

3. Word2vec (1)

2019년 1월 14일 월요일 오전 1:22

(a) $\frac{\partial CE(y, \hat{y})}{\partial v_c}$?

$$\begin{aligned} \frac{\partial CE(y, \hat{y})}{\partial v_c} &= \frac{\partial CE(y, \text{softmax}(U^T v_c))}{\partial v_c} \\ &= \underbrace{\frac{\partial CE(y, \text{softmax}(U^T v_c))}{\partial U^T v_c}}_{(\hat{y} - y)} \times \underbrace{\frac{\partial U^T v_c}{\partial v_c}}_U \quad \text{Denominator layout} \\ &= \underline{U(\hat{y} - y)} \end{aligned}$$

* 어떤 scalar 함수의 편미분 결과는 변수의 형태를 따른다.

즉, $\frac{\partial CE(y, \hat{y})}{\partial v_c}$ 은 v_c 의 형태, $(D \times 1)$ vector이다.

U 는 $(D \times W)$, y 와 \hat{y} 은 $(W \times 1)$ 이므로 $U(\hat{y} - y)$ 는 $(D \times 1)$ vector가 된다.

(b) $\frac{\partial CE(y, \hat{y})}{\partial U}$?

$$\frac{\partial CE(y, \hat{y})}{\partial U} = \frac{\partial CE(y, \text{softmax}(U^T v_c))}{\partial U^T v_c} \times \frac{\partial U^T v_c}{\partial U}$$

$$= (\hat{y} - y) \times v_c$$

$$= \underline{v_c (\hat{y} - y)^T}$$

$$\begin{aligned} v_c &: (D \times 1) \\ y, \hat{y} &: (W \times 1) \\ U &: (D \times W) \end{aligned}$$