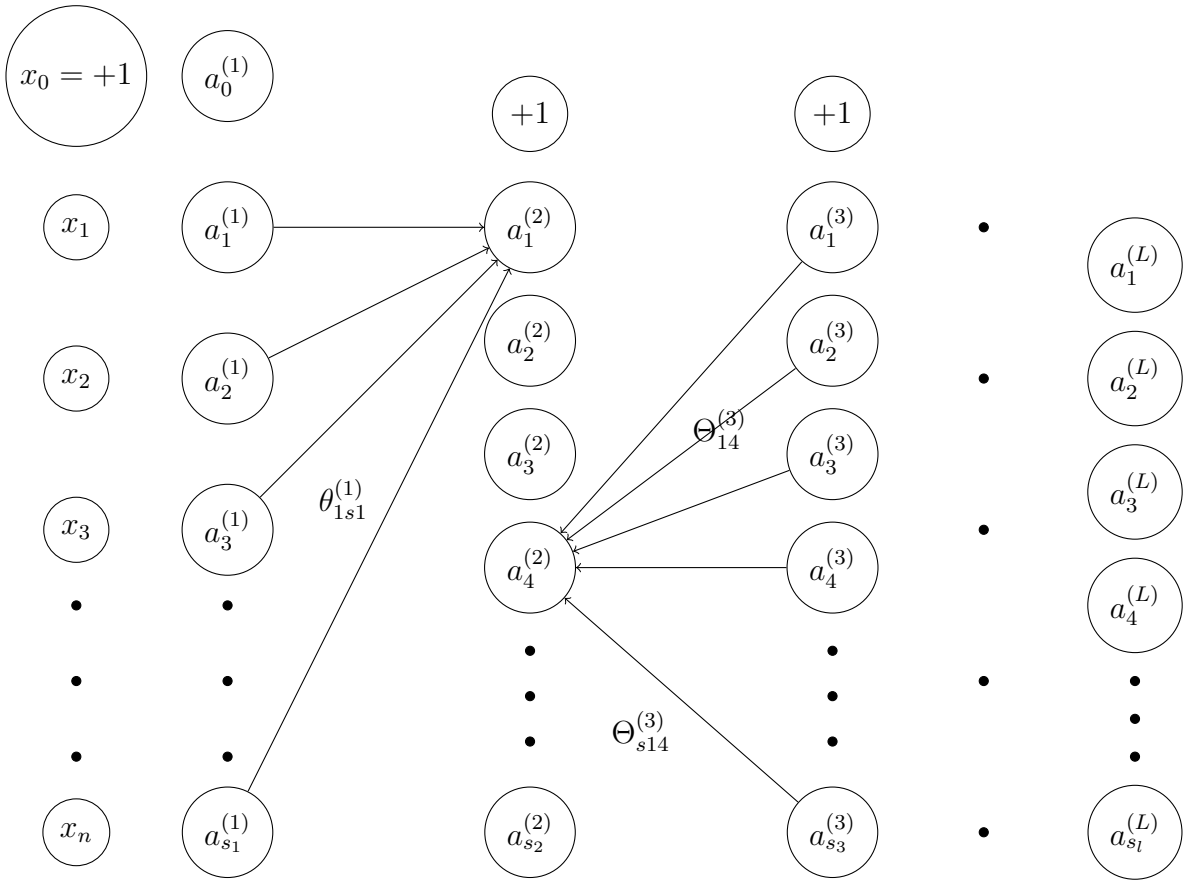
1 神经网络图例

1.1 神经网络示意图



1.2 神经网络 - 前向算法

1.2.1 X、θ、Θ、z、a

1. X

$$
\begin{aligned}
X &= \begin{pmatrix} (x^{(1)})^T \\ (x^{(2)})^T \\ (x^{(3)})^T \\ \vdots \\ (x^{(m)})^T \end{pmatrix} \\
&= \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots & x_n^{(2)} \\ x_1^{(3)} & x_2^{(3)} & x_3^{(3)} & \dots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & x_3^{(m)} & \dots & x_n^{(m)} \end{pmatrix} \\
&= \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots & x_n^{(2)} \\ x_1^{(3)} & x_2^{(3)} & x_3^{(3)} & \dots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & x_3^{(m)} & \dots & x_n^{(m)} \end{pmatrix} \Rightarrow (m, n)
\end{aligned} \tag{1}
$$

