Notes on Artificial Neural Networks

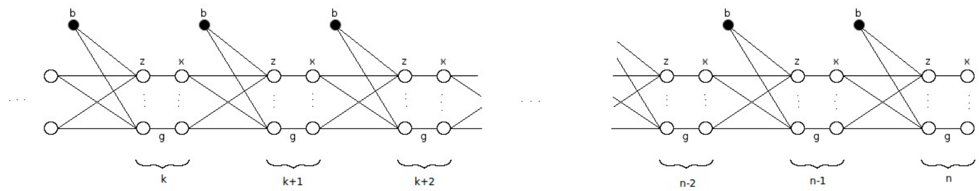

Figure 1. ANN

$$z_q^{k+1} = \sum_{i=0}^{N^k} x_i^k w_{i,q}^k + v_q^k b$$

$$E = \frac{1}{2}\sum_{i=0}^{N}(x_i^n - t_i)^2$$

$$\frac{\partial E}{\partial \omega_{p,q}^k} = \sum_{i_n=0}^{N^n}\Bigg(\frac{\partial E}{\partial x_{i_n}^n}\frac{\partial x_{i_n}^n}{\partial g}\frac{\partial g}{\partial z_{i_n}^n}\sum_{i_{n-1}=0}^{N^{n-1}}\Bigg(\frac{\partial z_{i_n}^n}{\partial x_{i_{n-1}}^{n-1}}\frac{\partial x_{i_{n-1}}^{n-1}}{\partial g}\frac{\partial g}{\partial z_{i_{n-1}}^{n-1}}\Bigg(\sum_{i_{n-2}=0}^{N^{n-2}}\frac{\partial z_{i_{n-1}}^{n-1}}{\partial x_{i_{n-2}}^{n-2}}\frac{\partial x_{i_{n-2}}^{n-2}}{\partial g}\frac{\partial g}{\partial z_{i_{n-2}}^{n-2}}\sum_{i_{n-3}=0}^{N^{n-3}}\Bigg(\cdots$$

$$\cdots\sum_{i_{k+2}=0}^{N^{k+2}}\Bigg(\frac{\partial z_{i_{k+3}}^{k+3}}{\partial x_{i_{k+2}}^{k+2}}\frac{\partial x_{i_{k+2}}^{k+2}}{\partial g}\frac{\partial g}{\partial z_{k+2}^{k+2}}\Bigg(\frac{\partial z_{i_{k+2}}^{k+2}}{\partial x_p^{k+1}}\frac{\partial x_p^{k+1}}{\partial g}\frac{\partial g}{\partial z_p^{k+1}}\Bigg(\frac{\partial z_p^{k+1}}{\partial \omega_{p,q}^k}\Bigg)\Bigg)\Bigg)\Bigg)\cdots\Bigg)\Bigg)\Bigg)\Bigg)$$

And the bias weights v come out with the same process:

$$\frac{\partial E}{\partial v_q^k} = \sum_{i_n=0}^{N^n}\Bigg(\frac{\partial E}{\partial x_{i_n}^n}\frac{\partial x_{i_n}^n}{\partial g}\frac{\partial g}{\partial z_{i_n}^n}\sum_{i_{n-1}=0}^{N^{n-1}}\Bigg(\frac{\partial z_{i_n}^n}{\partial x_{i_{n-1}}^{n-1}}\frac{\partial x_{i_{n-1}}^{n-1}}{\partial g}\frac{\partial g}{\partial z_{i_{n-1}}^{n-1}}\Bigg(\sum_{i_{n-2}=0}^{N^{n-2}}\frac{\partial z_{i_{n-1}}^{n-1}}{\partial x_{i_{n-2}}^{n-2}}\frac{\partial x_{i_{n-2}}^{n-2}}{\partial g}\frac{\partial g}{\partial z_{i_{n-2}}^{n-2}}\sum_{i_{n-3}=0}^{N^{n-3}}\Bigg(\cdots$$

