, \hat{y}^{(t)}) = -\sum_{j=1}^{|V|} y_j^{(t)} \log \hat{y}_j^{(t)} == -\log \hat{y}_j^{(t)}$$

$$\frac{\partial \hat{y}^{(t)}}{\partial z^{(t)}} = \frac{(e^{z_t})' \cdot \sum_{t=1}^{|V|} e^{z_t} - e^{z_t} \cdot (\sum_{t=1}^{|V|} e^{z_t})'}{(\sum_{j=1}^{|V|} e^{z_j})^2} = \frac{e^{z_t} \cdot \sum_{t=1}^{|V|} e^{z_t} - e^{2z_t}}{(\sum_{j=1}^{|V|} e^{z_j})^2} = \hat{y}^{(t)}(1 - \hat{y}^{(t)})$$

$$\frac{\partial h^{(t)}}{\partial o^{(t)}} = h^{(t)}(1 - h^{(t)})$$

故：

$$\frac{\partial J^{(t)}}{\partial b_2} = \frac{\partial J^{(t)}}{\partial \hat{y}^{(t)}} \cdot \frac{\partial \hat{y}^{(t)}}{\partial z^{(t)}} \cdot \frac{\partial z^{(t)}}{\partial b_2} = -\frac{1}{\hat{y}^{(t)}} \hat{y}^{(t)}(y^{(t)} - \hat{y}^{(t)}) = \hat{y}^{(t)} - y^{(t)}$$

$$\frac{\partial J^{(t)}}{\partial L_x^{(t)}} = \frac{\partial J^{(t)}}{\partial \hat{y}^{(t)}} \cdot \frac{\partial \hat{y}^{(t)}}{\partial z^{(t)}} \cdot \frac{\partial z^{(t)}}{\partial h^{(t)}} \cdot \frac{\partial h^{(t)}}{\partial o^{(t)}} \cdot \frac{\partial o^{(t)}}{\partial e^{(t)}} \cdot \frac{\partial e^{(t)}}{\partial L_x^{(t)}}$$

$$= -\frac{1}{\hat{y}^{(t)}} \hat{y}^{(t)}(-\hat{y}^{(t)} + y^{(t)}) U^T h^{(t)}(1 - h^{(t)}) I^T$$
$$= (\hat{y}^{(t)} - y^{(t)}) U^T h^{(t)}(1 - h^{(t)}) I^T$$

$$\frac{\partial J^{(t)}}{\partial I} = \frac{\partial J^{(t)}}{\partial \hat{y}^{(t)}} \cdot \frac{\partial \hat{y}^{(t)}}{\partial z^{(t)}} \cdot \frac{\partial z^{(t)}}{\partial h^{(t)}} \cdot \frac{\partial h^{(t)}}{\partial o^{(t)}} \cdot \frac{\partial o^{(t)}}{\partial I}$$

$$= \frac{1}{\hat{y}^{(t)}} \hat{y}^{(t)}(\hat{y}^{(t)} - y^{(t)}) U^T h^{(t)}(1 - h^{(t)}) e^{(t)}$$
$$= (\hat{y}^{(t)} - y^{(t)}) U^T h^{(t)}(1 - h^{(t)}) e^{(t)}$$

$$\frac{\partial J^{(t)}}{\partial H} = \frac{\partial J^{(t)}}{\partial \hat{y}^{(t)}} \cdot \frac{\partial \hat{y}^{(t)}}{\partial z^{(t)}} \cdot \frac{\partial z^{(t)}}{\partial h^{(t)}} \cdot \frac{\partial h^{(t)}}{\partial o^{(t)}} \cdot \frac{\partial o^{(t)}}{\partial H}$$

$$= \frac{1}{\hat{y}^{(t)}} \hat{y}^{(t)}(\hat{y}^{(t)} - y^{(t)}) U^T h^{(t)}(1 - h^{(t)}) h^{(t-1)}$$
$$= (\hat{y}^{(t)} - y^{(t)}) U^T h^{(t)}(1 - h^{(t)}) h^{(t-1)}$$