Andrew Dodd

5/10/16

Computer Vision: PA5 Written Portion

A – Convolution Layer Part 1

Note: Numbered equations are from homework numbered equations. Letters are unique to this write-up.

We begin with Equation 3:

$$\frac{\partial L}{\partial x[a,b]} = \sum_e \sum_f \frac{\partial L}{\partial y[e,f]} \cdot \frac{\partial y[e,f]}{\partial x[a,b]} \qquad Eq.3$$

Our goal is to show that this is equivalent to:

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} * \widetilde{w} \qquad Eq.\ A$$

The forward propagation equation (convolution equation y = x * w) states:

$$y[e,f] = \sum_a \sum_b w[e-a, f-b] \cdot x[a,b] \qquad Eq.B$$

Rearranging terms using commutativity of convolution:

$$y[e,f] = \sum_a \sum_b w[a,b] \cdot x[e-a, f-b] \qquad Eq.C$$

Taking the partial derivative of Eq. B:

$$\frac{\partial y[e,f]}{\partial x[a,b]} = w[e-a, f-b] \qquad Eq.D$$

Substituting Eq. D into Eq. 3:

$$\frac{\partial L}{\partial x[a,b]} = \sum_e \sum_f \frac{\partial L}{\partial y[e,f]} \cdot w[e-a, f-b] \qquad Eq.E$$

We can rewrite Eq. B (which is the form of a convolution equation with x and w to get y) by switching the letters of the indices (a switch with e, b switch with f) and switching the w and x terms with commutativity:

$$y[a,b] = \sum_e \sum_f x[e,f] \cdot w[a-e, b-f] \qquad Eq.F$$

Note the similarity between Eq. E and F. The only difference between the general equation for a convolution and Eq. E is that the kernel indices of w are negatives of each other. We can reduce Eq. E now using the definition of convolution (Eq. F):

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} * \widetilde{w}$$

A – Convolution Layer Part 2

Convolution is a commutative and associative transform, meaning when convolving two matrices, one can use either as the "kernel". With the convolution equation, therefore we have:

$$y = w * x \quad or \quad y = x * w$$

In fact, because we proved Q-A1, Q-A2 follows due to the definition of convolution as a commutative and associative matrix operation.

This can also be seen below:

Beginning with the following Eq. B from Q-A1:

$$y[e, f] = \sum_a \sum_b w[e - a, f - b] \cdot x[a, b] \qquad Eq. B$$

We can use convolution commutativity:

$$y[e, f] = \sum_a \sum_b w[e - a, f - b] \cdot x[a, b] \qquad Eq. G$$

Taking the partial derivative:

$$\frac{\partial y[e, f]}{\partial w[a, b]} = x[e - a, f - b] \qquad Eq. H$$

Now, we can rewrite Eq 3 as the following, switching x with w due to convolution commutativity:

$$\frac{\partial L}{\partial w[a, b]} = \sum_e \sum_f \frac{\partial L}{\partial y[e, f]} \frac{\partial y[e, f]}{\partial w[a, b]} \qquad Eq. I$$

Substituting Eq. H into I:

$$\frac{\partial L}{\partial w[a, b]} = \sum_e \sum_f \frac{\partial L}{\partial y[e, f]} x[e - a, f - b] \qquad Eq. J$$

Comparing to the convolution equation:

$$y[a, b] = \sum_e \sum_f x[e, f] \cdot w[a - e, b - f] \qquad Eq.K$$

We see that the indices of the Eq. J x term and those of the Eq. K w term are switched. Thus to put the equation into convolution form, we must take the negative of the indices of Eq. J x. Thus:

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} * \tilde{x}$$

B – Fully Connected Layer

Apologies for the handwriting.

(2) $y[j] = \sum_i w[j, i] \, x[i]$

(3) $\frac{\partial L}{\partial x[i]} = \sum_j \frac{\partial L}{\partial y[j]} \frac{\partial y[j]}{\partial x[i]}$

$\frac{\partial}{\partial x[i]} \left(y[j] \right) = w[j, i] \leftarrow$ (derivative of (2) WRT $x[i]$)

$\frac{\partial L}{\partial x[i]} = \sum_j \frac{\partial L}{\partial y[j]} w[j, i]$

$\frac{\partial L}{\partial x} = w^T \frac{\partial L}{\partial y}$

C – Softmax Loss Layer

Substituting p_i into L:

$$L = -\sum_i t_i \log\left[\frac{e^{s_i}}{\sum_j e^{s_j}}\right]$$

$$L = -\sum_i t_i \left(\log(e^{s_i}) - \log\left(\sum_j e^{s_j}\right)\right)$$

$$= -\sum_i t_i \left(s_i - \log\left(\sum_j e^{s_j}\right)\right)$$

$$= -\sum_i t_i s_i + \sum_i t_i \log\left(\sum_j e^{s_j}\right)$$

$$L = -\sum t_i s_i + \sum_i t \log\left(\sum_j e^{s_j}\right)$$

$$L = -s_a + \log\left(\sum_j e^{s_j}\right)$$

Now taking the derivative of L as an indicator function:

$\frac{\partial}{s_k}\left(\sum_j e^{s_j}\right)$, Chain rule

Taking the derivative of L:

$$\frac{\partial L}{\partial s_k} = \begin{cases} \text{if } a = k & -1 + \frac{e^{s_k}}{\sum_j e^{s_j}} \\ \text{else} & 0 + \frac{e^{s_k}}{\sum_j e^{s_j}} \end{cases} = \begin{cases} p_k - 1 \\ p_k \end{cases}$$

We notice that t is a vector of zeros except for a 1 so we can expand this to:

$$\frac{\partial L}{\partial s} = p - t$$