CHAPTER 5: NEURAL NETWORKS

(5.1) $y_k(x, w) = \sigma\left[\sum_{j=1}^{M} w_{kj}^{(2)} h\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right]$ (5.1)

+) $\tanh(x) = 2\sigma(2x) - 1$

$$\Rightarrow \sigma\left[\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) = \tanh\left(\frac{\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}}{2}\right) + 1$$

$$\Rightarrow y_k(x, w) = \sigma\left[\sum_{j=1}^{M} w_{kj}^{(2)} \frac{\left(\tanh\left(\frac{\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}}{2}\right) + 1\right)}{2} + w_{k0}^{(2)}\right]$$

$$y_k(x, w) = \sigma\left[\sum_{j=1}^{M} \frac{w_{kj}^{(2)}}{2} \tanh\left(\sum_{i=1}^{D} \frac{w_{ji}^{(1)}}{2} x_i + \frac{w_{j0}^{(1)}}{2}\right) + \sum_{j=1}^{M} \frac{w_{kj}^{(2)}}{2} + w_{k0}^{(2)}\right]$$

$$\Rightarrow \begin{cases} w'^{(1)}_{ji} = \dfrac{w_{ji}^{(1)}}{2} \\ w'^{(1)}_{j0} = \dfrac{w_{j0}^{(1)}}{2} \end{cases} \text{ and } \begin{cases} w'^{(2)}_{kj} = \dfrac{w_{kj}^{(2)}}{2} \\ w'^{(2)}_{k0} = w_{k0}^{(2)} + \sum_{j=1}^{M} \dfrac{w_{kj}^{(2)}}{2} \end{cases}$$

(5.26)

+) $\Omega_n = \frac{1}{2} \sum_k \sum_i (J_{ki} \tau_i)^2$

$= \frac{1}{2} \sum \sum (\tau_i \cdot \nabla y_k)^2$

$$\ldots w_{j0}) = \tanh\left(\frac{\sum_{i=1} w_{ji} x_i + w_{j0}^{(1)}}{2}\right) + 1$$

$$\Rightarrow y_k(x,w) = \sigma\left[\sum_{j=1}^{M} w_{kj}^{(2)} \left(\frac{\tanh\left(\frac{\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}}{2}\right) + 1}{2}\right) + w_{k0}^{(2)}\right]$$

$$y_k(x,w) = \sigma\left[\sum_{j=1}^{M} \frac{w_{kj}^{(2)}}{2} \tanh\left(\sum_{i=1}^{D} \frac{w_{ji}^{(1)}}{2} x_i + \frac{w_{j0}^{(1)}}{2}\right) + \sum_{j=1}^{M} \frac{w_{kj}^{(2)}}{2} + w_{k0}^{(2)}\right]$$

$$\Rightarrow \begin{cases} w'^{(1)}_{ji} = \dfrac{w^{(1)}_{ji}}{2} \\ w'^{(1)}_{j0} = \dfrac{w^{(1)}_{j0}}{2} \end{cases} \text{ and } \begin{cases} w'^{(2)}_{kj} = \dfrac{w^{(2)}_{kj}}{2} \\ w'^{(2)}_{k0} = w^{(1)}_{k0} + \displaystyle\sum_{j=1}^{M} \frac{w_{kj}}{2} \end{cases}$$

5.26

+) $$\Omega_n = \frac{1}{2}\sum_k \sum_i (J_{ki}\,\tau_i)^2$$

$$= \frac{1}{2}\sum_k \sum_i \left(\tau_i \frac{\partial y_k}{\partial x_i}\right)^2$$

$$= \frac{1}{2}\sum_k (G y_k)^2 \quad \left(\text{because } G y_k = \sum_i \tau_i \frac{\partial y_k}{\partial x_i}\right)$$

f) Forward propagation:

So to calculate $G y_h$ we have to find;

