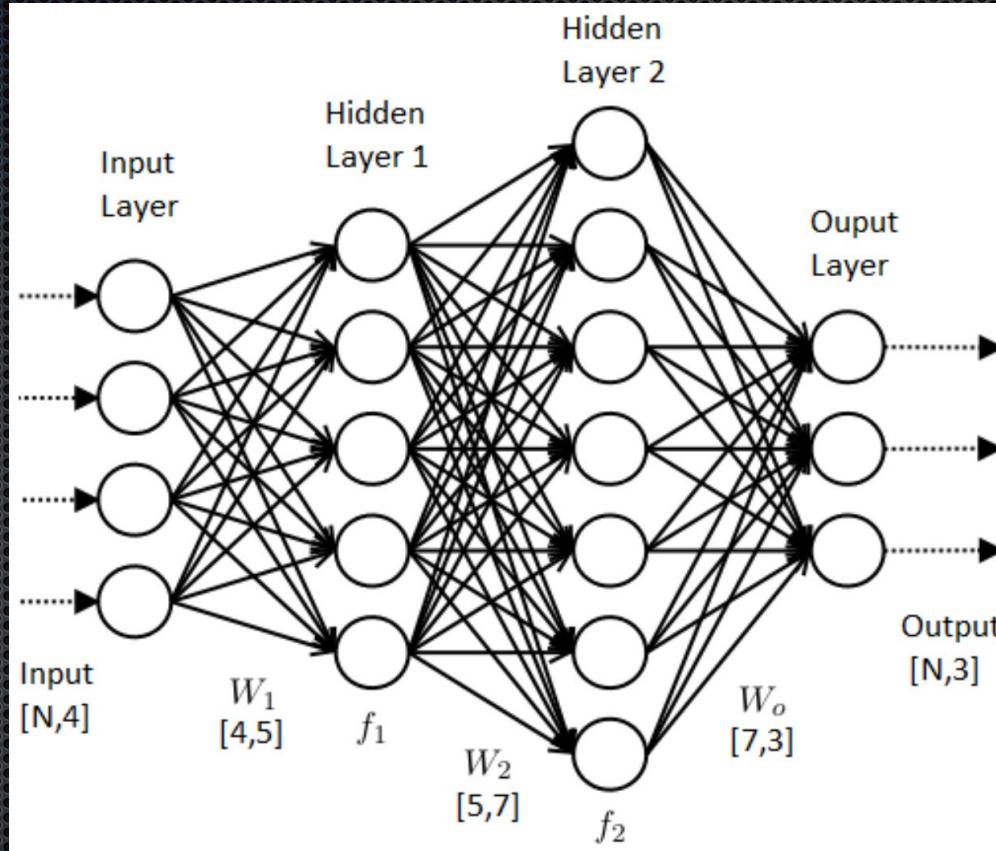
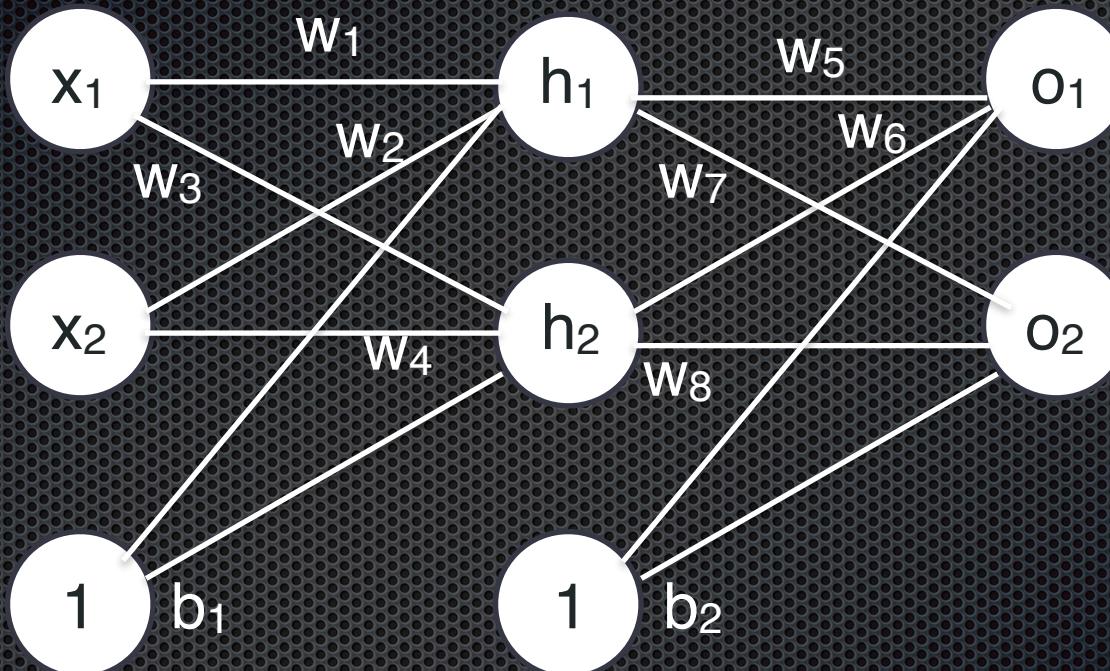


Multilayer Perceptrons

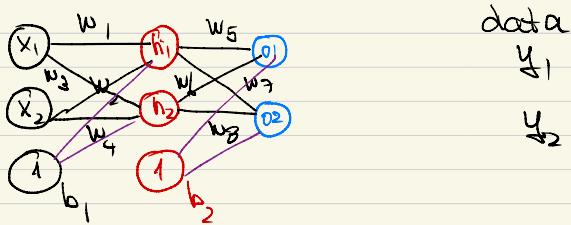


Example network with simple cost function



$$J(\mathbf{w}) = \frac{1}{2}((y_1 - o_1)^2 + (y_2 - o_2)^2)$$

Example of back propagation



Hidden: $z_1 = w_1 x_1 + w_2 x_2 + b_1$

$$h_1 = \phi(z_1) = \frac{1}{1 + e^{-z_1}}$$

$$z_2 = w_3 x_1 + w_4 x_2 + b_2$$

$$h_2 = \phi(z_2) = \frac{1}{1 + e^{-z_2}}$$

Output: $z_3 = w_5 h_1 + w_6 h_2 + b_3$

$$o_1 = \phi(z_3)$$

$$z_4 = w_7 h_1 + w_8 h_2 + b_4$$

$$o_2 = \phi(z_4)$$

$$J(\underline{w}) = \frac{1}{2} \left[\left(y_1 - o_1 \right)^2 + \left(y_2 - o_2 \right)^2 \right] \quad \boxed{9}$$

$$\underline{w} = (w_1, \dots, w_8, b_1, b_2)^T$$

i) Consider first the weights of links connecting the hidden and the output layer, e.g. w_5 .

$$\frac{\partial J}{\partial w_5} = \frac{\partial J}{\partial o_1} \frac{\partial o_1}{\partial z_3} \frac{\partial z_3}{\partial w_5}$$

$$\frac{\partial J}{\partial o_1} = -(y_1 - o_1)$$

$$\begin{aligned} \frac{\partial o_1}{\partial z_3} &= \phi'(z_3) = \phi(z_3)(1 - \phi(z_3)) \\ &= o_1(1 - o_1) \end{aligned}$$

$$\frac{\partial z_3}{\partial w_5} = h_1$$

10

hence

$$\frac{\partial \hat{E}}{\partial w_5} = - (y_1 - o_1) o_1 (1 - o_1) h_1 \\ = \delta_{o_1} h_1$$

Hence, $\overset{1}{w}_5 = w_5 - 1) \frac{\partial \hat{E}}{\partial w_5}$ update

The same can be done for w_6, w_7, w_8 .

2) Now consider the update for w_1 :

$$\frac{\partial \hat{E}}{\partial w_1} = \frac{\partial \hat{E}}{\partial h_1} \frac{\partial h_1}{\partial z_1} \frac{\partial z_1}{\partial w_1}$$

$$\frac{\partial \hat{E}}{\partial h_1} = \frac{\partial \hat{E}}{\partial o_1} \frac{\partial o_1}{\partial h_1} + \frac{\partial \hat{E}}{\partial o_2} \frac{\partial o_2}{\partial h_1}$$

$$\frac{\partial o_1}{\partial z_1} = h_1 (1 - h_1)$$

$$\frac{\partial z_1}{\partial w_1} = x_1$$

$$\text{and } \frac{\partial o_1}{\partial h_1} = o_1(1-o_1) w_5$$

II

$$\frac{\partial o_2}{\partial h_1} = o_2(1-o_2) w_7$$

hence we can write:

