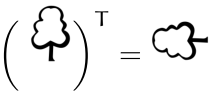
Тятя! Тятя! Нейросети заменили продавца!

Ппилиф Ульянкин

Листочек 4: матричное дифФфФфФференцирование



«Джек и бобовый стебель» (1890)

Упражнение 1

Найдите следующие производные:

а. $f(x) = x^2$, где x скаляр

б. $f(x) = a^T x$, где a и x векторы размера $1 \times n$

в. $f(x) = x^T A x$, где x вектор размера $1 \times n$, A матрица размера $n \times n$

г. $f(x) = \ln(x^T A x)$, где x вектор размера $1 \times n$, A матрица размера $n \times n$

д. $f(x) = a^T X A X a$, где x вектор размера $1 \times n$, A матрица размера $n \times n$

е. $f(x) = x x^T x$, где x вектор размера $1 \times n$

Упражнение 2

Давайте пополним таблицу дифференциалов несколькими новыми функциями, специфичными для матриц. Найдём матричные дифференциалы функций:

а. $f(X) = X^{-1}$, где матрица X размера $n \times n$

б. $f(X) = \det X$, где матрица X размера $n \times n$

в. $f(X) = \mathrm{tr}(X)$, где матрица X размера $n \times n$

г. Ещё больше матричных производных можно найти в книге The Matrix Cookbook[1]

[1] https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf

Упражнение 3

Рассмотрим задачу линейной регресии

$$L(w) = (y - Xw)^T(y - Xw) \to \min_w.$$

а. Найдите $L(w)$, выведите формулу для оптимального w.

б. Как выглядит шаг градиентного спуска в матричном виде?

в. Найдите $d^2L(w)$. Убедитесь, что мы действительно в точке минимума.

Упражнение 4

Найдите следующие производные:

а. $f(X) = \operatorname{tr}(AXB)$, где матрица A размера $p \times m$, матрица B размера $n \times p$, матрица X размера $m \times n$.

б. $f(X) = \operatorname{tr}(AX^TX)$, где матрица A размера $n \times n$, матрица X размера $m \times n$.

в. $f(X) = \ln \det X$

г. $f(X) = \operatorname{tr}(AX^TXBX^{-T})$

д. $f(X) = \det(X^TAX)$

е. $f(x) = x^TAb$, где матрица A размера $n \times n$, вектора x и b размера $n \times 1$.

ж. $f(A) = x^TAb$.

Упражнение 5

В случае Ridge-регрессии минимизируется функция со штрафом:

$$L(w) = (y - Xw)^T(y - Xw) + \lambda w^Tw,$$

где λ — положительный параметр, штрафующий функцию за слишком большие значения w.

а. Найдите $dL(w)$, выведите формулу для оптимального w.

б. Как выглядит шаг градиентного спуска в матричном виде?

в. Найдите $d^2L(w)$. Убедитесь, что мы действительно в точке минимума.

Упражнение 6

Пусть x_i — вектор-столбец $k \times 1$, y_i — скаляр, равный $+1$ или -1, w — вектор-столбец размера $k \times 1$. Рассмотрим логистическую функцию потерь с l_2 регуляризацией

$$L(w) = \sum_{i=1}^{n} \ln(1 + \exp(-y_i x_i^T w)) + \lambda w^T w$$

а. Найдите dL;

б. Найдите вектор-столбец ∇L.

в. Как для этой функции потерь выглядит шаг градиентного спуска в матричном виде?

Упражнение 7

Упражняемся в матричном методе максимального правдоподобия. Допустим, что выборка размера n пришла к нам из многомерного нормального распределения с неизвестными вектором средних μ и ковариационной матрицей Σ. В этом задании нужно найти оценки максимального правдоподобия для $\hat{\mu}$ и $\hat{\Sigma}$. Обратите внимание, что выборкой здесь будет не $x_1, \ldots, x_n$, а

$$\begin{pmatrix} x_{11}, \ldots, x_{n1} \\ \ldots \\ x_{n1}, \ldots, x_{nm} \end{pmatrix}$$

Упражнение 8

Найдите симметричную матрицу X наиболее близкую к матрице A по норме Фробениуса, $\sum_{i,j}(x_{ij} - a_{ij})^2$. Тут мы просто из каждого элемента вычитаем каждый и смотрим на сумму квадратов таких разностей. То есть решите задачку условной матричной минимизации

$$\begin{cases} ||X - A||^2 \to \min_A \\ X^T = X \end{cases}$$

Hint: Надо будет выписать Лагранджиан. А ещё пригодится тот факт, что $\sum_{i,j}(x_{ij} - a_{ij})^2 = ||X - A||^2 = \mathrm{tr}((X - A)^T(X - A))$.