$G z_j$ of layer 1

$G z_j$ ——— 2

. ——— ——— ———

$G y_h$

$$\Rightarrow G z_j = \sum_i \delta_i \frac{\partial h(a_j)}{\partial x_i}$$

$$= \sum_i \delta_i \frac{\partial h(a_j)}{\partial a_j} \cdot \frac{\partial a_j}{\partial x_i}$$

$$= h'(a_j) \sum_i \delta_i \frac{\partial a_j}{\partial x_i}$$

Previous layer

$$= h'(a_j) \sum_i \delta_i \frac{\partial}{\partial x_i} \left(\sum_m w_{jm} z_m \right)$$

$$= h'(a_j) \sum_i \delta_i \left(\sum_m w_{jm} \frac{\partial z_m}{\partial x_i} \right)$$

$h'(a_j)\sum_i \delta_i \frac{\partial a_j}{\partial x_i}$

Previous layer

$= h'(a_j)\sum_i \delta_i \frac{\partial}{\partial x_i}\left(\sum_m w_{jm} z_m\right)$

$= h'(a_j)\sum_i \delta_i \left(\sum_m w_{jm} \underbrace{\frac{\partial z_m}{\partial x_i}}_{\varphi_m}\right)$

$\Rightarrow G z_j = h'(a_j)\sum_i \delta_i \underbrace{\left(\sum_m w_{jm}\varphi_m\right)}_{B_j}$

$\Rightarrow$ $\boxed{G z_j = h'(a_j) B_j \quad \text{with} \quad B_j = \sum_i w_{ji}\varphi_i}$

+) ~~Backward propagation~~

+) Derivatives of Ω_n with respect to a weight w_{rs}:

$$\frac{\partial R_n}{\partial w_{rs}} = \frac{\partial}{\partial w_{rs}}\left[\frac{1}{2}\sum_h (G_{y_h})^2\right]$$

$$= \frac{1}{2}\sum_h \frac{\partial (G_{y_h})^2}{\partial w_{rs}}$$

$$= \cancel{\frac{1}{2}}\sum_h G_{y_h}\frac{\partial G_{y_h}}{\partial w_{rs}}$$

$$= \sum_h G_{y_h}\frac{\partial}{\partial w_{rs}}\left(\sum_i z_i \frac{\partial y_h}{\partial x_i}\right)$$

$$= \sum_h G_{y_h}\sum_i z_i \frac{\partial^2 y_h}{\partial w_{rs}\partial x_i}$$

$$= \sum_h G_{y_h}\sum_i z_i \frac{\partial}{\partial x_i}\left(\frac{\partial y_h}{\partial w_{rs}}\right)$$

$$= \sum G \ldots \sum z_i \frac{\partial}{\ldots}\left(\frac{\partial y_h}{\ldots}\right.$$

$$= \sum_h G_{y_h} \sum_i \sigma_i \frac{\partial^2 y_h}{\partial w_{rs} \partial x_i}$$

$$= \sum_h G_{y_h} \sum_i \sigma_i \frac{\partial}{\partial x_i}\left(\frac{\partial y_h}{\partial w_{rs}}\right)$$

$$= \sum_h G_{y_h} \sum_i \sigma_i \frac{\partial}{\partial x_i}\left(\frac{\partial y_h}{\partial a_r} \frac{\partial a_r}{\partial w_{rs}}\right)$$

because: $a_r = \sum_s w_{rs} z_s$ where z_s is from the previous layer. ~~We see that only a_r~~

$$= \sum_h G_{y_h} \sum_i \sigma_i \frac{\partial}{\partial x_i}\left(\delta_{hr} z_s\right)$$

$$= \sum_h G_{y_h} \left[\sum_i \sigma_i \left(\frac{\partial \delta_{hr}}{\partial x_i} z_s + \delta_{hr} \frac{\partial z_s}{\partial x_i}\right)\right]$$

$$= \sum_h G_{y_h} \left[\left(\sum_i \sigma_i \frac{\partial \delta_{hr}}{\partial x_i}\right) z_s + \delta_{hr} \left(\sum_i \frac{\partial z_s}{\partial x_i}\right.\right.$$

