hw4-pytorch

陈雨琳

a)

because $y^{(t)}$ is one-hot, only 1 dimension(suppose is dimension j) has value 1 while all the others are 0. suppose the corresponding dimension j of $\hat{y}^{(t)}$ has value m. then $PP = \sum_{j=1}^{|V|} y_j^{(t)} \hat{y}_j^{(t)} = m$, and $loss = -\log m$. therefore minimizing Loss is the same as minimizing Perplexity.

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

$J^{(t)}(\theta) = -\log \frac{1}{10000} = 4$

b)

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

$$\frac{\partial \hat{y}_j^{(t)}}{\partial h^{(t)}} = \hat{y}_j^{(t)}(\hat{y}^{(t)} - \hat{y}^{(t)})$$

$$\frac{\partial J^t}{\partial b_2} = \frac{\partial J}{\partial \hat{y}_j^{(t)}} * \frac{\partial \hat{y}_j^{(t)}}{\partial b_2} = \hat{y}^t - y^t$$

$$\frac{\partial J^t}{\partial L_{x^t}} = \frac{\partial J}{\partial \hat{y}_j^{(t)}} \frac{\partial \hat{y}_j^{(t)}}{\partial h^t} \frac{\partial h^t}{\partial e^t} \frac{\partial e^t}{\partial L_{x^t}} = (\hat{y}^t - y^t)Uh^t(1-h^t)I$$

$$\frac{\partial J^t}{\partial I} = frac\partial J\partial \hat{y}_j^{(t)} \frac{\partial \hat{y}_j^{(t)}}{\partial h^t} \frac{\partial h^t}{\partial I} = (\hat{y}^t - y^t)Uh^t(1-h^t)e^t$$

$$\frac{\partial J^t}{\partial H} = frac\partial J\partial \hat{y}_j^{(t)} \frac{\partial \hat{y}_j^{(t)}}{\partial h^t} \frac{\partial h^t}{\partial H} = (\hat{y}^t - y^t)Uh^t(1-h^t)h^{t-1}$$

$$\frac{\partial J^t}{\partial h^{t-1}} = frac\partial J\partial \hat{y}_j^{(t)} \frac{\partial \hat{y}_j^{(t)}}{\partial h^t} \frac{\partial h^t}{\partial h^{t-1}} = (\hat{y}^t - y^t)Uh^t(1-h^t)H$$