Neural Networks

Layer 1
Input
Layer

$x_1 = a_1^{(1)}$

$x_2 = a_2^{(1)}$

$\vdots$

$x_n = a_{s_1}^{(1)}$

$a_0^{(1)} = 1$

$$\Theta^{(1)} = \begin{bmatrix} \Theta_{10}^{(1)} & \Theta_{11}^{(1)} & \cdots & \Theta_{1s_1}^{(1)} \\ \Theta_{20}^{(1)} & \Theta_{21}^{(1)} & \cdots & \Theta_{2s_1}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \Theta_{s_20}^{(1)} & \Theta_{s_21}^{(1)} & \cdots & \Theta_{s_2s_1}^{(1)} \end{bmatrix}$$

Layer 2
Hidden Layer

$$a_1^{(2)} = g\left(z_1^{(2)}\right) \qquad z_1^{(2)} = \left(\Theta^{(1)}a^{(1)}\right)_1$$

$$a_2^{(2)} = g\left(z_2^{(2)}\right) \qquad z_2^{(2)} = \left(\Theta^{(1)}a^{(1)}\right)_2$$

$\vdots$

$$a_{s_2}^{(2)} = g\left(z_{s_2}^{(2)}\right) \qquad z_{s_2}^{(2)} = \left(\Theta^{(1)}a^{(1)}\right)_{s_2}$$

$a_0^{(2)} = 1$

$$\Theta^{(2)} = \begin{bmatrix} \Theta_{10}^{(2)} & \Theta_{11}^{(2)} & \cdots & \Theta_{1s_2}^{(2)} \\ \Theta_{20}^{(2)} & \Theta_{21}^{(2)} & \cdots & \Theta_{2s_2}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \Theta_{s_30}^{(2)} & \Theta_{s_31}^{(2)} & \cdots & \Theta_{s_3s_2}^{(2)} \end{bmatrix}$$

Layer 3
Hidden Layer

$$a_1^{(3)} = g\left(z_1^{(3)}\right) \qquad z_1^{(3)} = \left(\Theta^{(2)}a^{(2)}\right)_1$$

$$a_2^{(3)} = g\left(z_2^{(3)}\right) \qquad z_2^{(3)} = \left(\Theta^{(2)}a^{(2)}\right)_2$$

$\vdots$

$$a_{s_3}^{(3)} = g\left(z_{s_3}^{(3)}\right) \qquad z_{s_3}^{(3)} = \left(\Theta^{(2)}a^{(2)}\right)_{s_3}$$

$\cdots$

$\cdots$

$\cdots$

Layer L
Output Layer

$$a_1^{(L)} = g\left(z_1^{(L)}\right) = \hat{y}_1 \qquad z_1^{(L)} = \left(\Theta^{(L-1)}a^{(L-1)}\right)_1$$

$$a_2^{(L)} = g\left(z_2^{(L)}\right) = \hat{y}_2 \qquad z_2^{(L)} = \left(\Theta^{(L-1)}a^{(L-1)}\right)_2$$

$\vdots$

$$a_{s_L}^{(L)} = g\left(z_{s_L}^{(L)}\right) = \hat{y}_k \qquad z_{s_L}^{(L)} = \left(\Theta^{(L-1)}a^{(L-1)}\right)_{s_L}$$

Backpropagation Algorithm

$a_1^{(1)}$		$a_1^{(2)}$		$a_1^{(3)}$	$\cdots$	$a_1^{(L-1)}$		$a_1^{(L)}$
$a_2^{(1)}$	$\Theta^{(1)}_{s_2\times(s_1+1)}$	$a_2^{(2)}$	$\Theta^{(2)}_{s_3\times(s_2+1)}$	$a_2^{(3)}$	$\cdots$	$a_2^{(L-1)}$	$\Theta^{(L-1)}_{s_L\times(s_{L-1}+1)}$	$a_2^{(L)}$
$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
$a_{s_1}^{(1)}$		$a_{s_2}^{(2)}$		$a_{s_3}^{(3)}$	$\cdots$	$a_{s_{L-1}}^{(L-1)}$		$a_{s_L}^{(L-1)}$

$a^{(1)} = x$	$z^{(2)} = \Theta^{(1)}a^{(1)}$	$z^{(3)} = \Theta^{(2)}a^{(2)}$	$z^{(L-1)} = \Theta^{(L-2)}a^{(L-2)}$	$z^{(L)} = \Theta^{(L-1)}a^{(L-1)}$
$a_0^{(1)} = 1$	$a^{(2)} = g\left(z^{(2)}\right)$	$a^{(3)} = g\left(z^{(3)}\right)$	$a^{(L-1)} = g\left(z^{(L-1)}\right)$	$a^{(L)} = g\left(z^{(L)}\right) = h_\theta(x)$
	$a_0^{(2)} = 1$	$a_0^{(3)} = 1$	$a_0^{(L-1)} = 1$	

