Neural Networks

- Model Representation

$$a_1^{(2)} = g(\Theta_{10}^{(1)}x_0 + \Theta_{11}^{(1)}x_1 + \Theta_{12}^{(1)}x_2 + \Theta_{13}^{(1)}x_3)$$

$$a_2^{(2)} = g(\Theta_{20}^{(1)}x_0 + \Theta_{21}^{(1)}x_1 + \Theta_{22}^{(1)}x_2 + \Theta_{23}^{(1)}x_3)$$

$$a_3^{(2)} = g(\Theta_{30}^{(1)}x_0 + \Theta_{31}^{(1)}x_1 + \Theta_{32}^{(1)}x_2 + \Theta_{33}^{(1)}x_3)$$

$$\rightarrow h_\Theta(x) = a_1^{(3)} = g(\Theta_{10}^{(2)}a_0^{(2)} + \Theta_{11}^{(2)}a_1^{(2)} + \Theta_{12}^{(2)}a_2^{(2)} + \Theta_{13}^{(2)}a_3^{(2)})$$

- Cost function

$$h_\Theta(x) \in R^K, \quad \left(h_\Theta(x)\right)_i = i^{th}\ output$$

$$J(\Theta) = -\frac{1}{m}\left(\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} log\left(h_\Theta\left(x^{(i)}\right)\right)_k + (1 - y_k^{(i)})\log(1 - \left(h_\Theta\left(x^{(i)}\right)\right)_k)\right) + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}(\Theta_{ji}^{(l)})^2$$

- Backpropagation algorithm

Gradient computation:

$$J(\Theta) = -\frac{1}{m}\left(\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} log\left(h_\Theta\left(x^{(i)}\right)\right)_k + (1 - y_k^{(i)})\log(1 - \left(h_\Theta\left(x^{(i)}\right)\right)_k)\right) + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}(\Theta_{ji}^{(l)})^2$$

Given one training example (x, y):

Forward propagation:

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

$$\rightarrow z^{(2)} = \Theta^{(1)} a^{(1)}$$

$$\rightarrow a^{(2)} = g\left(z^{(2)}\right) \ (add\ a_0^{(2)})$$

$$\rightarrow z^{(3)} = \Theta^{(2)} a^{(2)}$$

$$\rightarrow a^{(3)} = g\left(z^{(3)}\right) \ (add\ a_0^{(3)})$$

$$\rightarrow z^{(4)} = \Theta^{(3)} a^{(3)}$$

$$\rightarrow a^{(4)} = h_\Theta(x) = g(z^{(4)})$$

Gradient computation: Backpropagation algorithm

Intuition: $\delta_j^{(l)}$ = "error" of node j in layer l.

For each output unit (layer L = 4)

$$\delta_j^{(4)} = a_j^{(4)} - y_j$$

$$\rightarrow \delta^{(3)} = (\Theta^{(3)})^T \delta^{(4)} .* \, g'(z^{(3)})$$

$$\rightarrow \delta^{(2)} = (\Theta^{(2)})^T \delta^{(3)} .* \, g'(z^{(2)})$$

Backpropagation algorithm

Training set $\{\left(x^{(1)}, y^{(1)}\right), \ldots, (x^{(m)}, y^{(m)})\}$

Set $\Delta_{ij}^{(l)} = 0 \ (for\ all\ l, i, j)$

For $i = 1\ to\ m$

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

Perform forward propagation to compute $a^{(l)}$, for $l = 2,3,\ldots,L$

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

Compute $\delta^{(L-1)}, \delta^{(L-2)}, \ldots, \delta^{(2)}$

$\Delta_{ij}^{(l)} \coloneqq \Delta_{ij}^{(l)} + a_j^{(l)} \delta_i^{(l+1)}$

$\rightarrow D_{ij}^{(l)} \coloneqq \frac{1}{m} \Delta_{ij}^{(l)} + \lambda \Theta_{ij}^{(l)},\ \ if\ j \neq 0,$

$\rightarrow D_{ij}^{(l)} \coloneqq \frac{1}{m} \Delta_{ij}^{(l)},\ \ if\ j = 0.$