$$\sum_h G_{y_h} \sum_i \tau_i \frac{\partial}{\partial x_i}\left(\frac{\partial y_h}{\partial a_r}\frac{\partial a_r}{\partial w_{rs}}\right)$$

because: $a_r = \sum_s w_{rs} z_s$ where z_s is from the previous layer. ~~We see that only~~ a_r

$$= \sum_h G_{y_h} \sum_i \tau_i \frac{\partial}{\partial x_i}\left(\delta_{hr} z_s\right)$$

$$= \sum_h G_{y_h}\left[\sum_i \tau_i\left(\frac{\partial \delta_{hr}}{\partial x_i} z_s + \delta_{hr}\frac{\partial z_s}{\partial x_i}\right)\right]$$

$$= \sum_h G_{y_h}\left[\underbrace{\left(\sum_i \tau_i \frac{\partial \delta_{hr}}{\partial x_i}\right)}_{} z_s + \delta_{hr}\underbrace{\left(\sum_i \frac{\partial z_s}{\partial x_i}\tau_i\right)}_{}\right]$$

$$= \sum_h G_{y_h}\left[\phi_{hr} z_s + \delta_{hr}\, \alpha_s\right].$$

$$= \sum_h G_{y_h}\left(\phi_{hr} z_s + \delta_{hr}\, \alpha_s\right) \quad (\alpha_h = G_{y_h}$$

$$= \sum_h \alpha_h\left(\phi_{hr} z_s + \delta_{hr}\, \alpha_s\right)$$

7) Write down the backprop equations for δ_{kr}:

We have; $a_r = \sum_s w_{rs} z_s$

$$\Rightarrow \frac{\partial y_k}{\partial a_s} = \sum_r \frac{\partial y_k}{\partial a_r} \frac{\partial a_r}{\partial z_s} \frac{\partial z_s}{\partial a_s}$$

$$\Rightarrow \delta_{ks} = \sum_r \delta_{kr} w_{rs} h'(a_s)$$

$$\Rightarrow \boxed{\underbrace{\delta_{ks}}_{\text{Layer}(s)} = h'(a_s) \sum_r w_{rs} \underbrace{\delta_{kr}}_{\text{Layer}(r)}}$$

) Derive a set of backprop equations for the evaluati[on of]

[t]he ϕ_{kr}:

W[e] ...

$$\Rightarrow \boxed{\delta_{ks} = h'(a_s) \sum_r w_{rs} \delta_{kr}}$$

$\delta_{ks} \rightarrow$ Layer (s) $\qquad \delta_{kr} \rightarrow$ Layer (r)

+) Derive a set of backprop equations for the evaluation of the ϕ_{kr}:

We have;

$$\phi_{kr} = \sum_i \zeta_i \frac{\partial \delta_{kr}}{\partial x_i}$$

So;

$$\phi_{ks} = \sum_i \zeta_i \frac{\partial \delta_{ks}}{\partial x_i} \qquad \left(a_r = \sum_s w_{rs} z_s \right.$$

$$\phi_{ks} = \sum_i \zeta_i \frac{\partial}{\partial x_i} \left[h'(a_s) \sum_r w_{rs} \delta_{kr} \right] \qquad \left(z_s = h(a_s) \right)$$

$$\Rightarrow \phi_{ks} = \sum_i \zeta_i \left[\frac{\partial h'(a_s)}{\partial x_i} \sum_r w_{rs} \delta_{kr} + h'(a_s) \sum_r w_{rs} \frac{\partial \delta_{kr}}{\partial x_i} \right.$$

$$\Rightarrow \phi_{ks} = \sum_i \zeta_i \left[h''(a_s) \frac{\partial a_s}{\partial x_i} \sum w_{rs} \delta_{kr} + h'(a_s) \sum w_{rs} \right.$$

layer(s)

ii) Derive a set of backprop equations for the evaluation of the Φ_{kr}:

