Matmul Operator

Rylan W. Yancey

10/1/2023

Definition & Gradient

When A is an M x N matrix, B is an N x P matrix, and C is an M x P matrix, the product AB for an element in C at some i, j is defined as:

$$C_{ij} = \sum_{k=0}^{N} A_{ik} B_{kj}$$

Visually, this is the vector product of row i in A and column j in B. Lets' begin by finding the derivative of the vector product, defined as. When A is a M x 1 vector, and B is an M x 1 vector, the vector product is defined as:

$$\sum_{i=0}^{M} A_i B_i$$

To find the gradient, we will make use of the sum rule and the product rule.

$$\frac{\delta}{\delta X} \sum_{i=0}^{M} \frac{\delta}{\delta X_i} (B_i (A_i \frac{\delta}{\delta X_i}) + A_i (B_i \frac{\delta}{\delta X_i}))$$

To find the gradient w.r.t. A, we will substitute X as A.

$$\frac{\delta}{\delta A} \sum_{i=0}^{M} \frac{\delta}{\delta A_i} (B_i \frac{\delta A_i}{\delta A_i} + A_i \frac{\delta B_i}{\delta A_i})$$

We now distribute $\frac{\delta}{\delta A_i}$ and simplify.

$$\sum_{i=0}^{M} B_i \frac{\delta A_i}{\delta A_i} + \sum_{i=0}^{M} A_i \frac{\delta B_i}{\delta A_i} = \sum_{i=0}^{M} B_i + \sum_{i=0}^{M} 0 = \sum_{i=0}^{M} B_i$$

Therefore, the gradient of the vector product with respect to A is the sum of the elements of B. The same logic is true for the gradient w.r.t. B, which is the sum of the elements of A. Applying this to the definition of an element of C in matrix multiplication, we can say that the gradient of C_{ij} w.r.t A is the sum of the elements in column B_j, and the gradient w.r.t. B is the sum of the elements in row A_i. For the purposes of gradients, we can conclude that the gradient of C w.r.t. A is B^T, and the gradient of C w.r.t. B. is A^T.

$$\frac{\delta}{\delta A}(AB) = B^T$$

$$\frac{\delta}{\delta B}(AB) = A^T$$

We get B^T and A^T becuase it is convention when working with gradients to transpose the derivatives of matrix multiplication. Under the jacobian convention, this is not so.