Туториал по матричному дифференцированию.

Машинное обучение 2017, ФЭН

September 8, 2017

В данной заметке я расскажу вам как быстро и без проблем дифференцировать сложные функции, которые отображают в вектора в вектора, вектора в числа и т.д.
Общий алгоритм выглядит следующим образом:

1. Вычисляем дифференциал функции, исходя из общих правил.

2. Вытаскиваем из дифференциала градиент, матрицу Якоби и все, что захочется.

Преимущество данного подхода состоит в том, что не нужно задумывать о том, кто строка, кто столбец. Также, не возникает промежуточных производных высоких размерностей (тензоров).
Пусть у нас есть функция $f(x)$, её дифференциал мы будем обозначать $df(x)$. При этом, если функция отображает вектор в число, то и дифференциал отображает вектор в число (для других типов отображений работает тоже самое). Таким образом будет удобно себя проверять. Теперь перейдем к правилам вычисления дифференциала, которые почти ничем не отличаются от дифференцирования функции одной переменной.

$$dA = 0$$
$$d(\alpha X) = \alpha(dX)$$
$$d(AXB) = A(dX)B$$
$$d(X^T) = (dX)^T$$
$$d(XY) = (dX)Y + X(dY)$$
$$d\left(\frac{X}{\phi}\right) = \frac{\phi dX - (d\phi)X}{\phi^2}$$

X, Y - дифференцируемые матричные функции, A, B - фиксированные матрицы, α - скаляр, ϕ - дифференцируемая скалярная функция. Как мы видим, константы выносятся (все как в обычном дифференцировании), также не забывайте о порядке! Фиксированные матрицы выносятся в соответствующую сторону. Производные сложных функций берутся как обычно:

$$d_x f(g(x)) = d_g f(g(x)) \cdot d_x g(x)$$

Почти все! Последний инструмент, который нам необходим - это таблица стандартных производных.

$$d(c^T x) = c^T dx$$
$$d(x^T Ax) = x^T(A + A^T)dx$$
$$d(Tr(X)) = Tr(dX)$$
$$d(Det(X)) = Det(X)Tr(X^{-1}dX)$$
$$d(X^{-1}) = -X^{-1}dXX^{-1}$$

Мы готовы.

Пример

Найти дифференциал функции $e^{x^T Ax}$, где A - симметричная матрица.
Решение

$$d(e^{x^T Ax}) = e^{x^T Ax} d(x^T Ax)$$

Мы воспользовались правилом для сложной функции.

$$e^{x^T Ax} d(x^T Ax) = e^{x^T Ax} x^T (A^T + A) dx = 2e^{x^T Ax} x^T A dx$$

Теперь, когда мы умеем находить дифференциалы, искать градиент очень просто. И делается это в соответствии со следующими правилами. В зависимости от того, куда и откуда работает отображение, дифференциал функции может быть представлен следующим образом:

- вектор в скаляр: $\nabla f(x)^T dx$
- вектор в вектор: $J_f dx$, где J_f – матрица Якоби.
- матрица в скаляр: $Tr(\nabla f(X)^T dX)$

Вернемся к нашему примеру. Функция отображает вектор в скаляр, значит нам нужно первое правило. Посмотрим на наш дифференциал:

$$2e^{x^T Ax} x^T A dx$$

Но он уже в нужном виде! Отсюда получаем, что $\nabla f(x)^T = 2e^{x^T Ax} x^T A$. Получаем, что градиент равен:

$$\nabla f(x) = 2e^{x^T Ax} A^T x$$