CS224 - Assignment 2

O.L.

September 2019

1

1.a

$$-logP(O = o|C = c) \xrightarrow{\text{Bayes' theorem}} -logP(O = o \bigcap C = c)P(C = c) \xrightarrow{\text{OneHot Vetor}} -log(\hat{y}_0)\underbrace{y_0}_{1} = -log\hat{y}_0$$

1.b

According to the chain rule:

$$\frac{\partial J}{\partial v_c} = \frac{\partial J}{\partial \theta}\frac{\partial \theta}{\partial v_c}$$

$$\frac{\partial J}{\partial \theta} = (\hat{y} - y)$$

Derivative according the inner vector:

$$\frac{\partial J}{\partial v_c} = \frac{\partial J}{\partial \theta}\frac{\partial \theta}{\partial v_c} = (\hat{y} - y)\frac{\partial U v_c}{\partial v_c} = U(\hat{y} - y)$$

1.c

Derivative according the outer vector:

$$\frac{\partial J}{\partial U} = \frac{\partial J}{\partial \theta}\frac{\partial \theta}{\partial U} = (\hat{y} - y)\frac{\partial U v_c}{\partial U} = v_c(\hat{y} - y)$$

1.d

$$\sigma'(x) = \sigma(x)(1 - \sigma(x))$$

1.e

Notations:

$\theta - u_o^T v_c$.
$\phi - \sigma(\theta)$.
$\delta - -log(\phi)$.

Derivative according the inner vector:

$$\frac{\partial J_{neg-sample}}{\partial v_c} = \frac{\partial \delta}{\partial \phi}\frac{\partial \phi}{\partial \theta}\frac{\partial \theta}{\partial\ v_c} + \sum_{i=1}^{K} \frac{\partial \delta}{\partial \phi}\frac{\partial \phi}{\partial \theta}\frac{\partial - \theta_k}{\partial\ v_c} =$$

$$(-\frac{1}{\phi})(\sigma(\theta)(1-\sigma(\theta))))(u_o^T) + \sum_{i=1}^{K}(-\frac{1}{\phi})(\sigma(-\theta)(1-\sigma(-\theta))))(-u_k^T) =$$

$$(-\frac{1}{\sigma(\theta)})(\sigma(u_o^T v_c)(1-\sigma(u_o^T v_c))))(u_o^T)+\sum_{i=1}^{K}(-\frac{1}{\sigma(-\theta)})(\sigma(-u_k^T v_c)(1-\sigma(-u_k^T v_c))))(-u_k^T) =$$

$$(-\frac{1}{\sigma(u_o^T v_c)})(\sigma(u_o^T v_c)(1-\sigma(u_o^T v_c))))(u_o^T)+\sum_{i=1}^{K}(-\frac{1}{\sigma(-u_k^T v_c)})(\sigma(-u_k^T v_c)(1-\sigma(-u_k^T v_c))))(-u_k^T) =$$

$$u_o^T(\sigma(u_o^T v_c) - 1) + \sum_{i=1}^{K}(u_k^T(1-\sigma(-u_k^T v_c)))$$

Derivative according the true word:

$$\frac{\partial J_{neg-sample}}{\partial u_o} = \frac{\partial \delta}{\partial \phi}\frac{\partial \phi}{\partial \theta}\frac{\partial \theta}{\partial\ u_o} =$$

$$(-\frac{1}{\phi})(\sigma(\theta)(1-\sigma(\theta))))(v_c) =$$

$$(-\frac{1}{\sigma(\theta)})(\sigma(u_o^T v_c)(1-\sigma(u_o^T v_c))))(v_c) =$$

$$(-\frac{1}{\sigma(u_o^T v_c)})(\sigma(u_o^T v_c)(1-\sigma(u_o^T v_c))))(v_c) =$$

$$v_c(\sigma(u_o^T v_c)-1)$$

Derivative according the negative word:

$$\frac{\partial J_{neg-sample}}{\partial u_k} = \sum_{i=1}^{K}\frac{\partial \delta}{\partial \phi}\frac{\partial \phi}{\partial \theta}\frac{\partial -\theta}{\partial u_k} =$$

$$\sum_{i=1}^{K}(-\frac{1}{\phi})(\sigma(-\theta)(1-\sigma(-\theta))))(-v_c) =$$

$$\sum_{i=1}^{K}(-\frac{1}{\sigma(-\theta)})(\sigma(-u_k^T v_c)(1-\sigma(-u_k^T v_c))))(-v_c) =$$

$$\sum_{i=1}^{K}(-\frac{1}{\sigma(-u_k^T v_c)})(\sigma(-u_k^T v_c)(1-\sigma(-u_k^T v_c))))(-v_c) =$$

$$\sum_{i=1}^{K}(v_c(1-\sigma(-u_k^T v_c)))$$