We have:

$$\Phi_{kr} = \sum_i \tau_i \frac{\partial \delta_{kr}}{\partial x_i}$$

So:

$$\Phi_{ks} = \sum_i \tau_i \frac{\partial \delta_{ks}}{\partial x_i}$$

$$\left(a_r = \sum_s w_{rs} z_s \qquad z_s = h(a_s) \right)$$

$$\Phi_{ks} = \sum_i \tau_i \frac{\partial}{\partial x_i}\left[h'(a_s) \sum_r w_{rs} \delta_{kr} \right]$$

$$\Rightarrow \Phi_{ks} = \sum_i \tau_i \left[\frac{\partial h'(a_s)}{\partial x_i} \sum_r w_{rs} \delta_{kr} + h'(a_s) \sum_r w_{rs} \frac{\partial \delta_{kr}}{\partial x_i} \right.$$

$$\Rightarrow \Phi_{ks} = \sum_i \tau_i \left[h''(a_s) \frac{\partial a_s}{\partial x_i} \sum_r w_{rs} \delta_{kr} + h'(a_s) \sum_r w_{rs} \frac{\partial \delta_{kr}}{\partial x_i} \right.$$

$$\Rightarrow \Phi_{ks} = h''(a_s) \left(\sum_i \tau_i \frac{\partial a_s}{\partial x_i} \right) \left(\sum_r w_{rs} \delta_{kr} \right) + h'(a_s) \sum_i \tau_i \left(\right.$$

$$\Rightarrow \boxed{\ldots}$$

(5.29)

$$\Rightarrow P(w) = \prod_i p(w_i) = \prod_i \sum_{j=1}^{M} \pi_j \, N(w_i | \mu_j, \sigma_j^2)$$

$$\Rightarrow N(w_i | \mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi}\,\sigma_j} \exp\left[\left(\frac{-1}{2}\right) \frac{(w_i - \mu_j)^2}{\sigma_j^2}\right]$$

$$\Rightarrow \frac{\partial \Omega(w)}{\partial w_i} = \frac{\partial}{\partial w_i}\left[(-1) \sum_j \ln p(w_j)\right]$$

$$= (-1) \frac{1}{p(w_i)} \frac{\partial p(w_i)}{\partial w_i}$$

$$\Rightarrow \frac{\partial p(w_i)}{\partial w_i} = \frac{\partial}{\partial w_i}\left[\sum_{j=1}^{M} \pi_j \, N(w_i | \mu_j, \sigma_j^2)\right]$$

$$= \frac{\partial}{\partial w_i}\left[\sum_{j=1}^{M} \pi_j \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{(-\frac{1}{2})(w_i - \mu_j)^2}{2\sigma_j^2}}\right]$$

$$\frac{(-\frac{1}{2})(w_i - \mu_j)^2}{\ldots}$$

$$\Rightarrow \boxed{\Phi_{ks} = h''(a_s)\,\beta_s \sum_r w_{rs}\,\delta_{kr} + h'(a_s) \sum_r w_{rs} \cdot \Phi_{kr}}$$

(5,29) ~~ln P(w~~

$$P(w) = \prod_i p(w_i) = \prod_i \sum_{j=1}^{m} \pi_j \, N(w_i \mid \mu_j, \sigma_j^2)$$

$$N(w_i \mid \mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi}\,\sigma_j} \exp\left[\left(\frac{-1}{2}\right) \frac{(w_i - \mu_j)^2}{\sigma_j^2}\right]$$

$$\frac{\partial \Omega(w)}{\partial w_i} = \frac{\partial}{\partial w_i}\left[(-1) \sum_j \ln p(w_j)\right]$$

$$= (-1) \frac{1}{p(w_i)} \frac{\partial p(w_i)}{\partial w_i}$$

$$\partial p(w_i) \quad \partial \quad \left[\sum^{M} \pi \; N(w_i \mid \mu_j, \sigma^2 \right.$$

$$\sqrt{2\pi}\,\sigma_j \qquad \left(\frac{1}{2}\right)\frac{(w_i-\mu_j)^2}{\sigma_j^2}$$

