Question 1a

We have

$$\begin{aligned} v(y) \cdot \phi(x;\theta) + \gamma_y &= v(y) \cdot (\alpha \times (Wx + b) + c) + \gamma_y \\ &= \alpha v(y) \cdot (Wx) + \alpha v(y) \cdot b + v(y) \cdot c + \gamma_y \\ &= v'(y) \cdot x + \gamma'_y \end{aligned}$$

where

$$v'(y) = \alpha W^\top v(y) \quad \gamma'_y = \alpha v(y) \cdot b + v(y) \cdot c + \gamma_y$$

Question 1b

We have

$$\begin{aligned} v(y) \cdot \phi(x;\theta) + \gamma_y &= v(y) \cdot (A(Wx + b) + c) + \gamma_y \\ &= v(y) \cdot (AWx) + v(y) \cdot (Ab + c) + \gamma_y \\ &= v'(y) \cdot x + \gamma'_y \end{aligned}$$

where

$$v'(y) = (AW)^\top v(y) \quad \gamma'_y = v(y) \cdot (AB + c) + \gamma_y$$

Question 2a

A useful property:

$$\frac{\partial}{\partial l_{y'}} \left(\log \sum_{y''} \exp\{l_{y''}\} \right) = \frac{\frac{\partial}{\partial l_{y'}} \left(\sum_{y''} \exp\{l_{y''}\} \right)}{\sum_{y''} \exp\{l_{y''}\}} = \frac{\exp\{l_{y'}\}}{\sum_{y''} \exp\{l_{y''}\}}$$

Question 2a (continued)

We have

$$\mathsf{LS}_y(l) = l_y - \log \sum_{y''} \exp\{l_{y''}\}$$

Hence

$$\frac{\partial \mathsf{LS}_y(l)}{\partial l_y} = 1 - \frac{\exp\{l_y\}}{\sum_{y'} \exp\{l_{y'}\}}$$

Similarly for y, y' such that $y \neq y'$,

$$\frac{\partial \mathsf{LS}_y(l)}{\partial l_{y'}} = -\frac{\exp\{l_{y'}\}}{\sum_{y''} \exp\{l_{y''}\}}$$

Question 2b

$$\frac{\partial l_{y'}}{\partial W_{j,k}} = \sum_{k=1}^{m} \frac{\partial l_{y'}}{\partial h_k} \frac{\partial h_k}{\partial W_{j,k}} = \sum_{k=1}^{m} \frac{\partial l_{y'}}{\partial h_k} \frac{\partial g_k(z_k)}{\partial z_k} x_j$$

$$\frac{\partial o}{\partial q_y} = \quad -1 \text{ if } y = y_i,\ 0 \text{ otherwise}$$

$$\frac{\partial q_y}{\partial l_{y'}} = [[y = y']] - \frac{\exp\{l_{y'}\}}{\sum_{y''} \exp\{l_{y''}\}}$$

$$\frac{\partial l_{y'}}{\partial h_k} = V_{y',k}$$

Question 4

Under this mapping, we have the following data points:

$$h = [0, 0, 0, 0, 0], \quad y = -1$$

$$h = [0, 1, 0, 0, 1], \quad y = +1$$

$$h = [1, 0, 0, 1, 0], \quad y = +1$$

$$h = [1, 1, 1, 1, 1], \quad y = +1$$

Now if we set $u = [1, 1, -4, 0, 0]$ we have the following data:

$$u.h = 0, \quad y = -1$$

$$u.h = 1, \quad y = +1$$

$$u.h = 1, \quad y = +1$$

$$u.h = -2, \quad y = -1$$

So setting $u = [1, 1, -4, 0, 0]$, $\gamma = -0.5$ gives $u.h + \gamma > 0$ for any example labeled $+1$, and $u.h + \gamma < 0$ for any example labeled -1.

It follows that if we choose $v(+1) - v(-1) = [1, 1, -4, 0, 0]$ and $\gamma_{+1} - \gamma_{-1} = -0.5$ we correctly model the data.