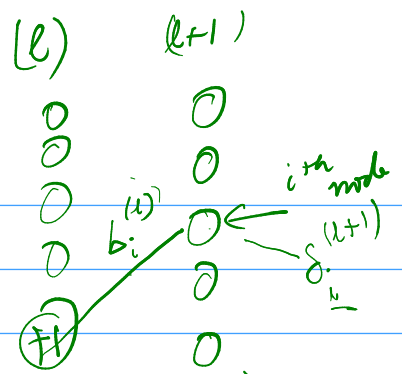


$$\frac{\partial J}{\partial b_i^{(l)}} = \frac{\partial J}{\partial z_i^{(l+1)}} \frac{\partial z_i^{(l+1)}}{\partial b_i^{(l)}} = 1$$

$\delta_i^{(l+1)}$

$$\boxed{\frac{\partial J}{\partial b_i^{(l)}} = \delta_i^{(l+1)}}$$

$$\frac{\partial J}{\partial \underline{b}} = \begin{pmatrix} \frac{\partial J}{\partial b_1} \\ \frac{\partial J}{\partial b_2} \\ \vdots \end{pmatrix} = \begin{pmatrix} \delta_1^{(l+1)} \\ \delta_2^{(l+1)} \\ \vdots \end{pmatrix} = \underline{\delta}^{(l+1)}$$



$$z_i^{(l+1)} = W^{(l)} a^{(l)} + b^{(l)}$$

$$z_i^{(l+1)} = \sum_j W_{ij}^{(l)} a_j^{(l)} + b_i^{(l)}$$

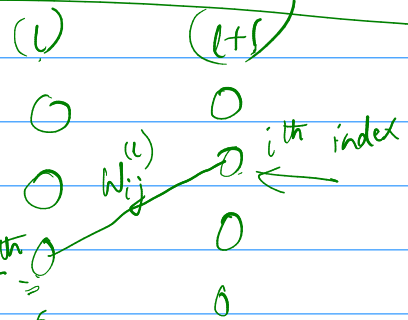
$$\frac{\partial z_i^{(l+1)}}{\partial b_i^{(l)}} = 1$$

$$\frac{\partial J}{\partial W_{ij}^{(l)}} = \frac{\partial J}{\partial z_i^{(l+1)}} \frac{\partial z_i^{(l+1)}}{\partial W_{ij}^{(l)}}$$

$$\boxed{\frac{\partial J}{\partial W_{ij}^{(l)}} = \delta_i^{(l+1)} a_j^{(l)}}$$

$$\frac{\partial J}{\partial \underline{W}^{(l)}} = \begin{pmatrix} \frac{\partial J}{\partial W_{11}^{(l)}} & \frac{\partial J}{\partial W_{12}^{(l)}} & \dots \\ \frac{\partial J}{\partial W_{21}^{(l)}} & \frac{\partial J}{\partial W_{22}^{(l)}} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} \delta_1^{(l+1)} a_1^{(l)} & \delta_1^{(l+1)} a_2^{(l)} & \dots \\ \delta_2^{(l+1)} a_1^{(l)} & \delta_2^{(l+1)} a_2^{(l)} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$$\boxed{\frac{\partial J}{\partial \underline{W}^{(l)}} = \underline{\delta}^{(l+1)} (\underline{a}^{(l)})^T}$$



$$z_i^{(l+1)} = W^{(l)} a^{(l)} + b^{(l)}$$

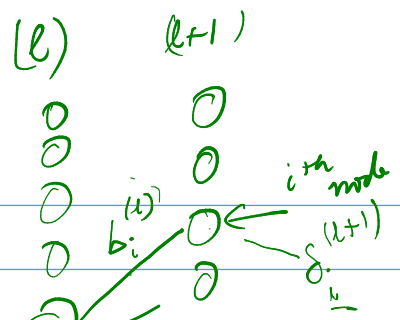
$$z_i^{(l+1)} = \sum_j W_{ij}^{(l)} a_j^{(l)} + b_i^{(l)}$$

$$\frac{\partial J}{\partial \underline{W}^{(l)}} = \begin{pmatrix} \delta_1^{(l+1)} a_1^{(l)} & \delta_1^{(l+1)} a_2^{(l)} & \dots \\ \delta_2^{(l+1)} a_1^{(l)} & \delta_2^{(l+1)} a_2^{(l)} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} = \underline{\delta}^{(l+1)} (\underline{a}^{(l)})^T$$