$\Delta_{ij}^{(l)} = 0$

For i from 1 *to* m *do*

$a^{(1)} = x^{(i)}$

Perform forward propagation to compute $a^{(l)}$ *for* $l = 2, \dots, L$

$\delta^{(L)} = a^{(L)} - y^{(i)}$

Perform back propagation to compute $\delta^{(l)}$ *for* $l = L-1, \dots, 2$

$$\delta^{(l)} = \left[\Theta^{(l)}\right]^T \delta^{(l+1)} .\times g'\left(z^{(l)}\right) = \left[\Theta^{(l)}\right]^T \delta^{(l+1)} .\times a^{(l)}\left(1 - a^{(l)}\right)$$

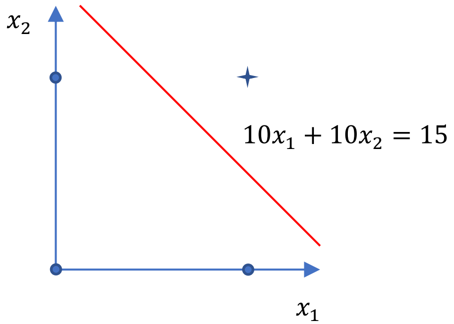
$\Delta_{ij}^{(l)} := \Delta_{ij}^{(l)} + a_j^{(l)}\delta_i^{(l+1)}$ *or* $\Delta^{(l)} := \Delta^{(l)} + \delta^{(l+1)}\left[a^{(l)}\right]^T$

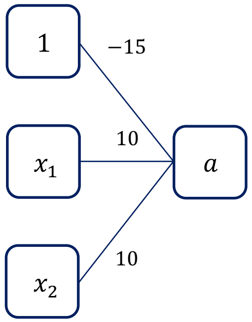
$$D_{ij}^{(l)} := \begin{cases} \dfrac{1}{m}\Delta_{ij}^{(l)} + \dfrac{\lambda}{m}\Theta_{ij}^{(l)} \ if\ j \neq 0 \\ \dfrac{1}{m}\Delta_{ij}^{(l)} \ if\ j = 0 \end{cases}$$

$$\frac{\partial}{\partial \Theta_{ij}^{(l)}} J(\Theta) = D_{ij}^{(l)}$$

Examples

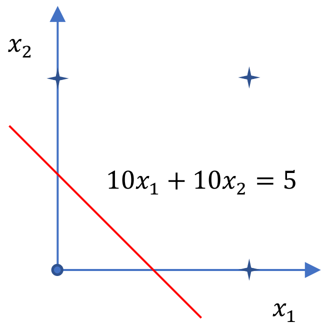
x_1	x_2	$x_1 \,\&\, x_2$
0	0	0
0	1	0
1	0	0
1	1	1

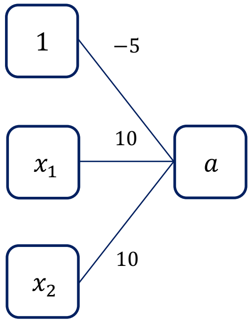




$$z = -15 + 10x_1 + 10x_2, a = g(z)$$

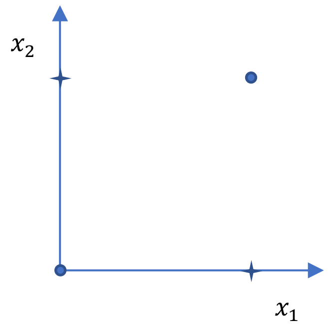
x_1	x_2	$x_1 \vee x_2$
0	0	0
0	1	1
1	0	1
1	1	1

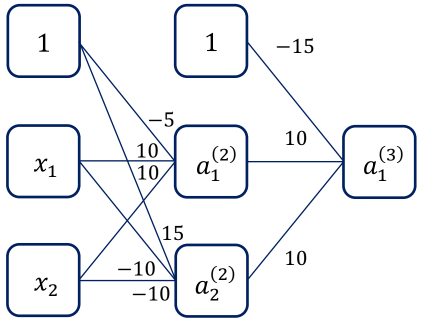




$$z = -5 + 10x_1 + 10x_2, a = g(z)$$

x_1	x_2	$x_1 \oplus x_2$
0	0	0
0	1	1
1	0	1
1	1	0





$$\Theta^{(1)} = \begin{bmatrix} -5 & 10 & 10 \\ 15 & -10 & -10 \end{bmatrix},$$
$$\Theta^{(2)} = [-15 \quad 10 \quad 10],$$
$$z^{(2)} = \Theta^{(1)}x, a^{(2)} = g\left(z^{(2)}\right),$$
$$z^{(3)} = \Theta^{(2)}a^{(2)}, a^{(3)} = g\left(z^{(3)}\right)$$