Q. 2(a)

Error function: $\frac{1}{2}\sum_{i=1}^{m}(\hat{y}(i) - y(i))^2$

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\ E(\theta)$$

$\nabla E(\theta)$

$$\frac{d}{dx}\text{sigmoid}(x) = \text{sigmoid}(x)(1-\text{sigmoid}(x))$$

$$\frac{\partial E}{\partial v} = \frac{\partial E}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial v} = \underbrace{\frac{1}{2} \times 2 \times \sum_{i=1}^{m}(\hat{y}(i) - y^{(i)})}_{\frac{\partial E}{\partial \hat{y}}} \underbrace{z^{(i)}}_{\frac{\partial y}{\partial v}}$$

$$= \sum_{i=1}^{m}(\hat{y}(i) - y(i))\, z(i)$$

gradient descent:

$$\boxed{v \leftarrow v - \eta \sum_{i=1}^{m}(\hat{y}(i) - y(i))\, z^{(i)}}$$

update for output layer parameters

$$\frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z_j} \frac{\partial z_j}{\partial w_j} = \underbrace{\sum_{i=1}^{m} (\hat{y}(i) - y(i))\, z(i)}_{\frac{\partial E}{\partial y}} \times \underbrace{v_j}_{\frac{\partial \hat{y}}{\partial z_j}} \times$$

$$\times \underbrace{z_j^{(i)} (1 - z_j^{(i)})\, x^{(i)}}_{\frac{\partial z_j}{\partial w_j}}$$

gradient descent:

$$w_j \leftarrow w_j - \eta \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)})\, v_j z_j (1 - z_j^{(i)})\, x^{(i)}$$

Algorithm:

1) Start with the guess for V & $\{w_i\}$
2) Use $x^{(i)}$ and current guess of V & $\{w_j\}$ to compute $z^{(i)}$ (output of hidden layer) then $\hat{y}^{(i)}$
3) Use $z^{(i)}$ and $\hat{y}^{(i)}$ to update V
4) Use $z^{(i)}$, $\hat{y}^{(i)}$ and V to update W
5) keep doing this until the objective changes.