+) $$\frac{\partial \Omega(w)}{\partial w_i} = \frac{\partial}{\partial w_i}\left[(-1)\sum_j^N \ln P(w_j)\right]$$

$$= (-1)\frac{1}{P(w_i)}\frac{\partial P(w_i)}{\partial w_i}$$

+) $$\frac{\partial P(w_i)}{\partial w_i} = \frac{\partial}{\partial w_i}\left[\sum_{j=1}^{M}\pi_j \, N(w_i \mid \mu_j, \sigma_j^2)\right]$$

$$= \frac{\partial}{\partial w_i}\left[\sum_{j=1}^{M}\pi_j \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2}\right)\frac{(w_i-\mu_j)^2}{2\sigma_j^2}}\right.$$

$$= \sum_{j=1}^{M}\pi_j \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2}\right)\frac{(w_i-\mu_j)^2}{2\sigma_j^2}} \frac{(-1)(w_i-\mu_j)}{\sigma_j^2}$$

$$\Rightarrow \frac{\partial P(w_i)}{\partial w_i} = \sum_{j=1}^{M}\pi_j \, N(w_i \mid \mu_j, \sigma_j^2)\frac{(-1)(w_i-\mu_j)}{\sigma_j^2}$$

$$\Rightarrow \frac{\partial \Omega(w)}{\partial w_i} = (-1)\frac{1}{P(w_i)}\frac{\partial P(w_i)}{\partial w_i}$$

$$= (-1)\frac{1}{\tilde{p}(w_i)} \frac{\partial \tilde{p}(w_i)}{\partial w_i}$$

$$+)\ \frac{\partial \tilde{p}(w_i)}{\partial w_i} = \frac{\partial}{\partial w_i}\left[\sum_{j=1}^{m} \pi_j N(w_i|\mu_j, \sigma_j^2)\right]$$

$$= \frac{\partial}{\partial w_i}\left[\sum_{j=1}^{m} \pi_j \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{(-1)(w_i-\mu_j)^2}{2\sigma_j^2}}\right]$$

$$= \sum_{j=1}^{m} \pi_j \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{(-1)(w_i-\mu_j)^2}{2\sigma_j^2}} \frac{(-1)(w_i-\mu_j)}{\sigma_j^2}$$

$$\Rightarrow \frac{\partial \tilde{p}(w_i)}{\partial w_i} = \sum_{j=1}^{m} \pi_j N(w_i|\mu_j, \sigma_j^2) \frac{(-1)(w_i-\mu_j)}{\sigma_j^2}$$

$$\Rightarrow \frac{\partial \Omega(w)}{\partial w_i} = (-1)\cdot\frac{1}{p(w_i)}\cdot\frac{\partial p(w_i)}{\partial w_i}$$

$$= (-1)\cdot\frac{\sum_{j=1}^{m} \pi_j N(w_i|\mu_j, \sigma_j^2)\frac{(-1)(w_i-\mu_j)}{\sigma_j^2}}{\sum_{h=1}^{m} \pi_h N(w_i|\mu_h, \sigma_h^2)}$$

$$w_{rs}\frac{\partial \delta_{rs\,hr}}{\partial x_i}$$

$$\Rightarrow \frac{\partial \Omega(w)}{\partial w_i} = \sum_{j=1}^{M} \left[\frac{\pi_j N(w_i \mid \mu_j, \sigma_j^2)}{\sum_{k=1}^{M} \pi_k N(w_k \mid \mu_k, \sigma_k^2)} \right.$$

$$\Rightarrow \frac{\partial \Omega(w)}{\partial w_i} = \sum_{j=1}^{M} \gamma_j(w_i) \frac{(w_i - \mu_j)}{\sigma_j^2}$$

