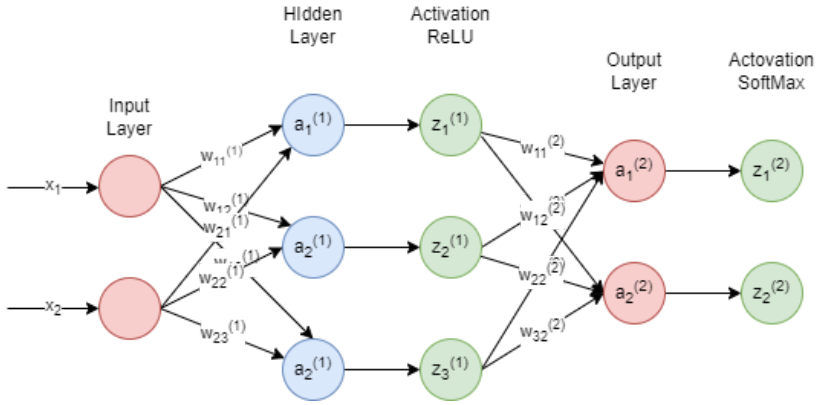
Let’s assume we have a neural network like this:



Loss function: Categorical Cross-Entropy

$$L = \sum_{i=1}^{2} -y_i \ln z_i^{(2)}$$

$$SoftMax: \; f(x) = \frac{e^x}{\sum_{y=c} e^y} \implies z_i^{(2)} = \frac{e^{a_i(2)}}{e^{a_1(2)} + e^{a_2(2)}}$$

$$\frac{dL}{dz_1^{(2)}} = \frac{d}{dz_1^{(2)}}\left[-y_1 \ln z_1^{(2)} - y_2 \ln z_2^{(2)}\right] = -\frac{y_1}{z_1^{(2)}} = \Delta z_1^{(2)}$$

$$\frac{dL}{dz_2^{(2)}} = \frac{d}{dz_2^{(2)}}\left[-y_1 \ln z_1^{(2)} - y_2 \ln z_2^{(2)}\right] = -\frac{y_2}{z_2^{(2)}} = \Delta z_2^{(2)}$$

$$\frac{dL}{da_1^{(2)}} = \frac{dL}{dz_1^{(2)}}\frac{dz_1^{(2)}}{da_1^{(2)}} = \Delta z_1^{(2)}\left[\frac{d}{da_1^{(2)}}\frac{e^{a_1(2)}}{e^{a_1(2)} + e^{a_2(2)}}\right] =$$

$$\Delta z_1^{(2)}\left[\frac{\left(e^{a_1(2)} + e^{a_2(2)}\right)\frac{de^{a_1(2)}}{da_1^{(2)}} - e^{a_1(2)}\frac{d}{da_1^{(2)}}\left(e^{a_1(2)} + e^{a_2(2)}\right)}{\left(e^{a_1(2)} + e^{a_2(2)}\right)^2}\right] =$$

$$\Delta z_1^{(2)}\left[\frac{e^{a_1(2)}\left(e^{a_1(2)} + e^{a_2(2)}\right) - e^{a_1(2)}e^{a_1(2)}}{\left(e^{a_1(2)} + e^{a_2(2)}\right)\left(e^{a_1(2)} + e^{a_2(2)}\right)}\right] =$$

$$\Delta z_1^{(2)}\left[\frac{e^{a_1(2)}}{\left(e^{a_1(2)} + e^{a_2(2)}\right)}\frac{e^{a_1(2)} + e^{a_2(2)} - e^{a_1(2)}}{\left(e^{a_1(2)} + e^{a_2(2)}\right)}\right] = \Delta z_1^{(2)}\left[z_1^{(2)}\left(\frac{\left(e^{a_1(2)} + e^{a_2(2)}\right)}{\left(e^{a_1(2)} + e^{a_2(2)}\right)} - \frac{e^{a_1(2)}}{\left(e^{a_1(2)} + e^{a_2(2)}\right)}\right)\right] =$$

$$\Delta z_1^{(2)}\left(z_1^{(2)}\left(1 - z_1^{(2)}\right)\right) = \Delta a_1^{(2)} \implies \frac{dL}{da_2^{(2)}} = \Delta z_2^{(2)}\left(z_2^{(2)}\left(1 - z_2^{(2)}\right)\right) = \Delta a_2^{(2)}$$

$$\frac{dL}{dw_{11}^{(2)}} = \frac{dL}{dz_1^{(2)}}\frac{dz_1^{(2)}}{da_1^{(2)}}\frac{da_1^{(2)}}{dw_{11}^{(2)}} = \Delta a_1^{(2)}\frac{d}{dw_{11}^{(2)}}\left[w_{11}^{(2)}z_1^{(1)} + w_{21}^{(2)}z_2^{(1)} + w_{31}^{(2)}z_3^{(2)} + b_1^{(2)}\right] = \Delta a_1^{(2)}z_1^{(1)}$$

$$\frac{dL}{dw_{21}^{(2)}} = \frac{dL}{dz_1^{(2)}}\frac{dz_1^{(2)}}{da_1^{(2)}}\frac{da_1^{(2)}}{dw_{21}^{(2)}} = \Delta a_1^{(2)}\frac{d}{dw_{21}^{(2)}}\left[w_{11}^{(2)}z_1^{(1)} + w_{21}^{(2)}z_2^{(1)} + w_{31}^{(2)}z_3^{(2)} + b_1^{(2)}\right] = \Delta a_1^{(2)}z_2^{(1)}$$

$$\frac{dL}{dw_{12}^{(2)}} = \frac{dL}{dz_2^{(2)}}\frac{dz_2^{(2)}}{da_2^{(2)}}\frac{da_2^{(2)}}{dw_{12}^{(2)}} = \Delta a_2^{(2)}\frac{d}{dw_{12}^{(2)}}\left[w_{12}^{(2)}z_1^{(1)} + w_{22}^{(2)}z_2^{(1)} + w_{32}^{(2)}z_3^{(2)} + b_2^{(2)}\right] = \Delta a_2^{(2)}z_1^{(1)}$$

$$\frac{dL}{db_1^{(2)}} = \frac{dL}{dz_1^{(2)}}\frac{dz_1^{(2)}}{da_1^{(2)}}\frac{da_1^{(2)}}{db_1^{(2)}} = \Delta a_1^{(2)}\frac{d}{db_1^{(2)}}\left[w_{11}^{(2)}z_1^{(1)} + w_{21}^{(2)}z_2^{(1)} + w_{31}^{(2)}z_3^{(2)} + b_1^{(2)}\right] = \Delta a_1^{(2)}$$

$$\frac{dL}{db_2^{(2)}} = \frac{dL}{dz_2^{(2)}}\frac{dz_2^{(2)}}{da_2^{(2)}}\frac{da_2^{(2)}}{db_2^{(2)}} = \Delta a_2^{(2)}\frac{d}{db_2^{(2)}}\left[w_{12}^{(2)}z_1^{(1)} + w_{22}^{(2)}z_2^{(1)} + w_{32}^{(2)}z_3^{(2)} + b_2^{(2)}\right] = \Delta a_2^{(2)}$$

$$\frac{dL}{dw_{ij}^{(2)}} = \Delta a_j^{(2)} z_i^{(1)}, \quad \frac{dL}{db_i^{(2)}} = \Delta a_i^{(2)}$$

$$\frac{dL}{dz_1^{(1)}} = \frac{d}{dz_1^{(1)}}\left[-y_1 \ln z_1^{(2)} - y_2 \ln z_2^{(2)}\right] = \frac{d}{dz_1^{(1)}}\left[-y_1 \ln z_1^{(2)}\right] + \frac{d}{dz_1^{(1)}}\left[-y_2 \ln z_2^{(2)}\right] =$$

$$- y_1\frac{d\ln z_1^{(2)}}{dz_1^{(1)}} - y_2\frac{d\ln z_2^{(2)}}{dz_1^{(1)}} = - y_1\left[\frac{d\ln\left(\frac{e^{a_1(2)}}{e^{a_1(2)}+e^{a_2(2)}}\right)}{da_1^{(2)}}\frac{da_1^{(2)}}{dz_1^{(1)}}\right] - y_2\left[\frac{d\ln\left(\frac{e^{a_2(2)}}{e^{a_1(2)}+e^{a_2(2)}}\right)}{da_2^{(2)}}\frac{da_2^{(2)}}{dz_1^{(1)}}\right] =$$

$$- y_1\frac{e^{a_1(2)}+e^{a_2(2)}}{e^{a_1(2)}}\frac{d}{da_1^{(2)}}\left[\frac{e^{a_1(2)}}{e^{a_1(2)}+e^{a_2(2)}}\right]\frac{da_1^{(2)}}{dz_1^{(1)}} - y_2\frac{e^{a_1(2)}+e^{a_2(2)}}{e^{a_2(2)}}\frac{d}{da_2^{(2)}}\left[\frac{e^{a_2(2)}}{e^{a_1(2)}+e^{a_2(2)}}\right]\frac{da_2^{(2)}}{dz_1^{(1)}} =$$