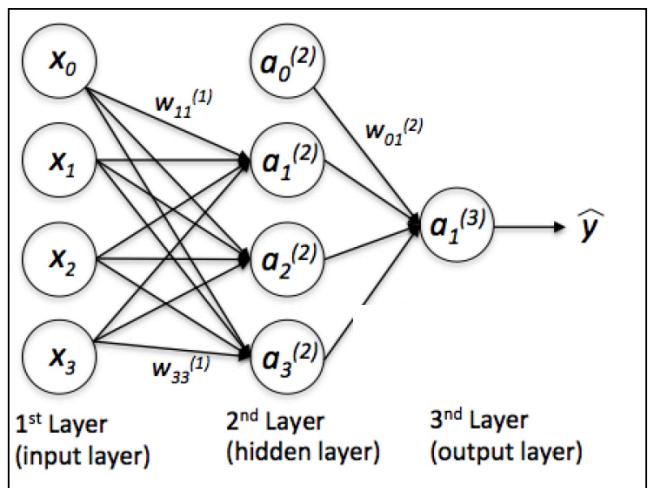
$$\frac{\partial E}{\partial w_1} = \bar{o}_{h_1} \cdot x_1 , \text{ where}$$

$$\begin{aligned} \bar{o}_{h_1} = & - \left[(y_1 - o_1) o_1(1-o_1) w_5 + \right. \\ & \left. (y_2 - o_2) o_2(1-o_2) w_7 \right] * \\ & * h_1 / (1-h_1) \end{aligned}$$

and this can also be done for

the other weights w_2, w_3 and w_4 .

General



$$a^{(1)} = \begin{bmatrix} a_0^{(1)} \\ a_1^{(1)} \\ \vdots \\ a_m^{(1)} \end{bmatrix} = \begin{bmatrix} 1 \\ x_1^{(1)} \\ \vdots \\ x_m^{(1)} \end{bmatrix}$$

$$z_1^{(2)} = a_0^{(1)} w_{1,0}^{(1)} + a_1^{(1)} w_{1,1}^{(1)} + \cdots + a_m^{(1)} w_{1,m}^{(1)}$$

$$a_1^{(2)} = \phi(z_1^{(2)})$$

$w_{j,k}^{(l)}$ weight connecting
unit j, level l+1 to
unit k, level l

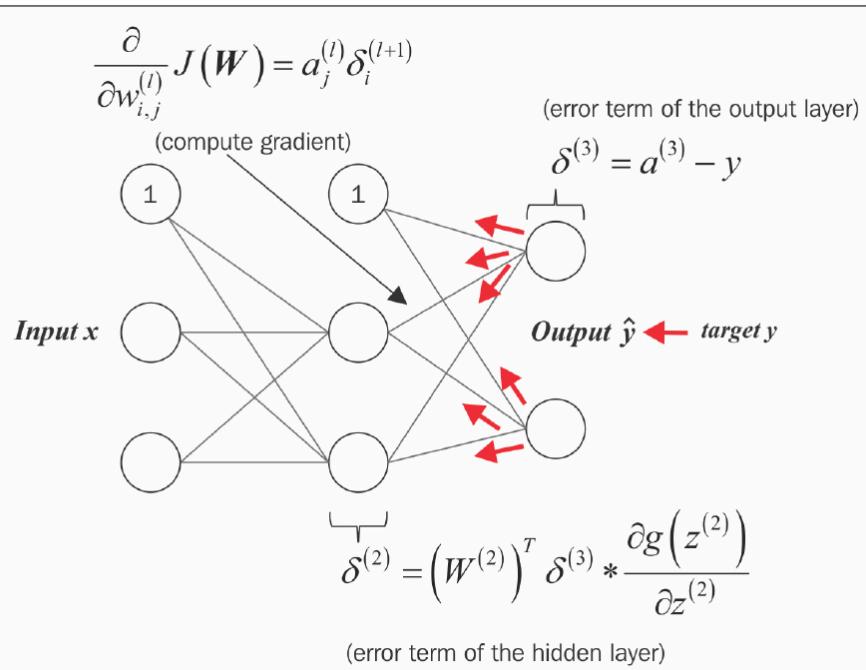
$$\phi(z) = \frac{1}{1 + e^{-z}}$$

Cost function

$$J(\mathbf{w}) = - \left[\sum_{i=1}^n \sum_{j=1}^m y_j^{(i)} \log \left(\phi \left(z_j^{(i)} \right) \right) + \left(1 - y_j^{(i)} \right) \log \left(1 - \phi \left(z_j^{(i)} \right) \right) \right] \\ + \frac{\lambda}{2} \sum_{l=1}^{L-1} \sum_{i=1}^{u_l} \sum_{j=1}^{u_{l+1}} \left(w_{j,i}^{(l)} \right)^2$$

Task: $\min_{\mathbf{w}} J(\mathbf{w})$

General: back propagation



$$\Delta^{(l)} = \Delta^{(l)} + \delta^{(l+1)} (A^{(l)})^T$$

$$\mathbf{W}_{k+1}^{(l)} = \mathbf{W}_k^{(l)} - \eta \Delta^{(l)}$$

Exercise B4

