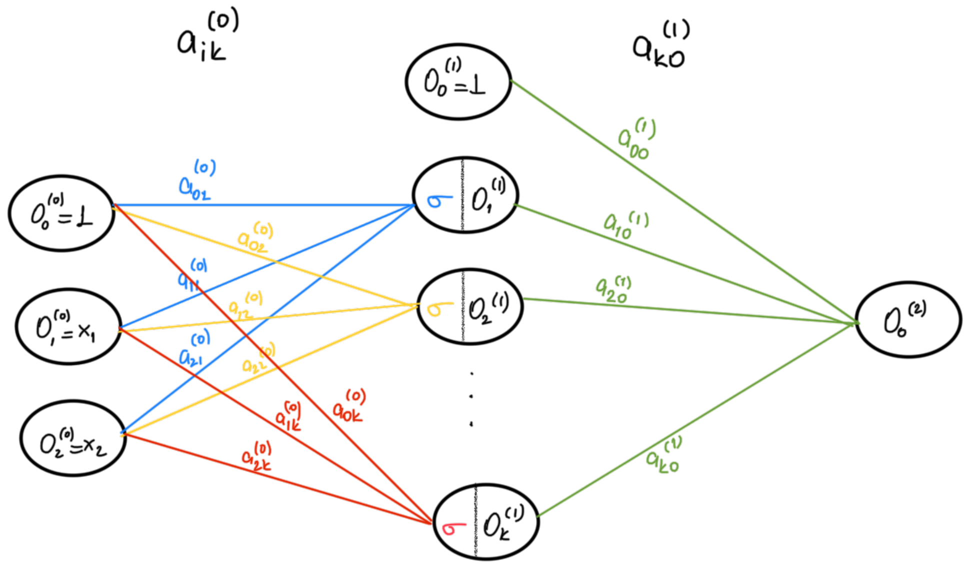
part 1 - regression



$$O_1^{(1)} = \sigma\left(a_{01}^{(0)} + a_{11}^{(0)}\cdot x_1 + a_{21}^{(0)} x_2\right) = O_1^{(1)} = \sigma\left(a_{01}^{(0)} + \sum_i a_{i1}^{(0)} x_i\right)$$

$$O_2^{(1)} = \sigma\left(a_{02}^{(0)} + a_{12}^{(0)} x_1 + a_{22}^{(0)} x_2\right) = O_2^{(1)} = \sigma\left(a_{02}^{(0)} + \sum_i a_{i2}^{(0)} x_i\right)$$

$$\vdots$$

$$O_k^{(1)} = \sigma\left(a_{0k}^{(0)} + a_{1k}^{(0)} x_1 + a_{2k}^{(0)} x_2\right) = \boxed{O_k^{(1)} = \sigma\left(a_{0k}^{(0)} + \sum_i a_{ik}^{(0)} x_i\right)}$$

$$O_0^{(2)} = a_{00}^{(1)} + a_{10}^{(1)}\cdot O_1^{(1)} + a_{20}^{(1)}\cdot O_2^{(1)} + \ldots + a_{k0}^{(1)} O_k^{(1)} = \boxed{a_{00}^{(1)} + \sum_k a_{k0}^{(1)}\cdot O_k^{(1)}}$$

$$O_0^{(2)} = a_{00}^{(1)} + \sum_k a_{k0}^{(1)} \overbrace{\left(\sigma\left(a_{0k}^{(0)} + \sum_i a_{ik}^{(0)} x_i\right)\right)}^{O_k^{(1)}}$$

$$SSE(y, O_0^{(2)}) = (y - O_0^{(2)})^2$$

- $$a_{ik}^{(0)} = a_{ik}^{(0)} - \alpha \frac{\partial SSE(y, O_0^{(2)})}{\partial a_{ik}^{(0)}}$$

$$\frac{\partial SSE(y, O_0^{(2)})}{\partial a_{ik}^{(0)}} = 2\cdot(y - O_0^{(2)})\cdot \frac{\partial (y - O_0^{(2)})}{\partial a_{ik}^{(0)}} \quad \text{(chain rule)}$$

$$= 2\left(O_0^{(2)} - y\right) \frac{\partial O_0^{(2)}}{\partial a_{ik}^{(0)}} \qquad \boxed{\sigma'(x) = \sigma(x)\,(1 - \sigma(x))}$$

$$\boxed{= 2\cdot(O_0^{(2)} - y)\cdot \sum_k a_{k0}^{(1)} \left(\sigma'\left(a_{0k}^{(0)} + \sum_i a_{ik}^{(0)} x_i\right)\cdot \sum_i x_i\right)}$$