$$- y_1\frac{1}{z_1^{(2)}}\frac{e^{a_1(2)}\left(e^{a_1(2)}+e^{a_2(2)}\right)-e^{a_1(2)}e^{a_1(2)}}{\left(e^{a_1(2)}+e^{a_2(2)}\right)^2}\frac{da_1^{(2)}}{dz_1^{(1)}} - y_2\frac{1}{z_2^{(2)}}\frac{e^{a_2(2)}\left(e^{a_1(2)}+e^{a_2(2)}\right)-e^{a_2(2)}e^{a_2(2)}}{\left(e^{a_1(2)}+e^{a_2(2)}\right)^2}\frac{da_2^{(2)}}{dz_1^{(1)}} =$$

$$- y_1\frac{1}{z_1^{(2)}}z_1^{(2)}\left(1-z_1^{(2)}\right)w_{11}^{(2)} - y_2\frac{1}{z_2^{(2)}}z_2^{(2)}\left(1-z_2^{(2)}\right)w_{12}^{(2)} = \Delta z_1^{(1)}$$

$$\frac{dL}{dz_2^{(1)}} = -y_1 \left(1 - z_1^{(2)}\right) w_{21}^{(2)} - y_2 \left(1 - z_2^{(2)}\right) w_{22}^{(2)} = \Delta z_2^{(1)}$$
$$\frac{dL}{dz_3^{(1)}} = -y_1 \left(1 - z_1^{(2)}\right) w_{31}^{(2)} - y_2 \left(1 - z_2^{(2)}\right) w_{32}^{(2)} = \Delta z_3^{(1)}$$

$$\frac{dL}{da_1^{(1)}} = \frac{dL}{dz_1^{(1)}} \frac{dz_1^{(1)}}{da_1^{(1)}} = \Delta z_1^{(1)} \left(0 \; if \; z_1^{(1)} \leq 0, \; 1 \; otherwise\right) = \Delta a_1^{(1)}$$
$$\frac{dL}{da_2^{(1)}} = \frac{dL}{dz_2^{(1)}} \frac{dz_2^{(1)}}{da_2^{(1)}} = \Delta z_2^{(1)} \left(0 \; if \; z_2^{(1)} \leq 0, \; 1 \; otherwise\right) = \Delta a_2^{(1)}$$
$$\frac{dL}{da_3^{(1)}} = \frac{dL}{dz_3^{(1)}} \frac{dz_3^{(1)}}{da_3^{(1)}} = \Delta z_3^{(1)} \left(0 \; if \; z_3^{(1)} \leq 0, \; 1 \; otherwise\right) = \Delta a_3^{(1)}$$

$$\frac{dL}{dw_{11}^{(1)}} = \frac{dL}{da_1^{(1)}} \frac{da_1^{(1)}}{dw_{11}^{(1)}} = \Delta a_1^{(1)} \frac{d}{dw_{11}^{(1)}} \left[w_{11}^{(1)} x_1 + w_{21}^{(1)} x_2 + b_1^{(1)}\right] = \Delta a_1^{(1)} x_1 = \Delta w_{11}^{(1)}$$
$$\frac{dL}{dw_{12}^{(1)}} = \frac{dL}{da_2^{(1)}} \frac{da_2^{(1)}}{dw_{12}^{(1)}} = \Delta a_2^{(1)} \frac{d}{dw_{12}^{(1)}} \left[w_{12}^{(1)} x_1 + w_{22}^{(1)} x_2 + b_2^{(1)}\right] = \Delta a_2^{(1)} x_1 = \Delta w_{12}^{(1)}$$
$$\frac{dL}{dw_{21}^{(1)}} = \frac{dL}{da_1^{(1)}} \frac{da_1^{(1)}}{dw_{21}^{(1)}} = \Delta a_1^{(1)} \frac{d}{dw_{21}^{(1)}} \left[w_{11}^{(1)} x_1 + w_{21}^{(1)} x_2 + b_1^{(1)}\right] = \Delta a_1^{(1)} x_2 = \Delta w_{21}^{(1)}$$

$$\frac{dL}{dw_{ij}^{(1)}} = \Delta a_j^{(1)} x_i$$

$$\frac{dL}{db_1^{(1)}} = \frac{dL}{da_1^{(1)}}\frac{da_1^{(1)}}{db_1^{(1)}} = \Delta a_1^{(1)}\frac{d}{db_1^{(1)}}\left[w_{11}^{(1)}x_1 + w_{12}^{(1)}x_2 \ + b_1^{(1)}\right] = \Delta a_1^{(1)} = \Delta b_1^{(1)}$$

$$\frac{dL}{db_2^{(1)}} = \frac{dL}{da_2^{(1)}}\frac{da_2^{(1)}}{db_2^{(1)}} = \Delta a_2^{(1)}\frac{d}{db_2^{(1)}}\left[w_{12}^{(1)}x_1 \ + w_{22}^{(1)}x_2 + b_2^{(1)}\right] = \Delta a_2^{(1)} = \Delta b_2^{(1)}$$

$$\frac{dL}{db_i^{(1)}} = \Delta a_i^{(1)}$$