Forward propagation:

Input layer: x

Output layer: y

Hidden layers: $a_j^{(i)} = \left(\theta_j^{(i)} \cdot a_{j-1}^{(i)}\right) - \text{bias}$

Sigmoid

$$6\left(\begin{bmatrix} \theta_{0,0} & \theta_{0,1} & \dots & \theta_{0,n} \\ \vdots & \ddots & & \vdots \\ \theta_{\ell,0} & \theta_{\ell,1} & \dots & \theta_{\ell,n} \end{bmatrix} \begin{bmatrix} a_0^{(0)} \\ a_1^{(0)} \\ \vdots \\ a_n^{(0)} \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}\right) = \begin{bmatrix} a_0^{(1)} \\ a_1^{(1)} \\ \vdots \\ a_n^{(1)} \end{bmatrix}$$

$$\boxed{6(\theta\, a^{(i)} - b) = a^{(i+1)}}$$

Cost funtion: $\min\left(\sum (\text{prediction} - \text{real})^2\right)$

Backpropagation: algorithm to compute gradient.