3(c) $J = -\log\left(\sigma\left(\mathbf{u}_o^{\mathrm{T}}\mathbf{v}_c\right)\right) - \sum_{k=1}^{K}\log\left(\sigma\left(-\mathbf{u}_k^{\mathrm{T}}\mathbf{v}_c\right)\right)$

NOTE: From 2(a), $\sigma'(x) = \sigma(x)\left(1-\sigma(x)\right)$

$\frac{\partial J}{\partial \mathbf{v}_c} =$	$-\frac{1}{\sigma\left(\mathbf{u}_o^{\mathrm{T}}\mathbf{v}_c\right)}\sigma\left(\mathbf{u}_o^{\mathrm{T}}\mathbf{v}_c\right)\left(1-\sigma\left(\mathbf{u}_o^{\mathrm{T}}\mathbf{v}_c\right)\right)\mathbf{u}_o - \sum_{k=1}^{K}\frac{1}{\sigma\left(-\mathbf{u}_k^{\mathrm{T}}\mathbf{v}_c\right)}\sigma\left(-\mathbf{u}_k^{\mathrm{T}}\mathbf{v}_c\right)\left(1-\sigma\left(-\mathbf{u}_k^{\mathrm{T}}\mathbf{v}_c\right)\right)\left(-\mathbf{u}_k\right)$	(1)
=	$-\left(1-\sigma\left(\mathbf{u}_o^{\mathrm{T}}\mathbf{v}_c\right)\right)\mathbf{u}_o - \sum_{k=1}^{K}\left(1-\sigma\left(-\mathbf{u}_k^{\mathrm{T}}\mathbf{v}_c\right)\right)\left(-\mathbf{u}_k\right)$	(2)
=	$\left(\sigma\left(\mathbf{u}_o^{\mathrm{T}}\mathbf{v}_c\right)-1\right)\mathbf{u}_o - \sum_{k=1}^{K}\left(\sigma\left(-\mathbf{u}_k^{\mathrm{T}}\mathbf{v}_c\right)-1\right)\mathbf{u}_k$	(3)

$\frac{\partial J}{\partial \mathbf{u}_o} =$	$-\frac{1}{\sigma\left(\mathbf{u}_o^{\mathrm{T}}\mathbf{v}_c\right)}\sigma\left(\mathbf{u}_o^{\mathrm{T}}\mathbf{v}_c\right)\left(1-\sigma\left(\mathbf{u}_o^{\mathrm{T}}\mathbf{v}_c\right)\right)\mathbf{v}_c$	(4)
=	$-\left(1-\sigma\left(\mathbf{u}_o^{\mathrm{T}}\mathbf{v}_c\right)\right)\mathbf{v}_c$	(5)
=	$\left(\sigma\left(\mathbf{u}_o^{\mathrm{T}}\mathbf{v}_c\right)-1\right)\mathbf{v}_c$	(6)

$\frac{\partial J}{\partial \mathbf{u}_k} =$	$-\frac{1}{\sigma\left(-\mathbf{u}_k^{\mathrm{T}}\mathbf{v}_c\right)}\sigma\left(-\mathbf{u}_k^{\mathrm{T}}\mathbf{v}_c\right)\left(1-\sigma\left(-\mathbf{u}_k^{\mathrm{T}}\mathbf{v}_c\right)\right)\left(-\mathbf{v}_c\right)$	(7)
=	$-\left(1-\sigma\left(-\mathbf{u}_k^{\mathrm{T}}\mathbf{v}_c\right)\right)\left(-\mathbf{v}_c\right)$	(8)
=	$-\left(\sigma\left(-\mathbf{u}_k^{\mathrm{T}}\mathbf{v}_c\right)-1\right)\mathbf{v}_c$	(9)

This cost function seems to be faster than the softmax-CE loss function by a factor of $\sum_{w=1}^{W}\exp(\mathbf{u}_w^{\mathrm{T}}\mathbf{v}_c)$