2. $a^{(1)}$

$$
a^{(1)} = X \Rightarrow (m, n) \tag{2}
$$

3. $\Theta^{(1)}$

$$
\Theta^{(1)} = \begin{pmatrix} \theta_{10}^{(1)} & \theta_{11}^{(1)} & \theta_{12}^{(1)} & \dots & \theta_{1,s_1}^{(1)} \\ \theta_{20}^{(1)} & \theta_{21}^{(1)} & \theta_{22}^{(1)} & \dots & \theta_{2,s_1}^{(1)} \\ \theta_{30}^{(1)} & \theta_{31}^{(1)} & \theta_{32}^{(1)} & \dots & \theta_{3,s_1}^{(1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{s_20}^{(1)} & \theta_{s_21}^{(1)} & \theta_{s_22}^{(1)} & \dots & \theta_{s_2,s_1}^{(1)} \end{pmatrix} \Rightarrow (s_2, s_1 + 1) = (s_2, n + 1) \tag{3}
$$

4. $z^{(2)}$

给$a^{(1)}$的每个数据均添加上$a_0=1$后与$\Theta^{(1)}$计算,得到$z^{(2)}$[1] $=(1,a^{(1)})(\Theta^{(1)})^T$

$$
\begin{aligned}
z^{(2)} &= a^{(1)}(\Theta^{(1)})^T \Rightarrow (m,n+1)*(n+1,s_2) \\
&= \begin{pmatrix} 1 & x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots & x_n^{(2)} \\ 1 & x_1^{(3)} & x_2^{(3)} & x_3^{(3)} & \dots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(m)} & x_2^{(m)} & x_3^{(m)} & \dots & x_n^{(m)} \end{pmatrix} \begin{pmatrix} \theta_{10}^{(1)} & \theta_{11}^{(1)} & \theta_{12}^{(1)} & \dots & \theta_{1,n}^{(1)} \\ \theta_{20}^{(1)} & \theta_{21}^{(1)} & \theta_{22}^{(1)} & \dots & \theta_{2,n}^{(1)} \\ \theta_{30}^{(1)} & \theta_{31}^{(1)} & \theta_{32}^{(1)} & \dots & \theta_{3,n}^{(1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{s_2,0}^{(j)} & \theta_{s_2,1}^{(j)} & \theta_{s_2,2}^{(j)} & \dots & \theta_{s_2,n}^{(1)} \end{pmatrix}^T \\
&= \begin{pmatrix} 1 & x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots & x_n^{(2)} \\ 1 & x_1^{(3)} & x_2^{(3)} & x_3^{(3)} & \dots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(m)} & x_2^{(m)} & x_3^{(m)} & \dots & x_n^{(m)} \end{pmatrix} \begin{pmatrix} \theta_{10}^{(1)} & \theta_{20}^{(1)} & \theta_{30}^{(1)} & \dots & \theta_{s_2,0}^{(1)} \\ \theta_{11}^{(1)} & \theta_{21}^{(1)} & \theta_{31}^{(1)} & \dots & \theta_{s_2,1}^{(1)} \\ \theta_{12}^{(1)} & \theta_{22}^{(1)} & \theta_{32}^{(1)} & \dots & \theta_{s_2,2}^{(1)} \\ \theta_{13}^{(1)} & \theta_{23}^{(1)} & \theta_{33}^{(1)} & \dots & \theta_{s_2,3}^{(1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{1,n}^{(1)} & \theta_{2,n}^{(1)} & \theta_{3,n}^{(1)} & \dots & \theta_{s_2,n}^{(1)} \end{pmatrix} \\
&= \begin{pmatrix} z^{(2)(1)} \\ z^{(2)(2)} \\ z^{(2)(3)} \\ \vdots \\ z^{(2)(m)} \end{pmatrix} \\
&= \begin{pmatrix} z_1^{(2)(1)} & z_2^{(2)(1)} & z_3^{(2)(1)} & \dots & z_{s_2}^{(2)(1)} \\ z_1^{(2)(2)} & z_2^{(2)(2)} & z_3^{(2)(2)} & \dots & z_{s_2}^{(2)(2)} \\ z_1^{(2)(3)} & z_2^{(2)(3)} & z_3^{(2)(3)} & \dots & z_{s_2}^{(2)(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_1^{(2)(m)} & z_2^{(2)(m)} & z_3^{(2)(m)} & \dots & z_{s_2}^{(2)(m)} \text{注[2]} \end{pmatrix} \Rightarrow (m,n+1)*(n+1,s_2)=(m,s_2)
\end{aligned}
\tag{4}
$$

