Вычисление производной сложной матричной функции

Постановка задачи

Пусть

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}$$

X - матрица измерений. W - матрица весов входных параметров

$$W = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{pmatrix}$$

Будем вычислять производную функции $S = \sigma(\nu(X, W))$, где ν - функция произвдедения матриц.

Определимся со значением функции $\sigma : (3 \times 2) \to (3 \times 2)$. То есть

$$\sigma(\nu(X, W)) = \sigma \begin{pmatrix} X_1W_1 & X_1W_2 \\ X_2W_1 & X_2W_2 \\ X_3W_1 & X_3W_2 \end{pmatrix} = \begin{pmatrix} \sigma(X_1W_1) & \sigma(X_1W_2) \\ \sigma(X_2W_1) & \sigma(X_2W_2) \\ \sigma(X_3W_1) & \sigma(X_3W_2) \end{pmatrix}$$

Где X_i - $i-ая$ строка матрицы X. W_j - $j-ый$ столбец матрицы W. X_iW_j - скалярное произвдение строки на столбец

Определение градиента

Для определния <<градиента>> необходимо ответить на вопрос: "Насколько изменение элемента каждой матрицы X повлияет на конечный результат?". Найдем сумму всех всех элементов $\sigma(\nu(X, W))$, обозначим ее через Λ:

$$\Lambda(\sigma(\nu(X, W))) = \sigma(X_1W_1) + \sigma(X_1W_2) + \sigma(X_2W_1) + \sigma(X_2W_2) + \sigma(X_3W_1) + \sigma(X_3W_2)$$

Найти градиент нашей сложной функции $\sigma(\nu(X, W))$ можно найдя градиент числовой функции Λ:

$$\frac{\partial \Lambda}{\partial u}(X) = \begin{pmatrix} \frac{\partial \Lambda}{\partial u}(x_{11}) & \frac{\partial \Lambda}{\partial u}(x_{12}) & \frac{\partial \Lambda}{\partial u}(x_{13}) \\ \dots & \dots & \dots \\ \frac{\partial \Lambda}{\partial u}(x_{31}) & \dots & \frac{\partial \Lambda}{\partial u}(x_{33}) \end{pmatrix}$$

Вычислять производную сложной функции будем по цепному правилу:

$$\frac{\partial \Lambda}{\partial X}(X) = \frac{\partial \nu}{\partial X}(X, W) \cdot \frac{\partial \sigma}{\partial t}(\nu(X, W)) \cdot \frac{\partial \Lambda}{\partial u}(\sigma(\nu(X, W)))$$

При этом увеличение каждого из компонент суммы Λ на единицу даст увеличение самого Λ на единицу. А значит:

$$\frac{\partial \Lambda}{\partial u}(\dots) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Вторая(слева) компонента произведения равна

$$\left(\frac{\partial \sigma}{\partial t}(X_jW_i) \right) i \in [1, 2] \; j \in [1, 3]$$

Теперь для вычисления $\frac{\partial \Lambda}{\partial u}(N = \nu(X, W))$ необходимо **поэлементно** умножить $\frac{\partial \Lambda}{\partial u}$ на $\frac{\partial \sigma}{\partial t}$:

$$\frac{\partial \Lambda}{\partial u}(N) = \mid \frac{\partial \sigma}{\partial t}(X_j W_i) \cdot 1 \mid$$

Логично, что как и в случае со столбцами(Вычисление производной сложной вектор функции):

$$\frac{\partial \nu}{\partial X}(X, W) = W^T = \begin{pmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \end{pmatrix}$$

А значит искомый градиент выражается по формуле:

$$\frac{\partial \Lambda}{\partial X}(X, W) = \begin{pmatrix} {\sigma'}_{X_1}(X_1 W_1) & {\sigma'}_{X_1}(X_1 W_2) \\ {\sigma'}_{X_2}(X_2 W_1) & {\sigma'}_{X_2}(X_2 W_2) \\ {\sigma'}_{X_3}(X_3 W_1) & {\sigma'}_{X_3}(X_3 W_2) \end{pmatrix} \cdot \begin{pmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \end{pmatrix}$$

Реализация

Реализация вычисления градиента для матричной функции