$$\frac{\partial J}{\partial b_i^{(l)}} = \frac{\partial J}{\partial z_i^{(l+1)}} \frac{\partial z_i^{(l+1)}}{\partial b_i^{(l)}} = 1$$

$$\frac{\partial J}{\partial b_i^{(l)}} = \delta_i^{(l+1)}$$

$$\frac{\partial J}{\partial \underline{b}} = \begin{pmatrix} \frac{\partial J}{\partial b_1} \\ \frac{\partial J}{\partial b_2} \\ \vdots \end{pmatrix} = \begin{pmatrix} \delta_1^{(l+1)} \\ \delta_2^{(l+1)} \\ \vdots \end{pmatrix} = \underline{\delta}^{(l+1)}$$



$$z_i^{(l+1)} = W^{(l)} a^{(l)} + b^{(l)}$$

$$z_i^{(l+1)} = \sum_j W_{ij}^{(l)} a_j^{(l)} + b_i^{(l)}$$

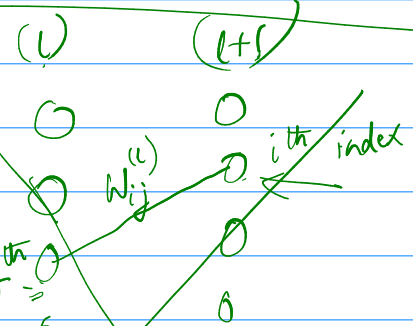
$$\frac{\partial z_i^{(l+1)}}{\partial b_i^{(l)}} = 1$$

$$\frac{\partial J}{\partial W_{ij}^{(l)}} = \frac{\partial J}{\partial z_i^{(l+1)}} \frac{\partial z_i^{(l+1)}}{\partial W_{ij}^{(l)}}$$

$$\frac{\partial J}{\partial W_{ij}^{(l)}} = \delta_i^{(l+1)} a_j^{(l)}$$

$$\frac{\partial J}{\partial \underline{W}^{(l)}} = \begin{pmatrix} \frac{\partial J}{\partial W_{11}^{(l)}} & \frac{\partial J}{\partial W_{12}^{(l)}} & \dots \\ \frac{\partial J}{\partial W_{21}^{(l)}} & \frac{\partial J}{\partial W_{22}^{(l)}} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} \delta_1^{(l+1)} a_1^{(l)} & \delta_1^{(l+1)} a_2^{(l)} & \dots \\ \delta_2^{(l+1)} a_1^{(l)} & \delta_2^{(l+1)} a_2^{(l)} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$$\frac{\partial J}{\partial \underline{W}^{(l)}} = \underline{\delta}^{(l+1)} (\underline{a}^{(l)})^T$$



$$z_i^{(l+1)} = W^{(l)} a^{(l)} + b^{(l)}$$

$$z_i^{(l+1)} = \sum_j W_{ij}^{(l)} a_j^{(l)} + b_i^{(l)}$$

$$\frac{\partial J}{\partial \underline{W}^{(l)}} = \underline{\delta}^{(l+1)} (\underline{a}^{(l)})^T$$

$$\delta_i^{(L)} = \left(\sum_j \frac{\partial J^{(L+1)}}{\partial z_j} \frac{\partial z_j^{(L)}}{\partial a_i} \right) \frac{\partial a_i^{(L)}}{\partial z_i} = f'(z_i^{(L)})$$

The diagram illustrates a single layer of a neural network. On the left, there are input nodes labeled (l) . These are connected to hidden nodes labeled $(l+1)$. The connections are labeled with weights w_{ij}^l . From the hidden nodes, connections lead to output nodes labeled $(l+1')$. The connections are labeled with weights w_{ij}^{l+1} . A bias term b_j^{l+1} is shown for the hidden nodes. An activation function σ is indicated for the output nodes. The diagram also shows a bias term b_i^l for the input nodes.

$$z_j^{(l+1)} = \sum_i w_{ji}^{(l)} a_i^{(l)} + b_j^{(l)}$$

$$\underline{\delta}^{(e)} = \left(W^{(e)T} \underline{\delta}^{(l+1)} \right) \odot f'(\underline{z}^{(e)})$$