Homework 2

Zhengyun Zhao

October 31, 2019

Part 1

1.

$$-\sum_{w\in vocab} y_w \log(\hat{y}_w) = -\sum_{w\in vocab, w\neq o} y_w \log(\hat{y}_w) - y_o \log(\hat{y}_o) = -\log(\hat{y}_o)$$

2.

$$\begin{aligned}
\frac{\partial J}{\partial v_c} &= \frac{\partial J}{\partial \hat{y}_0}\frac{\partial \hat{y}_0}{\partial v_c} \\
&= -\frac{1}{\hat{y}_o}\frac{\exp(u_o^T v_c)\sum_{w\in Vocab}\exp(u_w^T v_c)u_o - \exp(u_o^T v_c)\sum_{w\in Vocab}\exp(u_w^T v_c)u_w}{\left(\sum_{w\in Vocab}\exp(u_w^T v_c)\right)^2} \\
&= \frac{1}{\hat{y}_o}\hat{y}_o(U\hat{\boldsymbol{y}} - u_o) \\
&= \boldsymbol{U}(\hat{\boldsymbol{y}} - \boldsymbol{y})
\end{aligned}$$

3.

$$\begin{aligned}
\frac{\partial J}{\partial u_o} &= \frac{\partial J}{\partial \hat{y}_0}\frac{\partial \hat{y}_0}{\partial u_o} \\
&= -\frac{1}{\hat{y}_o}\frac{\exp(u_o^T v_c)\sum_{w\in Vocab}\exp(u_w^T v_c)v_c - \left(\exp(u_o^T v_c)\right)^2 v_c}{\left(\sum_{w\in Vocab}\exp(u_w^T v_c)\right)^2} \\
&= -\frac{1}{\hat{y}_o}\hat{y}_o(1-\hat{y}_o)v_c \\
&= (\hat{y}_o - 1)v_c
\end{aligned}$$

$$\begin{aligned}
\frac{\partial J}{\partial u_i} &= \frac{\partial J}{\partial \hat{y}_0}\frac{\partial \hat{y}_0}{\partial u_i} \\
&= -\frac{1}{\hat{y}_o}\frac{-\exp(u_o^T v_c)\exp(u_i^T v_c)v_c}{\left(\sum_{w\in Vocab}\exp(u_w^T v_c)\right)^2} \\
&= \frac{1}{\hat{y}_o}\hat{y}_o\hat{y}_i v_c \\
&= \hat{y}_i v_c
\end{aligned}$$

$$\frac{\partial J}{\partial \boldsymbol{U}} = v_c(\hat{\boldsymbol{y}} - \boldsymbol{y})^T$$

4.

$$\begin{aligned}
\sigma'(\boldsymbol{x}) &= [\frac{\partial \sigma(\boldsymbol{x})}{x_1}, \frac{\partial \sigma(\boldsymbol{x})}{x_2}, \ldots, \frac{\partial \sigma(\boldsymbol{x})}{x_n}]^T \\
&= [\sigma'(x_1), \sigma'(x_2), \ldots, \sigma'(x_n)]^T \\
&= [\sigma(x_1)(1-\sigma(x_1)), \sigma(x_2)(1-\sigma(x_2)), \ldots \sigma(x_n)(1-\sigma(x_n))]^T \\
&= \sigma(\boldsymbol{x})(1-\sigma(\boldsymbol{x}))
\end{aligned}$$