3d) ... me as 5,29) (almost ;)

$$\frac{\partial}{\partial \sigma_j} \left[(-1) \sum_i \ln p(w_i) \right]$$

$\partial w_i \qquad \sigma_j^2$

5,30) The same as 5,29) (almost ;)

5,31) $\frac{\partial \ln(w)}{\partial \sigma_j} = \frac{\partial}{\partial \sigma_j}\left[(-1)\sum_i \ln p(w_i)\right]$

$= \frac{(-1)}{p(w_i)} \frac{\partial p(w_i)}{\partial \sigma_j}$

$\rightarrow) \frac{\partial p(w_i)}{\partial \sigma_j} = \frac{\partial}{\partial \sigma_j}\left[\sum_{j=1}^{M} \pi_j \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2}\right)\frac{(w_i - \mu_j)^2}{\sigma_j^2}}\right]$

$\sum_{j=1}^{M} \pi_j \frac{-1}{\sqrt{2\pi}\sigma^2} e^{\left(-\frac{1}{2}\right)\frac{(w_i - \mu_j)^2}{\sigma_j^2}} + \pi_j \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2}\right)\frac{(w_i - \mu_j)^2}{\sigma_j^2}} \cdot$

$\sum_{j=1}^{M} \pi_j N(w_i|\mu_j, \sigma_j^2) \cdot \left(\frac{-1}{\sigma} + \frac{(w_i - \mu_j)^2}{\ldots}\right.$

$$+)\ \frac{\partial \ln(w_i)}{\partial \sigma_j} = \frac{\partial}{\partial \sigma_j}\left[\sum_{j=1}^{M} \pi_j \frac{1}{\sqrt{2\pi}\sigma} e^{(-\frac{1}{2})\frac{(w_i - \mu_j)^2}{\sigma_j^2}} \right]$$

$$= \sum_{j=1}^{M} \pi_j \frac{-1}{\sqrt{2\pi}\sigma^2} e^{(-\frac{1}{2})\frac{(w_i-\mu_j)^2}{\sigma_j^2}} + \pi_j \frac{1}{\sqrt{2\pi}\sigma} e^{(-\frac{1}{2})\frac{(w_i-\mu_j)^2}{\sigma_j^2}} \cdot \frac{(w_i-}{\sigma^3}$$

$$= \sum_{j=1}^{M} \pi_j N(w_i \mid \mu_j, \sigma_j^2) \left(\frac{-1}{\sigma} + \frac{(w_i - \mu_j)^2}{\sigma_j^3} \right)$$

$$\Rightarrow \frac{\partial l(w)}{\partial \sigma_j} = (-1) \frac{\sum_{j=1}^{M} \pi_j N(w_i \mid \mu_j, \sigma_j^2) \left(\frac{-1}{\sigma_j} + \frac{(w_i - \mu_j)^2}{\sigma_j^3} \right)}{\sum_{h=1}^{M} \pi_h N(w_i \mid \mu_h, \sigma_h^2)}$$

$$= \sum_{j=1}^{M} \left[\frac{\pi_j N(w_i \mid \mu_j, \sigma_j^2) \left(\frac{1}{\sigma_j} - \frac{(w_i - \mu_j)^2}{\sigma_j^3} \right)}{\sum_{h=1}^{M} \pi_h N(w_i \mid \mu_h, \sigma_h^2)} \right]$$

$$\Rightarrow \frac{\partial \Omega(w)}{\partial \sigma_j^2} = \sum_j \partial_j (w_i) \left[\frac{1}{\sigma^2_j} - \frac{(w_i - x_j)^2}{\sigma^3_j} \right.$$

$$(5,32)\quad \pi_k = \frac{e^{\eta_k}}{\sum_m e^{\eta_m}}$$

$j \neq k$:

$$\frac{\partial \pi_k}{\partial \eta_j} = e^{\eta_k} \cdot \frac{(-1)}{\left(\sum_m e^{\eta_m}\right)^2} \cdot \frac{\partial\left(\sum_m e^{\eta_m}\right)}{\partial \eta_j}$$

