Математика для Data Science. Линейная алгебра.
Условия задач

Содержание

Замечание. Вот этим цветом отмечены ссылки на страницы внутри этого файла.

Матричное дифференцирование

Задача 1

Пусть $A \in \mathbb{R}^{n \times n}$. Найдите $\nabla_x \ x^T Ax$. Ответ постарайтесь записать в матричном виде.

Задача 2

Определение. *Следом* матрицы $A \in \mathbb{R}^{n \times n}$ называется число $trA = \sum\limits_{i=1}^{n} a_{ii}$.

Пусть $A \in \mathbb{R}^{n \times n}$ найдите $\nabla_A tr(AB)$. Ответ также попробуйте записать в матричном виде.

Задача 3

Пусть $x \in \mathbb{R}^n, A \in \mathbb{R}^{n \times m}, y \in \mathbb{R}^m$. Найдите $\nabla_A x^T Ay$.

Замечание. Тут будет полезно циклическое свойство следа матрицы: $tr(ABC) = tr(BCA) = tr(CAB)$ для случаев, когда размеры матриц позволяют делать такие циклические сдвиги.

Подсказка. Воспользуйтесь предыдущей задачей.

Точное решение для линейной регрессии с MSE

Задача 1

Чтобы найти градиент $(y - Xw)^T(y - Xw)$ по w, для начала нужно раскрыть скобки в выражении. Эта задача подскажет, как.

Докажите, что для всех матриц $A, B \in \mathbb{R}^{n \times m}$ и $C \in \mathbb{R}^{m \times k}$:

1. $(A + B)^T = A^T + B^T$
2. $(AC)^T = C^T A^T$

Чему будет равняться $(ACD)^T$ для $D \in \mathbb{R}^{k \times r}$? $(A_1 A_2 \cdot \dots \cdot A_l)^T$? Размеры матриц $A_1, A_2, \cdot \dots, \cdot A_l$ согласованы.

Замечание. Запись вида "$A \in \mathbb{R}^{n \times m}$" означает, что матрица A составлена из действительных чисел и у неё n строк и m столбцов.

Задача 2

Пусть ранг матрицы $A \in \mathbb{R}^{m \times n}$ равен n. Докажите, что матрица $A^T A$ имеет полный ранг.

Подсказки.

1. Если ранг $A^T A$ меньше n, то существует ненулевой вектор v такой, что $A^T Av = 0$. Покажите, что из этого следует, что столбцы матрицы A будут линейно зависимы.
2. $A^T Av = 0 \implies v^T AA^T v = 0$

Backpropagation в общем случае

Задача 1

Рассмотрим нейронную сеть с одним линейным слоем и без функции активации (другими словами, с тождественной функцией активации). Веса линейного слоя задаются матрицей $W \in \mathbb{R}^{n \times m}$, m — размер входа с учетом нейрона сдвига, n — размер выхода.

Пусть функция потерь $L(y, \hat{y})$ — это длина вектора $(y - \hat{y})$, то есть скалярное произведение $\langle y - \hat{y}, y - \hat{y} \rangle$. Найдите $\nabla_x L$ двумя способами:

1. Через матрицу Якоби для композиции сложных функций

2. И напрямую: раскрыв скобки и применив полученные нами ранее правила дифференцирования.

Сравните ответы.

Замечание. Нам интереснее искать в этом случае $\nabla_W L$, поскольку оптимизируются в нейросети именно веса. Чтобы найти $\nabla_W L$ нужно обобщить на матрицы ряд уже доказанных фактов про векторное дифференцирование и сделать шаги аналогичные поиску $\nabla_W L$. Чтобы не тратить много времени, ограничимся поиском $\nabla_x L$.

Задача 2

Опишите, как будет работать backpropagation для сети ниже и функции потерь L: