TTIC 31230 Fundamentals of Deep Learning, winter 2019

CNN Problems

In these problems, as in the lecture notes, capital letter indeces are used to indicate subtensors (slices) so that, for example, $M[I,J]$ denotes a matrix while $M[i,j]$ denotes one element of the matrix, $M[i,J]$ denotes the ith row, and $M[I,j]$ denotes the jth collumn.

We also adopt the convention, similar to true Einstein notation, that repeated capital indeces in a product of tensors are implicitly summed. We can then write the inner product $e[w,I]^\top h[t,I]$ as $e[w,I]h[t,I]$. Using this implicit summation notation we can avoid ever using transpose.

Problem 1. Consider convolving a kernel $K[n_{\text{out}}, \Delta x, \Delta y, n_{\text{in}}]$ with thresholds $B[n_{\text{out}}]$ on a layer $L[b,x,y,n_{\text{in}}]$ where $B,\ X,\ Y, N_{\text{out}}, N_{\text{in}}, \Delta X,\ \Delta Y$ are the number of possible values for b, x, y, n_{out}, n_{in}, Δx and Δy respectively. How many floating point multiplies are required in computing the convolution on the batch (without any activation function)?

Solution:

$$BXY\ \Delta X\ \Delta Y\ N_{\text{out}}\ N_{\text{in}}$$

Problem 2: Suppose that we want a video CNN producing layers of the form $L[b,x,y,t,n]$ which are the same as the layers of an image CNN but with an additional time index. Write the equation for computing $L_{\ell+1}[b,x,y,t,j]$ from the tensor $L_\ell[B,X,Y,T,I]$. Your filter should include an index Δt and handle a stride s applied to both space and time. Use the repeated index notation for summation.

Solution:

$$L_{\ell+1}[b,x,y,t,n_{\text{out}}] = \sigma(K_{\ell+1}[n_{\text{out}}\Delta X, \Delta Y, \Delta T, N_{\text{in}}]L_\ell[b, sx{+}\Delta X, sy{+}\Delta Y, st{+}\Delta T, N_{\text{in}}]{-}B[n_{\text{out}}])$$