Chapter theory 5 solution

Artem Puzanov

2016-09-27

1 Backpropagation for convolutinal networks

Intuition   Usual layers require no re-definition. The only potentially interesting ones are the ones that are either connected to the layers of new type, or layers of the new type themselves. Let's go backwards from the final layer:

Max-pooling - Fully connected This connection is updated as usual. Fully connected layer updates weights for all neurons from each feature map, no change here.

Max-pooling δ: no changes here. For each δ is connected only to some neurons, but that changes nothing in terms of equations. Max-pooling b: doesn't exist (according to the textbook definition) Max-pooling w: doesn't exist (there is no weights, only (the force) maximum function)

Convolution layer is the interesting part of the problem:

Certain neurons in convolution layer are switched on/off in the next layer, in a discrete way. A direct consequence of this is that they are not updated unless they were on. We can implement that either by nullifying weights in conv-to-maxpool matrix, or nullyfing non-activated neurons (and therefore their derivatives). It seems the second approach is more computationally sound.

Another new thing is shared weigths and biases. A derivative of shared weight is a sum of individual neuron's derivatives. They all have different δ, but it all goes to the same parameter. Same is true for bias

Each convolution/max-pooling layer is optimized separately ofc. Our equations are for one such layer, but that doesn't lead to a loss of generality.

Mathematical notation   Here are basic backpropagation equations:

$$\delta^L = \sum{}'(z^L)\nabla_a C = \nabla_a C \odot \sigma'(z^L)$$

$$\delta^l = (w^{l+1})^T \delta^{l+1} \odot \sigma'(z^l)$$

$$\frac{\partial C}{\partial b_i} = \delta_i^l$$

$$\frac{\partial C}{\partial w_{ij}} = \delta_i a_j^{l-1}$$

For a (single) maxpool layer, delta equation stays the same:

$$\delta_{maxpool}^l = (w^{l+1})^T \delta^{l+1} \odot \sigma'(z^l)$$

Equations for b and w are not needed (the function is just maximum).

For a (single) convolution layer, delta equation is modified slightly:

$$\delta_{conv}^l = (w^{l+1})^T \delta^{l+1} \odot (\sigma'(z^l) \odot [ismax(z^l)])$$

w^{l+1} matrix consists of 1 and 0, and stays constant. where $[]$ is an indicator function, and $ismax$ is a function that determines whether particular neuron was maximum in it's pool.

Shared bias equation becomes:

$$\frac{\partial C}{\partial b} = \sum_i \delta_i^l$$

And shared weight:

$$\frac{\partial C}{\partial w} = \sum_i \delta_i^l a_i^{l-1}$$

where i is an input index.