$$\frac{\partial SSE(y, O_0^{(2)})}{\partial a_{0k}^{(0)}} = 2\cdot(y - O_0^{(2)})\cdot \frac{\partial (y - O_0^{(2)})}{\partial a_{0k}^{(0)}} \quad \text{(chain rule)}$$

$$= 2\left(O_0^{(2)} - y\right) \frac{\partial O_0^{(2)}}{\partial a_{0k}^{(0)}}$$

$$\boxed{= 2\cdot(O_0^{(2)} - y)\cdot \sum_k a_{k0}^{(1)} \left(\sigma'\left(a_{0k}^{(0)} + \sum_i a_{ik}^{(0)} x_i\right)\right)}$$

↳ for $i > 0$ $\quad a_{ik}^{(0)} = a_{ik}^{(0)} - \alpha . 2 . (O_0^{(2)} - y) . \sum_k a_{k0}^{(1)} \left(\sigma' \left(a_{0k}^{(0)} + \sum_i a_{ik}^{(0)} x_i \right) \sum_i x_i \right)$

$i = 0$ $\quad a_{0k}^{(0)} = a_{0k}^{(0)} - \alpha . 2 . (O_0^{(2)} - y) . \sum_k a_{k0}^{(1)} \left(\sigma' \left(a_{0k}^{(0)} + \sum_i a_{ik}^{(0)} x_i \right) \right)$

- $a_{k0}^{(1)} = a_{k0}^{(1)} - \alpha \dfrac{\partial SSE(y, O_0^{(2)})}{\partial a_{k0}^{(1)}}$

$$\frac{\partial SSE(y, O_0^{(2)})}{\partial a_{k0}^{(1)}} = 2 . (y - O_0^{(2)}) . \frac{\partial (y - O_0^{(2)})}{\partial a_{k0}^{(1)}} \quad \text{(chain rule)}$$

$$= 2 . (O_0^{(2)} - y) . \frac{\partial O_0^{(2)}}{\partial a_{k0}^{(1)}}$$

$$\boxed{= 2 (O_0^{(2)} - y) \left(\sum_k \left(\sigma \left(a_{0k}^{(0)} + \overbrace{\sum_i a_{ik}^{(0)} x_i}^{a_{1k}^{(0)} x_1 + a_{2k}^{(0)} x_2} \right) \right) \right)}$$

$$\frac{\partial SSE(y, O_0^{(2)})}{\partial a_{00}^{(1)}} = 2 . (y - O_0^{(2)}) . \frac{\partial (y - O_0^{(2)})}{\partial a_{00}^{(1)}}$$

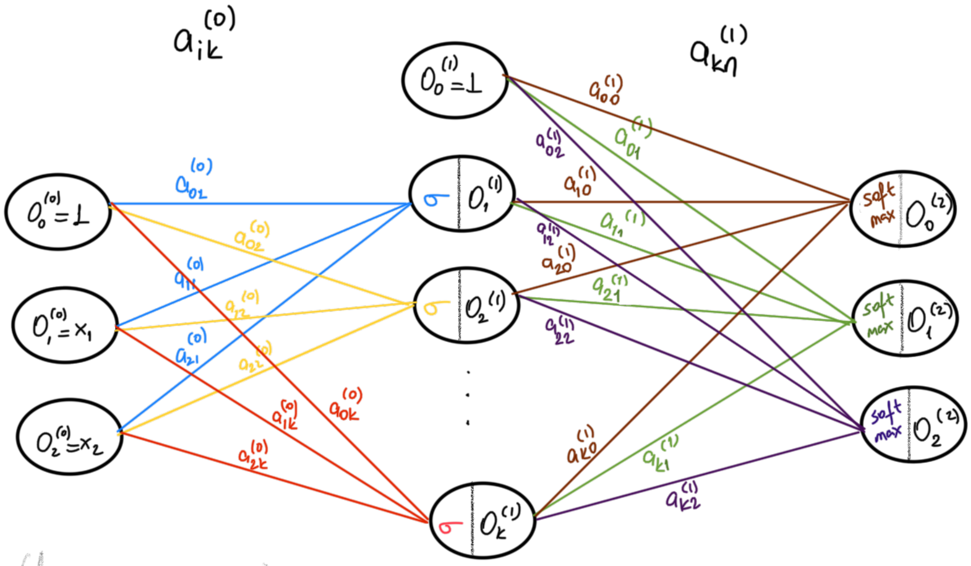
$$= 2 . (O_0^{(2)} - y) . \frac{\partial O_0^{(2)}}{\partial a_{00}^{(1)}}$$

$$\boxed{= 2 (O_0^{(2)} - y) . 1}$$

↳ for $k > 0$ $\quad a_{k0}^{(1)} = a_{k0}^{(1)} - \alpha . 2 . (O_0^{(2)} - y) \sum_k \left(\sigma \left(a_{0k}^{(0)} + a_{1k}^{(0)} x_1 + a_{2k}^{(0)} x_2 \right) \right)$

