Derivation of BP Algorithm

Lifu Huang

Feb. 2016

1 Derivation

1.1 Cost function

$$\begin{aligned} J(W,b) &= \frac{1}{m}\sum_{i=1}^{m} J(W,b;x,y) \\ J(W,b;x,y) &= \frac{1}{2}||h_{W,b}(x)-y||^2 \end{aligned} \tag{1}$$

1.2 Computation of δ

$$\begin{aligned} \delta_i^{(n_l)} &= \frac{\partial}{\partial z_i^{(n_l)}} J(W,b;x,y) \\ &= \frac{\partial}{\partial z_i^{(n_l)}} \frac{1}{2}||h_{W,b}(x)-y||^2 \\ &= \frac{\partial}{\partial z_i^{(n_l)}} \frac{1}{2}\sum_{j=1}^{s_{n_l}} (f(z_j^{(n_l)})-y_j)^2 \\ &= (f(z_i^{(n_l)})-y_i)f'(z_i^{(n_l)}) \\ &= -f'(z_i^{(n_l)})(y_i - f(z_i^{(n_l)})) \end{aligned} \tag{2}$$

$$\delta^{(n_l)} = -(y - a^{(n_l)}) \circ f'(z^{(n_l)}) \tag{3}$$

$$
\begin{aligned}
\delta_i^{(l)} &= \frac{\partial}{\partial z_i^{(l)}} J(W,b;x,y) \\
&= \sum_{j=1}^{s_{l+1}} \frac{\partial}{\partial z_j^{(l+1)}} J(W,b;x,y) \frac{\partial z_j^{(l+1)}}{\partial z_i^{(l)}} \\
&= \sum_{j=1}^{s_{l+1}} \delta_j^{(l+1)} \frac{\partial}{\partial z_i^{(l)}} \sum_{k=1}^{s_l} (W_{jk}^{(l)} f(z_k^{(l)})) \\
&= \sum_{j=1}^{s_{l+1}} \delta_j^{(l+1)} W_{ji}^{(l)} f'(z_i^{(l)}) \\
&= f'(z_i^{(l)}) (\sum_{j=1}^{s_{l+1}} \delta_j^{(l+1)} W_{ji}^{(l)})
\end{aligned} \tag{4}
$$

$$
\delta^{(l)} = ((W^{(l)})^T \delta^{(l+1)}) \circ f'(z^{(l)}) \tag{5}
$$

1.3 Derivative wrt W

$$
\begin{aligned}
\frac{\partial}{\partial W_{ij}^{(l)}} J(W,b;x,y) &= \sum_{k=1}^{s_{l+1}} \frac{\partial}{\partial z_k^{(l+1)}} J(W,b;x,y) \frac{\partial z_k^{(l+1)}}{\partial W_{ij}^{(l)}} \\
&= \delta_i^{(l+1)} a_j^{(l)}
\end{aligned} \tag{6}
$$

$$
\nabla_{W^{(l)}} J(W,b;x,y) = \delta^{(l+1)} (a^{(l)})^T \tag{7}
$$

1.4 Derivative wrt b

$$
\begin{aligned}
\frac{\partial}{\partial b_i^{(l)}} J(W,b;x,y) &= \sum_{k=1}^{s_{l+1}} \frac{\partial}{\partial z_k^{(l+1)}} J(W,b;x,y) \frac{\partial z_k^{(l+1)}}{\partial b_i^{(l)}} \\
&= \delta_i^{(l+1)}
\end{aligned} \tag{8}
$$

$$
\nabla_{b^{(l)}} J(W,b;x,y) = \delta^{(l+1)} \tag{9}
$$

2 Conclusion

$$
\delta^{(n_l)} = -(y - a^{(n_l)}) \circ f'(z^{(n_l)}) \tag{10}
$$

$$
\delta^{(l)} = ((W^{(l)})^T \delta^{(l+1)}) \circ f'(z^{(l)}) \tag{11}
$$

$$
\nabla_{W^{(l)}} J(W,b;x,y) = \delta^{(l+1)} (a^{(l)})^T \tag{12}
$$

$$
\nabla_{b^{(l)}} J(W,b;x,y) = \delta^{(l+1)} \tag{13}
$$