Sentence Vector Model

Yangfeng Ji

August 15, 2014
Modified on August 15, 2014

$$P(w_n = w|s, w_{n,c}) = \prod_i P(d_i|q_i, s, w_{n,c}) \tag{1}$$

$$P(d_i = 1|q_i, s, w_{n,c}) = \sigma\left(\hat{r}_{n,c}^\top \mathbf{U} q_i + s^\top \mathbf{V} q_i + b_i\right) \tag{2}$$

where

$$\hat{r}_{n,c} = \frac{1}{2c} \sum_{k=-c,\ldots,c} w_{n+k} \tag{3}$$

Parameter set:

- $q_i, \forall i$
- b_i
- s
- $w_{n,c}$
- $\mathbf{U}$
- $\mathbf{V}$

1 Gradient

Negative Log-likelihood:

$$\begin{aligned} NLL &= -\log P(w_n|s, w_{n,c}) & (4) \\ &= -\sum_i \log P(d_i|q_i, s, w_{n,c}) & (5) \\ &= -\sum_i \log \sigma\left(\hat{r}_{n,c}^\top \mathbf{U} q_i + s^\top \mathbf{V} q_i + b_i\right) & (6) \end{aligned}$$

Given that $\frac{\partial \log \sigma(x)}{\partial x} = 1 - \sigma(x)$, we have

$$\frac{\partial NLL}{\partial q_i} = -\Big(1 - \sigma\left(\hat{r}_{n,c}^\top \mathbf{U} q_i + s^\top \mathbf{V} q_i + b_i\right)\Big)\left(\mathbf{U}^\top \hat{r}_{n,c} + \mathbf{V}^\top s\right) \tag{7}$$

$$\frac{\partial NLL}{\partial b_i} = -\Big(1 - \sigma\left(\hat{r}_{n,c}^\top \mathbf{U} q_i + s^\top \mathbf{V} q_i + b_i\right)\Big) \tag{8}$$

$$\frac{\partial NLL}{\partial s} = -\sum_i \Big(1 - \sigma\left(\hat{r}_{n,c}^\top \mathbf{U} q_i + s^\top \mathbf{V} q_i + b_i\right)\Big)\mathbf{V} q_i \tag{9}$$

$$\frac{\partial NLL}{\partial \mathbf{U}} = -\sum_i \Big(1 - \sigma\left(\hat{r}_{n,c}^\top \mathbf{U} q_i + s^\top \mathbf{V} q_i + b_i\right)\Big)\hat{r}_{n,c} q_i^\top \tag{10}$$

$$\frac{\partial NLL}{\partial \mathbf{V}} = -\sum_i \Big(1 - \sigma\left(\hat{r}_{n,c}^\top \mathbf{U} q_i + s^\top \mathbf{V} q_i + b_i\right)\Big) s q_i^\top \tag{11}$$

$$\frac{\partial NLL}{\partial w_j} = -\sum_i \Big(1 - \sigma\left(\hat{r}_{n,c}^\top \mathbf{U} q_i + s^\top \mathbf{V} q_i + b_i\right)\Big)\frac{1}{2c}\mathbf{U} q_i \tag{12}$$

1.1 Building components

- $\Big(1 - \sigma\left(\hat{r}_{n,c}^\top \mathbf{U} q_i + s^\top \mathbf{V} q_i + b_i\right)\Big), \forall i$
- $\mathbf{U} q_i$
- $\mathbf{V} q_i$
- $\mathbf{U}^\top \hat{r}_{n,c}$
- $\mathbf{V}^\top s$
- $\hat{r}_{n,c} q_i^\top$
- $s q_i^\top$