$k = 0$ $\quad a_{00}^{(1)} = a_{00}^{(1)} - \alpha \, 2 (O_0^{(2)} - y)$

part 1 - classification



(from previous part)

$$\boxed{O_k^{(1)} = \sigma\left(a_{0k}^{(0)} + \sum_i a_{ik}^{(0)} x_i\right)}$$

$$X_0^{(2)} = a_{00}^{(1)} + \sum_k a_{k0}^{(1)} \cdot O_k^{(1)}$$
$$X_1^{(2)} = a_{01}^{(1)} + \sum_k a_{k1}^{(1)} \, O_k^{(1)}$$
$$X_2^{(2)} = a_{02}^{(1)} + \sum_k a_{k2}^{(1)} \, O_k^{(1)}$$

$$\boxed{X_n^{(2)} = a_{0n}^{(1)} + \sum_k a_{kn}^{(1)} \cdot O_k^{(1)}}$$

$$\left.\begin{aligned} O_0^{(2)} &= S(X_0^{(2)}, X) \\ O_1^{(2)} &= S(X_1^{(2)}, X) \\ O_2^{(2)} &= S(X_2^{(2)}, X) \end{aligned}\right\} O_n^{(2)} = S(X_n^{(2)}, X) = \frac{e^{x_n^{(2)}}}{\sum_m e^{x_m^{(2)}}}$$

$$CE(\ell = [\ell_0, \ell_1, \ell_2], \ell' = [O_0^{(2)}, O_0^{(1)}, O_0^{(2)}]) = -\sum_{n=0}^{2} \ell_n \log(O_n^{(2)})$$

- $$a_{ik}^{(0)} = a_{ik}^{(0)} - \alpha \boxed{\frac{CE(\ell, \ell')}{\partial a_{ik}^{(0)}}}$$

$$\frac{\partial CE(\ell,\ell')}{\partial a_{ik}^{(0)}} = \frac{\partial -\sum_{n=0}^{2} \ell_n \log(S(X_n^{(2)}, X))}{\partial a_{ik}^{(0)}} = -\sum_{n=0}^{2} \ell_n \frac{1}{S(X_n^{(2)}, X)} \cdot \boxed{\frac{\partial S(X_n^{(2)}, X)}{\partial a_{ik}^{(0)}}}$$

for $i > 0$:

$$\frac{\partial S(X_n^{(2)}, X)}{\partial a_{ik}^{(0)}} = \frac{\partial \dfrac{e^{x_n^{(2)}}}{\sum_m e^{x_m^{(2)}}}}{\partial a_{ik}^{(0)}}$$

$$\boxed{\sigma'(x) = \sigma(x) \cdot (1 - \sigma(x))}$$

$$= \frac{\frac{\partial x_n^{(2)}}{\partial a_{ik}^{(0)}} \cdot e^{x_n^{(2)}} \cdot \sum_m e^{x_m^{(2)}} - e^{x_n^{(2)}} \cdot \frac{\partial\left(\sum_m e^{x_m^{(2)}}\right)}{\partial a_{ik}^{(0)}}}{\left(\sum e^{x_m^{(2)}}\right)^2}$$

$$= \frac{\sum_k a_{kn}^{(1)} \cdot \boxed{\frac{\partial O_k^{(1)}}{\partial a_{ik}^{(0)}}} \cdot e^{x_n^{(2)}} \sum_m e^{x_m^{(2)}} - e^{x_n^{(2)}} \, \partial\left(\sum_m \left(\sum_k a_{kn}^{(1)} \boxed{\frac{\partial O_k^{(1)}}{\partial a_{ik}^{(0)}}} \cdot e^{x_m^{(2)}}\right)\right)}{\left(\sum e^{x_m^{(2)}}\right)^2}$$

$$\frac{\partial O_k^{(1)}}{\partial a_{ik}^{(0)}} = \sigma'\left(a_{0k}^{(0)} + \sum_i a_{ik}^{(0)} x_i\right) \cdot \sum_i x_i$$

