Neural networks

April 5, 2017

1 Back-propagation algorithm

Neural network consisting of L layers:

- C – cost function
- z_j^l – weighted input to neuron j in layer l of the neural network
- σ – neuron activation function
- w_{jk}^l – weight connecting neuron j in layer l to neuron k in layer $l-1$
- b_j^l – bias of neuron j in layer l
- a_j^l – output of neuron j in layer l

Neuron output is related to neuron weighted input as

$$a_j^l = \sigma\left(z_j^l\right) = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right). \tag{1}$$

Error of neuron j in last layer is

$$\delta_j^L = \frac{\partial C}{\partial z_j^L} = \sum_k \frac{\partial C}{\partial a_k^L}\frac{\partial a_k^L}{\partial z_j^L} = \frac{\partial C}{\partial a_j^L}\frac{\partial a_j^L}{\partial z_j^L} = \frac{\partial C}{\partial a_j^L}\,\sigma'\left(z_j^L\right) \tag{2}$$

Error of neuron j in layer l is

$$\delta_j^l = \frac{\partial C}{z_j^l} = \sum_k \frac{\partial C}{\partial z_k^{l+1}}\frac{\partial z_k^{l+1}}{\partial z_j^l}$$

Taking into account that

$$z_k^{l+1} = \sum_n w_{kn}^{l+1} a_n^l + b_k^{l+1} = \sum_n w_{kn}^{l+1}\sigma(z_n^l) + b_k^{l+1}$$

Finally

$$\delta_j^l = \sum_k \frac{\partial C}{\partial z_k^{l+1}}\frac{\partial}{\partial z_j^l}\left(\sum_n w_{kn}^{l+1}\sigma(z_n^l) + b_k^{l+1}\right) = \sum_k \delta_k^{l+1} w_{kj}^{l+1}\sigma'(z_j^l) \tag{3}$$