Методы оптимизации
Семинар 2. Матрично - векторное дифференцирование

Лобанов Александр Владимирович

Московский физико-технический институт
Факультет инноваций и высоких технологий

lobanov.av@mipt.ru

8 сентября 2022 г.

Что нового?

На github появилась таблица успеваемости:

Таблица успеваемости группы: Б05-027

№	Студент	Посещаемость															Д/З				К/Р		Текущая оценка за курс МО
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1	2	3	4	1	2	
1	Аджима Никита	+																					10
2	Алексеев Максим	+																					10
3	Белый Антон	н																					10
4	Вандакуров Артем	+																					10
5	Грачев Кирилл	+																					10
6	Егоров Гордей	+																					10
7	Мустафин Артём	н																					10
8	Мухаметгалин Артур	н																					10
9	Русскин Николай	н																					10
10	Рябухин Никита	н																					10
11	Турлыбеков Олжас	+																					10
12	Хоружий Тимофей	+																					10
13	Хрол Ариана	+																					10
14	Челышкин Артём	+																					10
15	Яфаров Владимир	+																					10

Обозначения: «+» – присутствовал на семинаре, «н» — отсутствовал на семинаре, «⊕» — сдано Д/З

Дата обновления: 6 сентября 2022 г.

Что нового?

На github появилась таблица успеваемости:

Таблица успеваемости группы: Б05-027

№	Студент	Посещаемость															Д/З				К/Р		Текущая оценка за курс МО
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1	2	3	4	1	2	
1	Аджима Никита	+																					10
2	Алексеев Максим	+																					10
3	Белый Антон	н																					10
4	Вандакуров Артем	+																					10
5	Грачев Кирилл	+																					10
6	Егоров Гордей	+																					10
7	Мустафин Артём	н																					10
8	Мухаметгалин Артур	н																					10
9	Русскин Николай	н																					10
10	Рябухин Никита	н																					10
11	Турлыбеков Олжас	+																					10
12	Хоружий Тимофей	+																					10
13	Хрол Ариана	+																					10
14	Челышкин Артём	+																					10
15	Яфаров Владимир	+																					10

Обозначения: «+» – присутствовал на семинаре, «н» — отсутствовал на семинаре, «⊕» — сдано Д/З

Дата обновления: 6 сентября 2022 г.

Зеленый столбец означает, что Д/З задано и можно сдавать.

Что нового?

На github появилась таблица успеваемости:

Таблица успеваемости группы: Б05-027

№	Студент	Посещаемость															Д/З				К/Р		Текущая оценка за курс МО
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1	2	3	4	1	2	
1	Аджима Никита	+																					10
2	Алексеев Максим	+																					10
3	Белый Антон	н																					10
4	Вандакуров Артем	+																					10
5	Грачев Кирилл	+																					10
6	Егоров Гордей	+																					10
7	Мустафин Артём	н																					10
8	Мухаметгалин Артур	н																					10
9	Русскин Николай	н																					10
10	Рябухин Никита	н																					10
11	Турлыбеков Олжас	+																					10
12	Хоружий Тимофей	+																					10
13	Хрол Ариана	+																					10
14	Челышкин Артём	+																					10
15	Яфаров Владимир	+																					10

Обозначения: «+» – присутствовал на семинаре, «н» — отсутствовал на семинаре, «⊕» — сдано Д/З

Дата обновления: 6 сентября 2022 г.

Красный столбец означает, что deadline наступил и сдавать Д/З нельзя.