$i=0$:

$$\frac{\partial S(x_n^{(2)},x)}{\partial a_{0k}^{(0)}} = \frac{\partial \dfrac{e^{x_n^{(2)}}}{\sum_m e^{x_m^{(2)}}}}{\partial a_{0k}^{(0)}}$$

$$= \frac{\frac{\partial x_n^{(2)}}{\partial a_{0k}^{(0)}} \cdot e^{x_n^{(2)}} \cdot \sum_m e^{x_m^{(2)}} - e^{x_n^{(2)}} \cdot \frac{\partial\left(\sum_m e^{x_m^{(2)}}\right)}{\partial a_{0k}^{(0)}}}{\left(\sum e^{x_m^{(2)}}\right)^2}$$

$$= \frac{\sum_k a_{kn}^{(1)} \cdot \boxed{\frac{\partial O_k^{(1)}}{\partial a_{0k}^{(0)}}} \cdot e^{x_n^{(2)}} \sum_m e^{x_m^{(2)}} - e^{x_n^{(2)}}\, \partial\left(\sum_m \left(\sum_k a_{kn}^{(1)} \boxed{\frac{\partial O_k^{(1)}}{\partial a_{0k}^{(0)}}} \cdot e^{x_m^{(2)}}\right)\right)}{\left(\sum e^{x_m^{(2)}}\right)^2}$$

$$\frac{\partial O_k^{(1)}}{\partial a_{0k}^{(0)}} = \sigma'\left(a_{0k}^{(0)} + \sum_i a_{ik}^{(0)} x_i\right)$$

- $a_{kn}^{(1)} = a_{kn}^{(1)} - \alpha \boxed{\frac{CE(\ell,\ell')}{\partial a_{kn}^{(1)}}}$

$$\frac{\partial CE(\ell,\ell')}{\partial a_{kn}^{(1)}} = \frac{\partial -\sum_{n=0}^{2} \ell_n \log\left(S(x_n^{(2)},x)\right)}{\partial a_{kn}^{(1)}} = -\sum_{n=0}^{2} \ell_n \frac{1}{S(x_n^{(2)},x)} \cdot \boxed{\frac{\partial S(x_n^{(2)},x)}{\partial a_{kn}^{(1)}}}$$

for $k>0$:

$$\frac{\partial S(x_n^{(2)},x)}{\partial a_{kn}^{(1)}} = \frac{\partial \dfrac{e^{x_n^{(2)}}}{\sum_m e^{x_m^{(2)}}}}{\partial a_{kn}^{(1)}}$$

$$= \frac{\frac{\partial x_n^{(2)}}{\partial a_{kn}^{(1)}} \cdot e^{x_n^{(2)}} \cdot \sum_m e^{x_m^{(2)}} - e^{x_n^{(2)}} \cdot \frac{\partial\left(\sum_m e^{x_m^{(2)}}\right)}{\partial a_{kn}^{(1)}}}{\left(\sum e^{x_m^{(2)}}\right)^2}$$

$$= \frac{\sum_k O_k^{(1)} \cdot e^{x_n^{(2)}} \cdot \sum_m e^{x_m^{(2)}} - e^{x_n^{(2)}} \left(\sum_m \left(\sum_k O_k^{(1)} e^{x_m^{(2)}}\right)\right)}{\left(\sum e^{x_m^{(2)}}\right)^2}$$

$k=0$:

$$\frac{\partial S(x_n^{(2)}, x)}{\partial q_{0n}^{(1)}} = \frac{\partial \dfrac{e^{x_n^{(2)}}}{\sum_m e^{x_m^{(2)}}}}{\partial q_{0n}^{(1)}}$$

$$= \frac{\frac{\partial x_n^{(2)}}{\partial q_{0n}^{(1)}} \cdot e^{x_n^{(2)}} \cdot \sum_m e^{x_m^{(2)}} - e^{x_n^{(2)}} \cdot \frac{\partial \left(\sum_m e^{x_m^{(2)}} \right)}{\partial q_{0n}^{(1)}}}{\left(\sum e^{x_m^{(2)}} \right)^2}$$

$$= \frac{e^{x_n^{(2)}} \cdot \sum_m e^{x_m^{(2)}} - e^{x_n^{(2)}} \left(\sum_m e^{x_m^{(2)}} \right)}{\left(\sum e^{x_m^{(2)}} \right)^2}$$