$$\Rightarrow \frac{\partial \pi_k}{\partial \eta_j} = (-1) \frac{e^{\eta_k}}{\left(\sum_m e^{\eta_m}\right)^2} \cdot e^{\eta_j}$$

$$\partial \pi_k = (-1)\left(\frac{e^{\eta_k}}{\ldots}\right) \cdot \left(\frac{e^{\eta_j}}{\ldots e^{\eta_m}}\right)$$

$$\sum_m e^{\eta_m}$$

+) $j \neq k$:

$$\frac{\partial \pi_k}{\partial \eta_j} = e^{\eta_k} \cdot \frac{(-1)}{\left(\sum_m e^{\eta_m}\right)^2} \cdot \frac{\partial\left(\sum_m e^{\eta_m}\right)}{\partial \eta_j}$$

$$\Rightarrow \frac{\partial \pi_k}{\partial \eta_j} = (-1)\,\frac{e^{\eta_k}}{\left(\sum_m e^{\eta_m}\right)^2} \cdot e^{\eta_j}$$

$$\Rightarrow \frac{\partial \pi_k}{\partial \eta_j} = (-1)\left(\frac{e^{\eta_k}}{\sum_m e^{\eta_m}}\right) \cdot \left(\frac{e^{\eta_j}}{\sum_m e^{\eta_m}}\right)$$

$$\Rightarrow \frac{\partial \pi_k}{\partial \eta_j} = (-1) \cdot \pi_k \cdot \pi_j$$

+) $j = k$:

$$\frac{\partial \pi_k}{\partial \eta_j} = \frac{\partial \pi_k}{\partial \eta_k} = \frac{\frac{\partial\left[e^{\eta_k}\right]}{\partial \eta_k} \cdot \sum_m e^{\eta_m} - e^{\eta_k} \cdot \frac{\partial\left(\sum_m e^{\eta_m}\right)}{\partial \eta_k}}{\left(\sum_m e^{\eta_m}\right)^2}$$

$$\Rightarrow \frac{\partial \pi_k}{\partial \eta_k} = \frac{e^{\eta_k} \cdot \left(\sum_m e^{\eta_m}\right) - e^{\eta_k} \cdot e^{\eta_k}}{\left(\sum e^{\eta_m}\right)^2}$$

$$\Rightarrow \frac{\partial \pi_h}{\partial \eta_j} = (-1)\left(\frac{e^{\eta_h}}{\sum_m e^{\eta_m}}\right)\cdot\left(\frac{e^{\eta_j}}{\sum_m e^{\eta_m}}\right)$$

$$\Rightarrow \frac{\partial \pi_h}{\partial \eta_j} = (-1)\cdot \pi_h \cdot \pi_j$$

+) $j = h$:

$$\frac{\partial \pi_h}{\partial \eta_j} = \frac{\partial \pi_h}{\partial \eta_h} = \frac{\frac{\partial[e^{\eta_h}]}{\partial \eta_h}\cdot \sum_m e^{\eta_m} - e^{\eta_h}\cdot \frac{\partial(\sum_m e^{\eta_m})}{\partial \eta_h}}{\left(\sum_m e^{\eta_m}\right)^2}$$

$$\Rightarrow \frac{\partial \pi_h}{\partial \eta_h} = \frac{e^{\eta_h}\cdot\left(\sum_m e^{\eta_m}\right) - e^{\eta_h}\cdot e^{\eta_h}}{\left(\sum_m e^{\eta_m}\right)^2}$$

$$= \frac{e^{\eta_h}}{\sum_m e^{\eta_m}} - \frac{e^{\eta_h}}{\sum_m e^{\eta_m}}\cdot\frac{e^{\eta_h}}{\sum_m e^{\eta_m}}$$

$$= \pi_h - \pi_h\,\pi_h$$