2(a) Given:

$J(\theta) =$	$CE(\mathbf{y},\hat{\mathbf{y}}) = -\sum_{i=1}^{5}\left(y_i \log(y_i)\right)$	(1)
$\mathbf{h} =$	$\tanh(\mathbf{x}^{(t)}\mathbf{W} + \mathbf{b_1})$	(2)
Let $\mathbf{z_1} =$	$\mathbf{x}^{(t)}\mathbf{W} + \mathbf{b_1}$	(3)
$\hat{\mathbf{y}} =$	$\mathrm{softmax}(\mathbf{hU} + \mathbf{b_2})$	(4)
Let $\mathbf{z_2} =$	$\mathbf{hU} + \mathbf{b_2}$	(5)

$\frac{\partial J}{\partial \mathbf{U}} =$	$\frac{\partial J}{\partial \mathbf{z_2}} \frac{\partial \mathbf{z_2}}{\partial \mathbf{U}}$	(6)
$\frac{\partial J}{\partial \mathbf{z_2}} =$	$-\frac{\partial}{\partial \mathbf{z_2}} \sum_i \left(y_i \log(y_i)\right)$	(7)
=	$\hat{\mathbf{y}} - \mathbf{y}$ from Assignment1 2(b), by substituting $\mathbf{z_2}$ for $\boldsymbol{\theta}$	(8)
$\frac{\partial \mathbf{z_2}}{\partial \mathbf{U}} =$	$\mathbf{h}$	(9)
$\frac{\partial J}{\partial \mathbf{U}} =$	$\mathbf{h}^{\mathbf{T}}\left(\hat{\mathbf{y}} - \mathbf{y}\right)$	(10)

$\frac{\partial J}{\partial \mathbf{b_2}} =$	$\frac{\partial J}{\partial \mathbf{z_2}} \frac{\partial \mathbf{z_2}}{\partial \mathbf{b_2}}$	(11)
$\frac{\partial J}{\partial \mathbf{z_2}} =$	$\hat{\mathbf{y}} - \mathbf{y}$ from Equation (8)	(12)
$\frac{\partial \mathbf{z_2}}{\partial \mathbf{b_2}} =$	1	(13)
$\frac{\partial J}{\partial \mathbf{b_2}} =$	$\left(\hat{\mathbf{y}} - \mathbf{y}\right)$	(14)

$\frac{\partial J}{\partial \mathbf{h}} =$	$\frac{\partial J}{\partial \mathbf{z_2}} \frac{\partial \mathbf{z_2}}{\partial \mathbf{h}}$	(15)
$\frac{\partial J}{\partial \mathbf{z_2}} =$	$\hat{\mathbf{y}} - \mathbf{y}$ from Equation (8)	(16)
$\frac{\partial \mathbf{z_2}}{\partial \mathbf{h}} =$	$\mathbf{U}$	(17)
$\frac{\partial J}{\partial \mathbf{b_2}} =$	$\left(\hat{\mathbf{y}} - \mathbf{y}\right)\mathbf{U}^{\mathbf{T}}$	(18)

$$\frac{\partial J}{\partial \mathbf{W}} = \frac{\partial J}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \mathbf{W}} \tag{19}$$

$$= \mathbf{x}^{(t)\mathbf{T}} \left(\frac{\partial J}{\partial \mathbf{h}} \odot \tanh^{'}(\mathbf{x}^{(t)}\mathbf{W} + \mathbf{b_1}) \right) \tag{20}$$

$$\frac{\partial J}{\partial \mathbf{b_1}} = \frac{\partial J}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \mathbf{b_1}} \tag{21}$$

$$= \left(\frac{\partial J}{\partial \mathbf{h}} \odot \tanh^{'}(\mathbf{x}^{(t)}\mathbf{W} + \mathbf{b_1}) \right) \tag{22}$$

$$\frac{\partial J}{\partial \mathbf{x}^{(t)}} = \frac{\partial J}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \mathbf{x}^{(t)}} \tag{23}$$

$$= \left(\frac{\partial J}{\partial \mathbf{h}} \odot \tanh^{'}(\mathbf{x}^{(t)}\mathbf{W} + \mathbf{b_1}) \right) \mathbf{W}^{\mathbf{T}} \tag{24}$$

$$\frac{\partial CE}{\partial \mathbf{W}_2} = \frac{\partial CE}{\partial \mathbf{z}_2}\frac{\partial \mathbf{z}_2}{\partial \mathbf{W}_2} \tag{25}$$

$$= \boldsymbol{\delta}_1 \frac{\partial\left(\mathbf{h}\mathbf{W}_2 + \mathbf{b}_2\right)}{\partial \mathbf{W}_2} \tag{26}$$

$$= \mathbf{h}^{\mathbf{T}}\boldsymbol{\delta}_1 \tag{27}$$

$$\frac{\partial CE}{\partial \mathbf{b}_2} = \frac{\partial CE}{\partial \mathbf{z}_2}\frac{\partial \mathbf{z}_2}{\partial \mathbf{b}_2} \tag{28}$$

$$= \boldsymbol{\delta}_1 \frac{\partial\left(\mathbf{h}\mathbf{W}_2 + \mathbf{b}_2\right)}{\partial \mathbf{b}_2} \tag{29}$$

$$= \mathbf{1}^{\mathbf{T}}\boldsymbol{\delta}_1 \tag{30}$$

$$\frac{\partial CE}{\partial \mathbf{W}_1} = \frac{\partial CE}{\partial \mathbf{h}}\frac{\partial \mathbf{h}}{\partial \mathbf{z}_1}\frac{\partial \mathbf{z}_1}{\partial \mathbf{W}_1} \tag{31}$$

$$= \boldsymbol{\delta}_1 \mathbf{W}_2{}^{\mathbf{T}} \sigma'\left(\mathbf{z}_1\right)\frac{\partial\left(\mathbf{x}\mathbf{W}_1 + \mathbf{b}_1\right)}{\partial \mathbf{W}_1} \tag{32}$$

$$= \mathbf{x}^{\mathbf{T}}\boldsymbol{\delta}_1 \mathbf{W}_2{}^{\mathbf{T}} \sigma'\left(\mathbf{z}_1\right) \tag{33}$$

$$\frac{\partial CE}{\partial \mathbf{b}_1} = \frac{\partial CE}{\partial \mathbf{h}}\frac{\partial \mathbf{h}}{\partial \mathbf{z}_1}\frac{\partial \mathbf{z}_1}{\partial \mathbf{b}_1} \tag{34}$$

$$= \boldsymbol{\delta}_1 \mathbf{W}_2{}^{\mathbf{T}} \sigma'\left(\mathbf{z}_1\right)\frac{\partial\left(\mathbf{x}\mathbf{W}_1 + \mathbf{b}_1\right)}{\partial \mathbf{b}_1} \tag{35}$$

$$= \mathbf{1}^{\mathbf{T}}\boldsymbol{\delta}_1 \mathbf{W}_2{}^{\mathbf{T}} \sigma'\left(\mathbf{z}_1\right) \tag{36}$$