A

$y_w = 0$ for all w expect $w = o$

B

$A_i = u_i^T v_c$
$A = U^T v_c$
$P = softmax(A)$
$L = -y * log(P)$

Derivative of softmax and cross entropy
$dP_i/dA_i = \begin{cases} P_i(1-P_i) & \text{if } i = j, \\ -P_i(P_j) & \text{if } i \neq j, \end{cases}$

$dL/A_i = -\sum_k \frac{y_k}{P_k} * \frac{dP_k}{dA_k}$
$dL/A_i = -y_i(1-P_i) - \sum_{k \neq i} y_k P_i$
$dL/A_i = P_i - y_i$
$dL/A = P - y$
$dL/v_c = U(P-y)$

C

$dL/du_i = dL/dA_i * dA_i/du_i$
$dL/du_i = (P_i - y_i)v_C^T$

D

$dL/U = (dL/dA)^T v_c^T$
$dL/U = (P-y)v_c^T$

E

$d\sigma(x)/dx = \sigma(x)(1-\sigma(x))$

F

$A_i = u_i^T v_c$
$L = -log(\sigma(A_o)) + \sum_k -log(\sigma(-A_k))$

$dL/dA_i = -\frac{\sigma(A_i)(1-\sigma(A_i))}{\sigma(A_i)} = \sigma(A_i) - 1$

$dL/du_o = dL/dA_o * v_c^T = (\sigma(u_o^T v_c) - 1)v_c^T$
$dL/du_k = dL/d(-A_k) * v_c^T = (1 - \sigma(-u_k^T v_c))v_c^T$
$dL/v_c = dL/dA_o * u_o^T + \sum_k dL/d(-A_k) * u_k^T$
$dL/v_c = (\sigma(u_o^T v_c) - 1)u_o^T + \sum_k (1 - \sigma(-u_k^T v_c))u_k^T$

G

$count =$ number of times u_k appears in sampled window
$dL/du_k = dL/d(-A_k) * v_c^T * count = (1 - \sigma(-u_k^T v_c))v_c^T * count$

H

$$dL_{sg}(v_c, w_{t-m}, \dots w_{t+m}, U)/dU = \sum_{\substack{-m \le j \le m \\ j \neq 0}} dL(v_c, w_{t+k}, U)/dU$$

$$dL_{sg}(v_c, w_{t-m}, \dots w_{t+m}, U)/dv_c = \sum_{\substack{-m \le j \le m \\ j \neq 0}} dL(v_c, w_{t+k}, U)/dv_c$$

$$dL_{sg}(v_c, w_{t-m}, \dots w_{t+m}, U)/dv_{w, w \neq c} = 0$$
