# 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:

$$\begin{aligned} NLL &= -\log P(w_n|s, w_{n,c}) && (4)\\ &= -\sum_i \log P(d_i|q_i, s, w_{n,c}) && (5)\\ & && (6) \end{aligned}$$

where

$$
\begin{aligned}
P(d_i = 1|q_i, s, w_{n,c}) &= \sigma\left(\hat{r}_{n,c}^\top \mathbf{U} q_i + s^\top \mathbf{V} q_i + b_i\right) && (7) \\
P(d_i = 0|q_i, s, w_{n,c}) &= 1 - \sigma\left(\hat{r}_{n,c}^\top \mathbf{U} q_i + s^\top \mathbf{V} q_i + b_i\right) && (8)
\end{aligned}
$$

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

$$\frac{\partial NLL}{\partial q_i} = -\partial \log P(d_i|q_i, s, w_{n,c}) \left(\mathbf{U}^\top \hat{r}_{n,c} + \mathbf{V}^\top s\right) \tag{9}$$

$$\frac{\partial NLL}{\partial b_i} = -\partial \log P(d_i|q_i, s, w_{n,c}) \tag{10}$$

$$\frac{\partial NLL}{\partial s} = -\sum_i \partial \log P(d_i|q_i, s, w_{n,c}) \mathbf{V} q_i \tag{11}$$

$$\frac{\partial NLL}{\partial \mathbf{U}} = -\sum_i \partial \log P(d_i|q_i, s, w_{n,c}) \hat{r}_{n,c} q_i^\top \tag{12}$$

$$\frac{\partial NLL}{\partial \mathbf{V}} = -\sum_i \partial \log P(d_i|q_i, s, w_{n,c}) s q_i^\top \tag{13}$$

$$\frac{\partial NLL}{\partial w_j} = -\sum_i \partial \log P(d_i|q_i, s, w_{n,c}) \frac{1}{2c} \mathbf{U} q_i \tag{14}$$

## 1.1 Building components

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