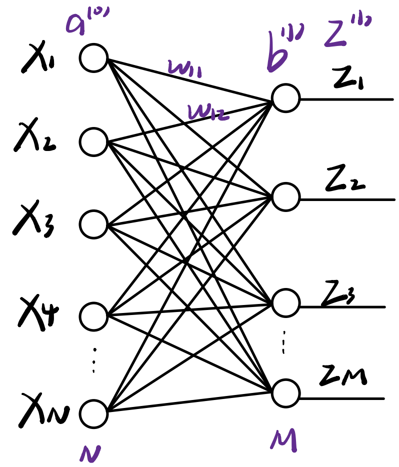
$$X = a^{(0)} \Rightarrow W^{(1)}X + b^{(1)} = Z^{(1)}$$

$$\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ \vdots & & & \vdots \\ w_{N1} & \cdots & \cdots & w_{NN} \end{bmatrix}_{M\times N}$$



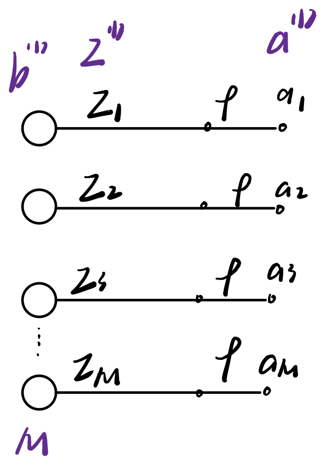

$X \in \mathbb{R}^N$

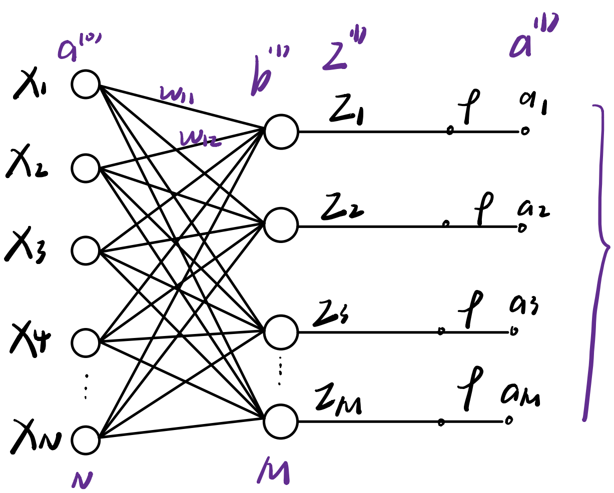
$W^{(1)} \in \mathbb{R}^{M\times N}$

$b^{(1)} \in \mathbb{R}^M$

$a^{(1)} \in \mathbb{R}^M$

$$Z^{(1)} \xrightarrow{f} a^{(1)} = f(Z^{(1)})$$




作为第二层的输入

即: 第一层 第二层

$$X \Rightarrow W^{(1)}X+b^{(1)}=Z^{(1)} \xrightarrow{f} a^{(1)}=f(Z^{(1)}) \to Z^{(2)}=W^{(2)}a^{(1)}+b^{(2)} \xrightarrow{f} a^{(2)}=f(Z^{(2)})$$

$$\to Z^{(3)}=W^{(3)}a^{(2)}+b^{(3)} \cdots \quad Z^{(m)}=W^{(m)}a^{(m-1)}+b^{(m)} \cdots$$

$$\to Z^{(L)}=W^{(L)}a^{(L-1)}+b^{(L)} \Rightarrow y=a^{(L)}=f(Z^{(L)})$$

定义:

① 网络共 $L$ 层

② $Z^{(k)}$ $a^{(k)}$ $b^{(k)}$ 是第k层的向量 与第k层神经元个数一致

③ $Z_i^{(k)}$ $a_i^{(k)}$ $b_i^{(k)}$ 表示 $Z_i^{(k)}$ $a_i^{(k)}$ $b_i^{(k)}$ 的第i个分量

④ 用 $y_i$ 表示 $y$ 的第i个分量

BP算法:

① 随机初始化 $W, b$

② 训练样本 $(X, Y)$ 代入网络 可求出

所有的 $(Z, a, y)$ 前向传播

③ 链式求偏导

最小化: $E=\frac{1}{2}\|y-Y\|^2=\frac{1}{2}\sum_{i=1}^{M}(y_i-Y_i)^2$ MSE

求: $\frac{\partial E}{\partial W}$ $\frac{\partial E}{\partial b}$

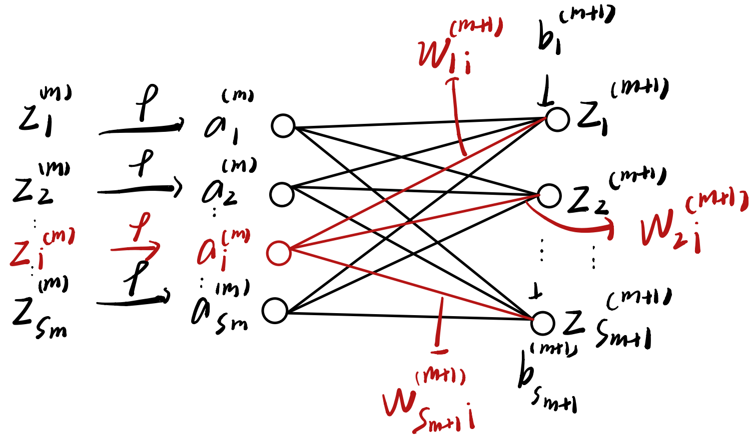
设 $\delta_i^{(m)}=\frac{\partial E}{\partial Z_i^{(m)}}$

① $\delta_i^{(L)}=\frac{\partial E}{\partial Z_i^{(L)}}=\frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial Z_i^{(L)}}=(y_i-Y_i)f'(Z_i^{(L)})$

② $\delta_i^{(m)}=\frac{\partial E}{\partial Z_i^{(m)}}=\frac{\partial E}{\partial a_i^{m}}\cdot\frac{\partial a_i^{m}}{\partial Z_i^{m}}=f'(Z_i^{(m)})\sum_{j=1}^{S_{m+1}}\delta_j^{(m+1)}\cdot W_{ji}$

$1 \le m \le (L-1)$

$$\left(\frac{\partial E}{\partial z_j^{(m+1)}}\right)$$




③ $$\frac{\partial E}{\partial w_{ij}^{(m)}} = \delta_j^{(m)} a_i^{(m-1)}$$

④ $$\frac{\partial E}{\partial b_i^{(m)}} = \delta_i^{(m)}$$

④ 更新:

$$W^{(new)} = W^{(old)} - \alpha \frac{\partial E}{\partial W}\bigg|_{W^{(old)}}$$

$$b^{(new)} = b^{(old)} - \alpha \frac{\partial E}{\partial b}\bigg|_{b^{(old)}}$$