

TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2020

Einstein Notation

Tensors

An array with more than two indices is called a **tensor**.

In a convolutional neural network (CNN) for vision a “layer” in the model is a three index (third order) tensor $L[b, x, y, i]$.

$L[b, x, y, i]$ is the value of feature i for batch element b at image position $\langle x, y \rangle$.

A Convolution Layer

$$W[\Delta x, \Delta y, i, j] \quad L_\ell[b, x, y, i] \quad L_{\ell+1}[b, x, y, j]$$

River Trail Documentation

$$L_{\ell+1}[b, x, y, j]$$

$$= \sigma \left(\left(\sum_{\Delta x, \Delta y, i} W[\Delta x, \Delta y, i, j] L_\ell[b, x + \Delta x, y + \Delta y, i] \right) - B[j] \right)$$

Einstein Notation

The convolution equation can be written more compactly in Einstein notation.

Capital letter indices will be 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 i th row, and $M[I, j]$ denotes the j th column.

By abuse of notation, capital letters will also be used to indicate the range of the indices. A vector $x[I]$ has I dimensions with values $x[0], \dots, x[I - 1]$. A Matrix $M[I, J]$ has $I \times J$ entries.

Einstein Notation

$$y = Wx \quad \text{abbreviates} \quad y[i] = \sum_j W[i, j]x[j].$$

$$y = x^\top W \quad \text{abbreviates} \quad y[j] = \sum_i W[i, j]x[i].$$

i is a “row index” and j is a “column index” of W .

Linear algebra suppresses indices.

Einstein notation uses explicit indices.

Einstein Notation

Einstein notation is the convention that repeated capital indices in a product of tensors are implicitly summed.

$$y = Wx \quad \text{abbreviates} \quad y[i] = \sum_j W[i, j]x[j] = W[i, J]x[J].$$

$$y = x^\top W \quad \text{abbreviates} \quad y[j] = \sum_i W[i, j]x[i] = W[I, j]x[I].$$

An MLP in Einstein Notation

$$h[j] = \sigma \left(W^0[j, I] \, x[I] - b^0[j] \right)$$

$$s[\hat{y}] = \sigma \left(W^1[\hat{y}, J] \, h[J] - b^1[\hat{y}] \right)$$

$$P_\Phi[\hat{y}] = \operatorname{softmax}_{\hat{y}} s[\hat{y}]$$

Convolution in Einstein Notation

$$L_{\ell+1}[b, x, y, j]$$

$$= \sigma \left(\left(\sum_{\Delta x, \Delta y, i} W[\Delta x, \Delta y, i, j] L_{\ell}[b, x + \Delta x, y + \Delta y, i] \right) - B[j] \right)$$

$$= \sigma (W[\Delta X, \Delta Y, I, j] L_{\ell}[b, x + \Delta X, y + \Delta Y, I] - B[j])$$

Why Einstein Notation?

We will need to work with tensors — arrays with more than two indices.

For 2D CNNs the weight tensor and the data tensor each have four indices (including the batch index of the data tensor).

For higher order tensors suppressing indices becomes confusing.

Einstein went back to explicit index notation (Einstein notation) when working with the higher order tensors in his theory of gravitation.

Why Einstein Notation?

Also, the indices of tensors generally have types such as a “time index”, “x coordinate”, “y coordinate”, “batch index”, or “feature index”.

Writing a matrix as $W[T, I]$ where T is a time index and I is a feature index makes the type of the matrix W clear and clarifies the order of the indices (disambiguates W from W^\top).

END