2(c) $CE(\mathbf{y},\hat{\mathbf{y}}) = -\sum_i \left(y_i \log(\widehat{y_i}) \right)$

$\boldsymbol{\delta}_1 = \dfrac{\partial CE}{\partial \mathbf{z}_2} =$	$-\dfrac{\partial}{\partial \mathbf{z}_2} \sum_i \left(y_i \log(\widehat{y_i}) \right)$	(1)
=	$\hat{\mathbf{y}} - \mathbf{y}$ from 2(b), by substituting $\mathbf{z}_2$ for $\boldsymbol{\theta}$	(2)

$\boldsymbol{\delta}_2 = \dfrac{\partial CE}{\partial \mathbf{h}} =$	$\dfrac{\partial CE}{\partial \mathbf{z}_2} \dfrac{\partial \mathbf{z}_2}{\partial \mathbf{h}}$	(3)
=	$\boldsymbol{\delta}_1 \dfrac{\partial \left(\mathbf{h}\mathbf{W}_2 + \mathbf{b}_2 \right)}{\partial \mathbf{h}}$	(4)
=	$\boldsymbol{\delta}_1 \mathbf{W}_2^{\mathbf{T}}$	(5)

$\boldsymbol{\delta}_3 = \dfrac{\partial CE}{\partial \mathbf{z}_1} =$	$\dfrac{\partial CE}{\partial \mathbf{h}} \dfrac{\partial \mathbf{h}}{\partial \mathbf{z}_1}$	(6)
=	$\boldsymbol{\delta}_1 \dfrac{\partial \left(\mathrm{sigmoid} \left(\mathbf{x}\mathbf{W}_1 + \mathbf{b}_1 \right) \right)}{\partial \mathbf{z}_1}$	(7)
=	$\boldsymbol{\delta}_1 \dfrac{\partial \left(\sigma \left(\mathbf{z}_1 \right) \right)}{\partial \mathbf{z}_1}$	(8)
=	$\boldsymbol{\delta}_1 \sigma' \left(\mathbf{z}_1 \right)$	(9)

$\dfrac{\partial CE}{\partial \mathbf{x}} =$	$\dfrac{\partial CE}{\partial \mathbf{z}_1} \dfrac{\partial \mathbf{z}_1}{\partial \mathbf{x}}$	(10)
=	$\boldsymbol{\delta}_3 \dfrac{\partial \left(\mathbf{x}\mathbf{W}_1 + \mathbf{b}_1 \right)}{\partial \mathbf{x}}$	(11)
=	$\boldsymbol{\delta}_3 \mathbf{W}_1^{\mathbf{T}}$	

$\frac{\partial CE}{\partial \mathbf{W_2}} =$	$\frac{\partial CE}{\partial \mathbf{z_2}} \frac{\partial \mathbf{z_2}}{\partial \mathbf{W_2}}$	(12)
=	$\boldsymbol{\delta}_1 \frac{\partial \left(\mathbf{h} \mathbf{W}_2 + \mathbf{b}_2 \right)}{\partial \mathbf{W}_2}$	(13)
=	$\mathbf{h}^{\mathbf{T}} \boldsymbol{\delta}_1$	(14)

$\frac{\partial CE}{\partial \mathbf{b_2}} =$	$\frac{\partial CE}{\partial \mathbf{z_2}} \frac{\partial \mathbf{z_2}}{\partial \mathbf{b_2}}$	(15)
=	$\boldsymbol{\delta}_1 \frac{\partial \left(\mathbf{h} \mathbf{W}_2 + \mathbf{b}_2 \right)}{\partial \mathbf{b}_2}$	(16)
=	$\mathbf{1}^{\mathbf{T}} \boldsymbol{\delta}_1$	(17)

$\frac{\partial CE}{\partial \mathbf{W_1}} =$	$\frac{\partial CE}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \mathbf{z_1}} \frac{\partial \mathbf{z_1}}{\partial \mathbf{W_1}}$	(18)
=	$\boldsymbol{\delta}_1 \mathbf{W}_2{}^{\mathbf{T}} \sigma' \left(\mathbf{z}_1 \right) \frac{\partial \left(\mathbf{x} \mathbf{W}_1 + \mathbf{b}_1 \right)}{\partial \mathbf{W}_1}$	(19)
=	$\mathbf{x}^{\mathbf{T}} \boldsymbol{\delta}_1 \mathbf{W}_2{}^{\mathbf{T}} \sigma' \left(\mathbf{z}_1 \right)$	(20)

$\frac{\partial CE}{\partial \mathbf{b_1}} =$	$\frac{\partial CE}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \mathbf{z_1}} \frac{\partial \mathbf{z_1}}{\partial \mathbf{b_1}}$	(21)
=	$\boldsymbol{\delta}_1 \mathbf{W}_2{}^{\mathbf{T}} \sigma' \left(\mathbf{z}_1 \right) \frac{\partial \left(\mathbf{x} \mathbf{W}_1 + \mathbf{b}_1 \right)}{\partial \mathbf{b}_1}$	(22)
=	$\mathbf{1}^{\mathbf{T}} \boldsymbol{\delta}_1 \mathbf{W}_2{}^{\mathbf{T}} \sigma' \left(\mathbf{z}_1 \right)$	(23)