Матричное и векторное произведение

Пусть A — матрица $m \times n$, B — матрица $n \times p$, и пусть произведение AB будет выглядить, как:

$$C = AB,$$

Матричное и векторное произведение

Пусть A — матрица $m \times n$, B — матрица $n \times p$, и пусть произведение AB будет выглядить, как:

$$C = AB,$$

тогда C — матрица $m \times p$ с элементом (i, j), заданным следующим образом:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Матричное и векторное произведение

Пусть A — матрица $m \times n$, B — матрица $n \times p$, и пусть произведение AB будет выглядить, как:

$$C = AB,$$

тогда C — матрица $m \times p$ с элементом (i, j), заданным следующим образом:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Пусть A — матрица $m \times n$, x — вектор $n \times 1$, тогда элемент произведения $z = Ax$ задан следующим образом:

$$z_i = \sum_{k=1}^{n} a_{ik} x_k.$$

Матричное и векторное произведение

Полезные свойства

- $C = AB \quad C^\mathsf{T} = B^\mathsf{T} A^\mathsf{T}$
- $AB \neq BA$
- $e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$
- $e^{A+B} \neq e^A e^B$
- $\langle x, Ay \rangle = \langle A^\mathsf{T} x, y \rangle$

Полезные определения и обозначения

Градиент

Пусть $f(x) : \mathbb{R}^n \to \mathbb{R}$, тогда вектор, который содержит все частные производные первого порядка:

$$\nabla f(x) = \frac{df}{dx} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Полезные определения и обозначения

Градиент

Пусть $f(x) : \mathbb{R}^n \to \mathbb{R}$, тогда вектор, который содержит все частные производные первого порядка:

$$\nabla f(x) = \frac{df}{dx} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Гессиан

Пусть $f(x) : \mathbb{R}^n \to \mathbb{R}$, тогда матрица, которая содержит все частные производные второго порядка:

$$f''(x) = \frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$$

Якобиан (Матрица Якоби)

Пусть $f(x) : \mathbb{R}^n \to \mathbb{R}^m$, тогда матрица, составленная из частных производных первого порядка этих функций:

$$f'(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

Полезные определения и обозначения

Якобиан (Матрица Якоби)

Пусть $f(x) : \mathbb{R}^n \to \mathbb{R}^m$, тогда матрица, составленная из частных производных первого порядка этих функций:

$$f'(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

Промежуточный Итог

$$f(x) : X \to Y; \qquad \frac{\partial f(x)}{\partial x} \in G$$

X	Y	G	Название
$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	$f'(x)$ (Производная)
$\mathbb{R}^n$	$\mathbb{R}$	$\mathbb{R}^n$	$\frac{\partial f}{\partial x_i}$ (Градиент)
$\mathbb{R}^n$	$\mathbb{R}^m$	$\mathbb{R}^{m \times n}$	$\frac{\partial f_i}{\partial x_j}$ (Матрица Якоби)
$\mathbb{R}^{m \times n}$	$\mathbb{R}$	$\mathbb{R}^{m \times n}$	$\frac{\partial f}{\partial x_{ij}}$

Основная Техника (Базовый Подход)

Матричное представление функции

$$f(x) = c^\top x$$

Основная Техника (Базовый Подход)

Матричное представление функции

$$f(x) = c^{\mathsf{T}} x$$

Скалярное представление функции

$$f(x) = \sum_{i=1}^{n} c_i x_i$$

Матричное представление функции

$$f(x) = c^{\mathsf{T}} x$$

Скалярное представление функции

$$f(x) = \sum_{i=1}^{n} c_i x_i$$

Вычисление производной

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial \left(\sum_{i=1}^{n} c_i x_i\right)}{\partial x_k}$$

Основная Техника (Базовый Подход)

Матричное представление функции

$$f(x) = c^\mathsf{T} x$$

Скалярное представление функции

$$f(x) = \sum_{i=1}^{n} c_i x_i$$

Вычисление производной

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial \left(\sum_{i=1}^{n} c_i x_i\right)}{\partial x_k}$$

$$\frac{\partial f(x)}{\partial x_k} = c_k$$

Основная Техника (Базовый Подход)

Матричное представление функции

$$f(x) = c^\top x$$

Скалярное представление функции

$$f(x) = \sum_{i=1}^{n} c_i x_i$$

Вычисление производной

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial \left(\sum_{i=1}^{n} c_i x_i\right)}{\partial x_k}$$

$$\frac{\partial f(x)}{\partial x_k} = c_k$$

Матричное представление градиента

$$\nabla f(x) = c$$

Основная Техника (Продвинутый Подход)

Данный подход подразумевает формулировку набора простых правил, позволяющих вычислять производные точно так же, как и в простом подходе.

