BGN: 2191A
P6
Q2

a. Let $(\vec{x}, y')$ be the labelled example

$$\delta = \frac{\partial E(\theta)}{\partial(a)} = \frac{\partial E(\theta)}{\partial y} \cdot \frac{\partial y}{\partial a} \quad \text{by the chain rule}$$

$$= \frac{\partial E(\theta)}{\partial y}\left(\frac{\partial}{\partial a}\sigma(a)\right)$$

$$= \sigma'(a)\frac{\partial E(\theta)}{\partial y}$$

b. Let $\vec{v} = (v_1, v_2, v_3, \dots, v_p)^T$

For $1 \leq i \leq p$,

$$\frac{\partial}{\partial v_i} E(\theta) = \delta \cdot \frac{\partial a}{\partial v_i} \quad \text{by the chain rule}$$

$$= \delta \cdot \frac{\partial}{\partial v_i} \sum_{i'=1}^{p} v_{i'}\sigma(a_{i'}) + v_0$$

$$= \delta \cdot \frac{\partial}{\partial v_i}(v_i \sigma(a_i))$$

$$= \underline{\delta\sigma(a_i)} = \sigma(a_i)\sigma'(a)\frac{\partial E(\theta)}{\partial y}$$

$$\frac{\partial}{\partial v_0} E(\theta) = \delta \cdot \frac{\partial a}{\partial v_0} \quad \text{by the chain rule}$$

$$= \delta \cdot \frac{\partial}{\partial v_0} \sum_{i'=1}^{p} v_{i'}\sigma(a_{i'}) + v_0$$

$$= \underline{\delta p} = p\sigma'(a)\frac{\partial E(\theta)}{\partial y}$$

c. $\delta_i = \frac{\partial E(\theta)}{\partial a_i} = \delta \cdot \frac{\partial a}{\partial a_i}$ by the chain rule

$$= \delta \cdot \frac{\partial}{\partial a_i} \sum_{i'=1}^{p} v_{i'} \sigma(a_{i'}) + v_0$$

$$= \delta \frac{\partial}{\partial a_i} (v_i \sigma(a_i))$$

$$= \delta v_i \sigma'(a_i)$$

d. Let $\vec{w} = (w_1, w_2, \ldots, w_n)^T$

$\frac{\partial E(\theta)}{\partial w_i} = \delta \cdot \frac{\partial a}{\partial w_i}$ by the chain rule

$$= \delta \cdot \frac{\partial}{\partial w_i} \sum_{i'=1}^{p} v_{i'} \sigma(a_{i'}) + v_0$$

$= \delta \cdot \sum_{i'=1}^{p} \frac{\partial}{\partial w_i} (v_{i'} \sigma(a_{i'}) + v_0)$ by linearity

$= \delta \cdot \sum_{i'=1}^{p} v_{i'} \sigma'(a_{i'}) \frac{\partial a_{i'}}{\partial w_i}$ by the chain rule

$$= \delta \cdot \sum_{i'=1}^{p} v_{i'} \sigma'(a_{i'}) \cdot \frac{\partial}{\partial w_i} (\vec{w}^T x^{(i')} + w_0)$$

For $i = 0$:

$$\frac{\partial E(\theta)}{\partial w_0} = \delta \cdot \sum_{i'=1}^{p} v_{i'} \sigma'(a_{i'})$$

For $1 \leq i \leq n$

$$\frac{\partial E(\theta)}{\partial w_i} = \delta \cdot \sum_{i'=1}^{p} v_{i'} \sigma'(a_{i'}) \frac{\partial}{\partial w_i} \sum_{k=1}^{n} w_k x_k^{(i')}$$

$$= \delta \cdot \sum_{i'=1}^{p} v_{i'} \sigma'(a_{i'}) x_i^{(i')}$$