$$y_i = \frac{\exp(a_i)}{\sum_{j=1}^{k} \exp(a_j)}$$

$i \neq j$

$$\frac{\partial y_i}{\partial a_j} = -\frac{\exp(a_i)}{\left(\sum_{l=1}^{k} \exp(a_l)\right)^2} \cdot \exp(a_j)$$

$$= -y_i \cdot y_j$$

$i = j$

$$\frac{\partial y_i}{\partial a_j} = \frac{\sum_{l=1}^{k} \exp(a_l) \cdot \exp(a_i) - \exp(a_i) \cdot \exp(a_j)}{\left(\sum_{l=1}^{k} \exp(a_l)\right)^2}$$

$$\left(\frac{f}{g}\right)' = \frac{gf' - fg'}{g^2}$$

$$= \frac{\exp(a_i) \cdot \left(\sum_{l=1}^{k} \exp(a_l) - \exp(a_j)\right)}{\left(\sum_{l=1}^{k} \exp(a_l)\right)^2}$$

$$= y_i (1 - y_j)$$

$$\frac{\partial y_i}{\partial a_j} = y_i (\sigma_{ij} - y_j)$$