Дифференциалы

После получения дефференциальной записи df мы можем получить градиент, используя следующую формулу:

$$df = \langle \nabla f(x), dx \rangle.$$

Тогда если у нас есть дифференциал указанной выше формы и нам нужно вычислить вторую производную матричной/векторной функции, мы рассматриваем «старое» dx как константу dx_1, затем вычисляем $d(df)$

$$d^2 f(x) = \langle \nabla^2 f(x) dx_1, dx_2 \rangle = \langle H_f(x) dx_1, dx_2 \rangle$$

Основная Техника (Продвинутый Подход)

Свойства

Пусть A и B будут фиксированными матрицами, α — фиксированный скаляр, X, Y — переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$
- $d(AXB) = A(dX)B$
- $d(X + Y) = dX + dY$
- $d(X)^{-1} = -X^{-1}(dX)X^{-1}$
- $d(X^\top) = (dX)^\top$
- $d(XY) = (dX)Y + X(dY)$
- $d\langle X, Y\rangle = \langle dX, Y\rangle + \langle X, dY\rangle$
- $d(\text{det}X) = \text{det}X\langle X^{-\top}, dX\rangle$
- $d(\text{tr}X) = \langle I, dX\rangle$

Примеры

1. Найти $\nabla f(x)$, если $f(x) = \frac{1}{2}x^{\mathsf{T}}Ax + b^{\mathsf{T}}x + c$.

Примеры

1. Найти $\nabla f(x)$, если $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c$.
2. Найти $\nabla f(x), f''(x)$, если $f(x) = -e^{-x^\top x}$.

Примеры

1. Найти $\nabla f(x)$, если $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c$.
2. Найти $\nabla f(x), f''(x)$, если $f(x) = -e^{-x^\top x}$.
3. Найти $df(X)$, $d^2f(X)$ и $\nabla f(X)$, если $f(X) = \ln(\det(X))$.

Примеры

1. Найти $\nabla f(x)$, если $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c$.
2. Найти $\nabla f(x)$, $f''(x)$, если $f(x) = -e^{-x^\top x}$.
3. Найти $df(X)$, $d^2 f(X)$ и $\nabla f(X)$, если $f(X) = \ln(\det(X))$.
4. Найти $df(x)$, $d^2 f(x)$, а также $\nabla f(x)$ и $\nabla^2 f(x)$, если $f(x) = \ln(1 + e^{\langle a,x\rangle})$.

Домашнее задание

<u>**deadline**: 23:59 (Московское время), 22 сентября.</u>

1. Найти $\nabla f(x)$ и $f''(x)$, если $f(x) = \frac{1}{2}\|Ax - b\|_2^2$, $x \in \mathbb{R}^n$.
2. Найти $\nabla f(x)$ и $f''(x)$, если $f(x) = \frac{1}{p}\|x\|_2^p$, $x \in \mathbb{R}^n \setminus \{0\}$, p — порядковый номер по списку группы (см. табл. успеваемости).
3. Найти $df(X)$ и $\nabla f(X)$, если $f(X) = \|AX - B\|_F$, $X \in \mathbb{R}^{k \times n}$.
4. Найти $df(X)$ и $\nabla f(X)$, если $f(X) = \mathrm{Tr}(AXBX^{-1})$, $X \in \mathbb{R}^{n \times n}$, $\det(X) \neq 0$.
5. Найти аналитическое выражение градиента, гессиана и сравнить с ответами, полученными любой системой автоматической дифференциации (autograd / jax / pytorch / tensorflow) для следующих функций:

1) $\boxed{f(x) = \dfrac{1}{2}x^\top Ax + b^\top x + c}$ 2) $\boxed{f(x) = \dfrac{1}{2}\|Ax - b\|_2^2}$ 3) $\boxed{f(x) = \ln(1 + \exp(\langle a, x\rangle)).}$