

$$D_j S_i = \frac{\partial \frac{e^{a_i}}{\sum_{k=0}^N e^{a_k}}}{\partial a_j}$$

The softmax derivative looks a cool candidate for quotient rule $f(x) = \frac{g(x)}{h(x)}$

$$g(x) = e_i^a$$

$$h(x) = \sum_{k=0}^N e^{a_k}$$

The following cases will help in our quest to compute jacobians.

1) for $N=0..2$,

$$\sum_{k=0}^N e^{a_k} = e^{a_0} + e^{a_1} + e^{a_2}$$

if we compute derivative of all elements w.r.t to a_0 all derivative except $\frac{\partial e^{a_0}}{\partial a_0}$ will be zero

$$\frac{\partial e^{a_0}}{\partial a_0} = e^{a_0}$$

Hence

$$\sum_{k=0}^N e^{a_k} = e^{a_i}$$