The objective described in the negative sampling part is:-

$$\log \sigma(V'^{T}_{w_O} V_{w_I}) + \sum E_{w_i \sim P_n(w)} [\log \sigma(-V'^{T}_{w_i} V_{w_I})]$$

Here, w_O is a context word, w_I is the target word. and w_i's are random words chosen on the basic of the distribution $P_n(w)$, that gives:-

$P_n(w_i) = 1 - \sqrt{\frac{t}{f(w_i)}}$ $f(w_i)$ is freq. of w_i.

BASICALLY,
we take a real context word from the target word's neighborhood, and do gradient ascent for $\log(\sigma(V'^{T}_{w_O} V_{w_I})) = y$ (say).

$$\frac{\partial y}{\partial V_{w_O}} = \frac{1}{\sigma(V'^{T}_{w_O} V_{w_I})} \sigma(V'^{T}_{w_O} V_{w_I})(1 - \sigma(V'^{T}_{w_O} V_{w_I})) \cdot V_{w_I}$$

$$\frac{\partial y}{\partial V_{w_O}} = (1 - \sigma(V'^{T}_{w_O} V_{w_I})) \cdot V_{w_I}$$

So, the parameter update for V_{w_O} :

$$V_{w_O} = V_{w_O} + \alpha (1 - \sigma(V'^{T}_{w_O} V_{w_I})) \cdot V_{w_I}$$

Same for the context-vector derivative and update:-

$$\frac{\partial y}{\partial V_{w_I}} = (1 - \sigma(V'^{T}_{w_O} V_{w_I})) \cdot V'^{T}_{w_O}$$

$$V_{w_I} = V_{w_I} + \alpha (1 - \sigma(V'^{T}_{w_O} V_{w_I})) V'^{T}_{w_O}$$

α is learning rate.

Now, for the negative samples:-
The gradient steps will be in the opposite direction, because the expressions differ only by a -1.
For each 'fake', 'false' context word:-

$$V_{w_i} = V_{w_i} - \alpha\left(1 - \sigma\left(V_{w_t}^{T} V_{w_i}\right)\right) V_{w_i}^{T}$$

$$V_{w_t} = V_{w_t} - \alpha\left(1 - \sigma\left(V_{w_t}^{T} V_{w_i}\right)\right) V_{w_i}$$