Тятя! Тятя! Наши сети притащили мертвеца!

Листочек с задачками №2: Матричное дифФфФфФфириенцирование*

https://github.com/FUlyankin/neural_nets_prob

РАНХ
осень 2020

Цитата про матрицы

Из фильма «Я, робот» (2004)

Упражнение 1

В этой задачке нужно просто найти НЕмного разных производных:

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

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

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

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

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

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

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

Упражнение 2

В этой задачке нужно просто найти много разных производных:

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

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

в. $f(X) = \ln \det X$

г. $f(X) = \ln AX^{-1}B$

д. $f(X) = tr(AX^T X B X^{-T})$

*Часть задач взята из прототипа задачника по ML Демешева, часть из Конспектов Соколова е. $f(X) = \ln\det(X^T A X)$

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

з. $f(A) = x^T A b$.

Упражнение 3

Рассмотрим задачу линейной регресии

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

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

б. Как выглядит шаг градиентного спуска в матричном виде?

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

Упражнение 4

В случае Ridge-регрессии минимизируется функция

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

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

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

б. Как выглядит шаг градиентного спуска в матричном виде?

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

В случае Lasso-регрессии мы имеем дело с функцией

$$Q(w) = (y - Xw)^T(y - Xw) + \lambda|w|,$$

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

б. Как выглядит шаг градиентного спуска в матричном виде?

Упражнение 5

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

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

а. Найдите dQ;

б. Найдите вектор-столбец ∇Q.

Упражнение 6

Упражняемся в матричном методе максимального правдоподобия! Допустим, что векторы $X_1, \ldots, X_m$ выбраны 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 $X_1, \ldots, X_m$, а

$$\begin{pmatrix} x_{11}, \ldots, x_{m1} \\ \ldots \\ x_{1n}, \ldots, x_{mn} \end{pmatrix}$$

Упражнение 7

Найдите симметричную матрицу X наиболее близкую а 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 $\sum_{i,j}(x_{ij} - a_{ij})^2 = ||X - A||^2 = tr((X - A)^T(X - A))$. То, что это так мы доказали на семинаре :) Вспоминайте!