1 Tensors

2 Backward pass for basic operations

Proposition 2.1. Let P, X, Q be three matrices of shape $n \times s$, $s \times r$ and $r \times m$.

$$\frac{\partial PXQ}{\partial X} = Q^T \otimes P \tag{1}$$

Proof. Any element w_{ij} of the product $W = PXQ$ is expressed by:

$$w_{ij} = \sum_{h=1}^{r} \sum_{t=1}^{s} p_{it} x_{th} b_{hj}, \tag{2}$$

It follows that:

$$\frac{\partial w_{ij}}{\partial x_{ab}} = a_{ia} b_{bj} \tag{3}$$

This is the definition of the Kronecker product of Q^T and P. $\square$

2.1 Tensor multiplication

Let A and B be two matrices and $C = AB$ the result of their matrix product. Suppose that the gradient of a scalar z with respect to C is given by G. Both $\frac{\partial z}{\partial A}$ and $\frac{\partial z}{\partial B}$ need to be computed in order to complete the backward pass.

Using Proposition 2.1, for $P = I$ the identity matrix, $X = A$ and $Q = B$, we get:

$$\frac{\partial IAB}{\partial A} = \frac{\partial AB}{\partial A} = B^T \otimes I = B^T \tag{4}$$

And similarly, for $P = A$, $X = B$ and $Q = I$:

$$\frac{\partial ABI}{\partial B} = \frac{\partial AB}{\partial B} = I \otimes A = A \tag{5}$$

Finaly, using the chain rule:

$$\frac{\partial z}{\partial A} = \frac{\partial z}{\partial C} \frac{\partial C}{\partial A} = GB^T \ ; \ \frac{\partial z}{\partial B} = \frac{\partial C}{\partial B} \frac{\partial z}{\partial C} = A^T G \tag{6}$$