形状

输入 X 的形状为 (M,K)

输入 Y 的形状为 (K,N)

输出 Z 的形状为 (M,N)

前向传播计算

$$Z = XY$$

计算 z_{ij}:

$$z_{ij} = \sum_{k=1}^{K} x_{ik} y_{kj}$$

反向传播

反向传播的目标是计算损失函数 L 对输入 X 的梯度 $\frac{\partial L}{\partial X}$,以及对输入 Y 的梯度 $\frac{\partial L}{\partial Y}$. 我们已知损失函数 L 对输出 Z 的梯度 $\frac{\partial L}{\partial Z}$.

且由于:

$$z_{ij} = \sum_{k=1}^{K} x_{ik} y_{kj}$$

计算 $\frac{\partial L}{\partial x_{ij}}$:

因为 x_{ij} 出现在 z 的第 i 行的所有元素中，而 z 的每一行共有 K 个元素，所以:

$$\frac{\partial L}{\partial x_{ij}} = \sum_{k=1}^{K} \frac{\partial L}{\partial z_{ik}} \frac{\partial z_{ik}}{\partial x_{ij}}$$

我们先计算 $\frac{\partial z_{ik}}{\partial x_{ij}}$:

$$\frac{\partial z_{ik}}{\partial x_{ij}} = \frac{\partial \left(\sum_{n=1}^{N} x_{in} y_{nk} \right)}{\partial x_{ij}}$$

只有当 $n = j$ 时，右侧偏导数才不为零，因此:

$$\frac{\partial z_{ik}}{\partial x_{ij}} = y_{jk}$$

代入上式:

$$\frac{\partial L}{\partial x_{ij}} = \sum_{k=1}^{K} \frac{\partial L}{\partial z_{ik}} y_{jk}$$

将其写成矩阵形式:

$$\boxed{\frac{\partial L}{\partial X} = \left(\frac{\partial L}{\partial Z}\right) Y^T}$$

$\frac{\partial L}{\partial Z}$ 的形状为 $[M, K]$, Y^T 的形状为 $[K, N]$,则 $\frac{\partial L}{\partial X}$ 的形状为 $[M, N]$,与 X 形状完成相同.

计算 $\frac{\partial L}{\partial y_{ij}}$:
因为 y_{ij} 出现在 z 的第 j 列的所有元素中，而 z 的每一列共有 M 个元素，所以:

$$\frac{\partial L}{\partial y_{ij}} = \sum_{k=1}^{M} \frac{\partial L}{\partial z_{kj}} \frac{\partial z_{kj}}{\partial y_{ij}}$$

我们先计算 $\frac{\partial z_{kj}}{\partial y_{ij}}$:

$$\frac{\partial z_{kj}}{\partial y_{ij}} = \frac{\partial \left(\sum_{n=1}^{N} x_{kn} y_{nj}\right)}{\partial y_{ij}}$$

只有当 $n == i$ 时，右侧偏导数才不为零，因此:

$$\frac{\partial z_{kj}}{\partial y_{ij}} = x_{ki}$$

代入上式:

$$\frac{\partial L}{\partial y_{ij}} = \sum_{k=1}^{M} \frac{\partial L}{\partial z_{kj}} x_{ki} = \sum_{k=1}^{M} x_{ki} \frac{\partial L}{\partial z_{kj}}$$

将其写成矩阵形式:

$$\boxed{\frac{\partial L}{\partial Y} = X^T \frac{\partial L}{\partial Z}}$$

X^T 的形状为 $[N, M]$, $\frac{\partial L}{\partial Z}$ 的形状为 $[M, K]$,则 $\frac{\partial L}{\partial Y}$ 的形状为 $[N, K]$,与 Y 形状完全相同。

其他类型的推导

如果

$$Z = XY^T$$

令 $\hat{Y} = Y^T$, 则可以转换为标准形式

$$\frac{\partial L}{\partial X} = \left(\frac{\partial L}{\partial Z}\right) \hat{Y}^T = \left(\frac{\partial L}{\partial Z}\right) \left(Y^T\right)^T = \left(\frac{\partial L}{\partial Z}\right) Y$$

$$\frac{\partial L}{\partial \hat{Y}} = X^T \frac{\partial L}{\partial Z}$$

通过变量代换 $\hat{Y} = Y^T$：

$$\frac{\partial L}{\partial Y} = \left(\frac{\partial L}{\partial \hat{Y}}\right)^T = \left(X^T \frac{\partial L}{\partial Z}\right)^T = \left(\frac{\partial L}{\partial Z}\right)^T X$$