$$\cdots\sum_{i_{k+2}=0}^{N^{k+2}}\Bigg(\frac{\partial z_{i_{k+3}}^{k+3}}{\partial x_{i_{k+2}}^{k+2}}\frac{\partial x_{i_{k+2}}^{k+2}}{\partial g}\frac{\partial g}{\partial z_{k+2}^{k+2}}\Bigg(\frac{\partial z_{i_{k+2}}^{k+2}}{\partial x_p^{k+1}}\frac{\partial x_p^{k+1}}{\partial g}\frac{\partial g}{\partial z_p^{k+1}}\Bigg(\frac{\partial z_p^{k+1}}{\partial v_p^k}\Bigg)\Bigg)\Bigg)\Bigg)\cdots\Bigg)\Bigg)\Bigg)\Bigg)$$

$$\frac{\partial E}{\partial x_i} = (x_i^n - t_i)$$

$$\frac{\partial z_q^{k+1}}{\partial \omega_{p,q}^k} = x_p^k$$

$$\frac{\partial z_q^{k+1}}{\partial v_q^k} = b$$

$$\frac{\partial z_{i_n}^n}{\partial x_{i_{n-1}}^{n-1}} = \omega_{i_n,i_{n-1}}^{n-1}$$

$$\frac{\partial x}{g} = 1$$

$$x^k = g(z^k;\beta) = \frac{1}{1+exp(-\beta z^k)}$$

$$D_g(z^k;\beta) = \frac{\partial g}{\partial z^k} = \beta \frac{\exp(-\beta z^k)}{(1+\exp(-\beta z^k))^2} = \beta g(z^k;\beta)\big(1-g(z^k;\beta)\big) = \beta x^k(1-x^k)$$

Then making these substitutions:

$$\frac{\partial E}{\partial \omega^k_{p,q}} = \sum_{i_{k+2}=0}^{N^{k+2}} \left((x^n_i - t_i) D_g(z^n_{i_n};\beta) \sum_{i_{n-1}=0}^{N^{n-1}} \left(\omega^{n-1}_{i_n,i_{n-1}} D_g(z^{n-1}_{i_{n-1}};\beta) \sum_{i_{n-2}=0}^{N^{n-2}} \left(\omega^{n-2}_{i_{n-1},i_{n-2}} D_g(z^{n-2}_{i_{n-2}};\beta) \sum_{i_{n-3}=0}^{N^{n-3}} \left(\ldots \right.\right.\right.\right.$$

$$\left.\left.\left.\left. \ldots \sum_{i_{k+2}=0}^{N^{k+2}} \left(\omega^{k+2}_{i_{k+3},i_{k+2}} D_g(z^{k+2}_{i_{k+2}};\beta) \left(\omega^{k+1}_{i_{k+2},p} D_g(z^{k+1}_p;\beta) x^k_q \right)\right)\right)\right) \ldots \right)\right)\right)\right)$$

(1)

$$\frac{\partial E}{\partial \omega^k_{p,q}} = \sum_{i_{k+2}=0}^{N^{k+2}} \left((x^n_i - t_i) \beta x^n_{i_n}(1-x^n_{i_n}) \sum_{i_{n-1}=0}^{N^{n-1}} \left(\omega^{n-1}_{i_n,i_{n-1}} \beta x^{n-1}_{i_{n-1}}(1-x^{n-1}_{i_{n-1}}) \sum_{i_{n-2}=0}^{N^{n-2}} \left(\omega^{n-2}_{i_{n-1},i_{n-2}} \beta x^{n-2}_{i_{n-2}}(1-x^{n-2}_{i_{n-2}}) \right.\right.\right.$$

$$\left.\left.\left. \ldots \sum_{i_{k+2}=0}^{N^{k+2}} \left(\omega^{k+2}_{i_{k+3},i_{k+2}} \beta x^{k+2}_{i_{k+2}}(1-x^{k+2}_{i_{k+2}}) \left(\omega^{k+1}_{i_{k+2},p} \beta x^{k+1}_p (1-x^{k+1}_p) x^k_q \right)\right)\right)\right) \ldots \right)\right)\right)\right)$$

The weights are updated as follows:

$$\omega^k_{p,q} - \eta \frac{\partial E}{\partial \omega^k_{p,q}} \to \omega^k_{p,q}$$

where η is the learning rate. In the case of three layers, either three layers deep in the network or the network is only three layers in length,

$$\frac{\partial E}{\partial \omega^{n-2}_{p,q}} = \sum_{i_n=0}^{N^n} \left((x^n_{i_n} - t_{i_n}) D_g(z^n_{i_n};\beta) \left(\omega^{n-1}_{i_n,i_p} D_g(z^{n-1}_p;\beta) x^{n-2}_q \right)\right)$$

For only two layers, either the network is two layers deep or we are only looking at the set of weights at the end of the network:

$$\frac{\partial E}{\partial w^{n-1}_{p,q}} = (x^n_p - t_p) D_g(z^n_p;\beta) x^{n-1}_q.$$

Define

$$
(2)\quad
\begin{aligned}
A^k &:= D_g(z^{k+1};\beta)\otimes x^k = \big(\beta x^{k+1}\circ(1-x^{k+1})\big)\otimes x^k\\
\varpi^k_{i,j} &:= \omega^k_{i,j}D_g(z^k_j) = \omega^k_{i,j}\big(\beta x^k_j(1-x^k_j))\big)\\
W^k &:= \varpi^{n-1}\varpi^{n-2}\cdots\varpi^k \Rightarrow W^k = W^{k+1}\varpi^k\\
\Delta &:= (x^n-t)\circ D_g(z^n) = (x^n-t)\circ\big(\beta x^n\circ(1-x^n)\big)
\end{aligned}
$$

Then (1) becomes

$$
(3)\quad
\begin{aligned}
\frac{\partial E}{\partial\omega^k_{p,q}} = \sum_{i_n=0}^{N^n}\Bigg(\Delta_{i_n}\sum_{i_{n-1}=0}^{N^{n-1}}\Bigg(\varpi^{n-1}_{i_n,i_{n-1}}\sum_{i_{n-2}=0}^{N^{n-2}}\Bigg(\varpi^{n-2}_{i_{n-1},i_{n-2}}\sum_{i_{n-3}=0}^{N^{n-3}}\Bigg(\cdots\\
\sum_{i_{k+2}=0}^{N^{k+2}}\Bigg(\varpi^{k+2}_{i_{k+3},i_{k+2}}\Bigg(\omega^{k+1}_{i_{k+2},p}a^{k+1}_{p,q}\Bigg)\Bigg)\Bigg)\Bigg)\cdots\Bigg)\Bigg)\Bigg)\Bigg).
\end{aligned}
$$

re-writing (3) using (2) and matrix notation:

$$\frac{\partial E}{\partial\omega^k} = \left[\begin{pmatrix}\Delta_1 & \cdots & \Delta_{N^n}\\ \vdots & \vdots & \vdots\\ \Delta_1 & \cdots & \Delta_{N^n}\end{pmatrix}W^{k+2}\omega^{k+1}\right]^T\circ A^k$$

Then

$$\omega^k - \eta\frac{\partial E}{\partial\omega^k}\to\omega^k$$

and

$$\frac{\partial E}{\partial v^k} = \left[\begin{pmatrix}\Delta_1 & \cdots & \Delta_{N^n}\\ \vdots & \vdots & \vdots\\ \Delta_1 & \cdots & \Delta_{N^n}\end{pmatrix}W^{k+2}\omega^{k+1}\right]^T\circ\big(x^{k+1}\circ(1-x^{k+1})\big)b$$

Then

$$v^k - \eta\frac{\partial E}{\partial v^k}\to v^k$$

For 3 layers:

$$\frac{\partial E}{\partial\omega^{n-2}} = \left[\begin{pmatrix}\Delta_1 & \cdots & \Delta_{N^n}\\ \vdots & \vdots & \vdots\\ \Delta_1 & \cdots & \Delta_{N^n}\end{pmatrix}\omega^{n-1}\right]^T\circ A^{n-2}$$

And for 2 layers:

$$\frac{\partial E}{\partial\omega^{n-1}} = \begin{pmatrix}\Delta_1 & \cdots & \Delta_{N^n}\\ \vdots & \vdots & \vdots\\ \Delta_1 & \cdots & \Delta_{N^n}\end{pmatrix}^T\circ A^{n-1}$$