CS 224n – Assignment #2 : word2vec

$\hat{y}$	dimensions	$1 \times V$
y	"	$1 \times V$
U	"	$V \times d$
V	"	$V \times d$
u_o	"	$1 \times d$
v_c	"	$1 \times d$

d – # embeddings
V = # words in the vocabulary

(a) Because y is one hot vector, y_i is only 1 when the word is 'o' $\Rightarrow$ Cross-entropy $= -\log(\hat{y}_o)$

(b) $\frac{\partial J}{\partial v_c} = (y - \hat{y}) U$ (J – naive-softmax)

(c) $\frac{\partial J}{\partial U} = (y - \hat{y})^T \cdot v_c$ (J – naive softmax)

(d) $\frac{d\,\sigma(x)}{dx} = \sigma(x)(1 - \sigma(x))$

(e)
$$\frac{\partial J}{\partial v_c} = (\sigma(u_o^T v_c) - 1) u_o - \sum_{k=1}^{K} (\sigma(-u_k^T v_c) - 1) u_k$$
$$\frac{\partial J}{\partial u_o} = (\sigma(u_o^T v_c) - 1) v_c$$
$$\frac{\partial J}{\partial u_k} = (1 - \sigma(-u_k^T v_c)) v_c$$
(J – neg sample)

(f) Common, this is very straightforward