5. $a^{(2)}$

$$a^{(2)} = g(z^{(2)}) \Rightarrow (m,s_2) \tag{5}$$

6. 一般式

注[1]从$a^{(1)}$得到$a^{(2)}$需要经过sigmoid()函数，后续的从$a^{(j)}$得到$a^{(j+1)}$均需要经过sigmoid()函数

后续同理：

$$
\begin{aligned}
\Theta^{(2)} &= \begin{pmatrix} \theta_{10}^{(2)} & \theta_{11}^{(2)} & \theta_{12}^{(2)} & \dots & \theta_{1,s_2}^{(2)} \\ \theta_{20}^{(2)} & \theta_{21}^{(2)} & \theta_{22}^{(2)} & \dots & \theta_{2,s_2}^{(2)} \\ \theta_{30}^{(2)} & \theta_{31}^{(2)} & \theta_{32}^{(2)} & \dots & \theta_{3,s_2}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{s_30}^{(2)} & \theta_{s_31}^{(2)} & \theta_{s_32}^{(2)} & \dots & \theta_{s_3,s_2}^{(2)} \end{pmatrix} \Rightarrow (s_3, s_2+1) \\
z^{(3)} &= (1, a^{(2)})(\Theta^{(2)})^T \Rightarrow (m, s_2+1) * (s_2+1, s_3) = (m, s_3) \\
a^{(3)} &= g(z^{(3)}) \Rightarrow (m, s_3) \\
&\vdots \\
a^{(j)} &= g(z^{(j-1)}) \Rightarrow (m, s_j) \\
\Theta^{(j)} &= \begin{pmatrix} \theta_{10}^{(j)} & \theta_{11}^{(j)} & \theta_{12}^{(j)} & \dots & \theta_{1,s_j}^{(j)} \\ \theta_{20}^{(j)} & \theta_{21}^{(j)} & \theta_{22}^{(j)} & \dots & \theta_{2,s_j}^{(j)} \\ \theta_{30}^{(j)} & \theta_{31}^{(j)} & \theta_{32}^{(j)} & \dots & \theta_{3,s_j}^{(j)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{s_{j+1}0}^{(j)} & \theta_{s_{j+1}1}^{(j)} & \theta_{s_{j+1}2}^{(j)} & \dots & \theta_{s_{j+1},s_j}^{(j)} \end{pmatrix} \Rightarrow (s_{j+1}, s_j+1) \\
z^{(j+1)} &= (1, a^{(j)})(\Theta^{(j)})^T \Rightarrow (m, s_j+1) * (s_j+1, s_{j+1}) = (m, s_3) \\
a^{(j+1)} &= g(z^{(j+1)}) \Rightarrow (m, s_{j+1})
\end{aligned}
\tag{6}
$$

1.2.2 y

$$
y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ \vdots \\ y^{(m)} \end{pmatrix}_{m*1} \tag{7}
$$

为进行矩阵运算，要将其转化为如下形式:[注[3]]

$$
y = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}_{m*s_L} \text{注[4]} \tag{8}
$$

注[3]y所对应的值所在的索引位置值为1，其他位置均为0