



Семинар 1

Матричное

дифференцирование

Титов Владислав

vladislav.v.titov@yandex.ru

ВШЭ, 2020

ЗАДАЧА 1.1

$$\frac{\partial a^T x}{\partial x_i} = \frac{\partial}{\partial x_i} \sum_j a_j x_j$$

ЗАДАЧА 1.1

$$\frac{\partial a^T x}{\partial x_i} = \frac{\partial}{\partial x_i} \sum_j a_j x_j = a_i$$

ЗАДАЧА 1.1

Поэтому

$$\nabla_x a^T x = a$$

ЗАДАЧА 1.1

Так как скалярное произведение симметрично $a^T x = x^T a$.
Следовательно,

$$\nabla_x x^T a = a$$

ЗАДАЧА 1.2

Воспользуемся теоремой Лапласа о разложении определителя по строке:

$$\frac{\partial}{\partial A_{ij}} \det A = \frac{\partial}{\partial A_{ij}} \left[\sum_k (-1)^{i+k} A_{ik} M_{ik} \right] = (-1)^{i+j} M_{ij},$$

где M_{ik} - дополнительный минор матрицы A .

ЗАДАЧА 1.2

Вспомним формулу для элементов обратной матрицы:

$$(A^{-1})_{ij} = \frac{1}{\det A} (-1)^{i+j} M_{ji}$$

ЗАДАЧА 1.2

Подставляя выражение для дополнительного минора,
получаем ответ:

$$\nabla_A \det A = (\det A) A^{-T}$$

ЗАДАЧА 1.3

$$\frac{\partial}{\partial A_{ij}} \text{tr}(AB) = \frac{\partial}{\partial A_{ij}} \sum_k (AB)_{kk}$$

ЗАДАЧА 1.3

$$\frac{\partial}{\partial A_{ij}} \text{tr}(AB) = \frac{\partial}{\partial A_{ij}} \sum_k (AB)_{kk} = \frac{\partial}{\partial A_{ij}} \sum_{k,l} A_{kl} B_{lk}$$

ЗАДАЧА 1.3

$$\frac{\partial}{\partial A_{ij}} \text{tr}(AB) = \frac{\partial}{\partial A_{ij}} \sum_k (AB)_{kk} = \frac{\partial}{\partial A_{ij}} \sum_{k,l} A_{kl} B_{lk} = B_{ji}$$

ЗАДАЧА 1.3

Таким образом, получаем ответ:

$$\nabla_A \text{tr}(AB) = B^T$$

ЗАДАЧА 1.4

Так как произведение вектора на матрицу на вектор в результате дает число:

$$\nabla_A x^T A y = \nabla_A \text{tr}(x^T A y)$$

ЗАДАЧА 1.4

Вспомним циклическое свойство следа матрицы (для матриц подходящего размера):

$$\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)$$

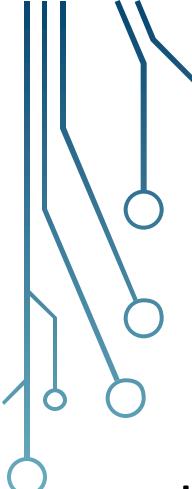
Применим:

$$\nabla_A x^T A y = \nabla_A \text{tr}(x^T A y) = \nabla_A \text{tr}(A y x^T)$$

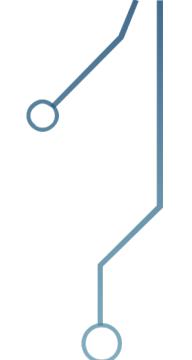
ЗАДАЧА 1.4

Воспользуемся результатом предыдущей задачи:

$$\nabla_A x^T A y = \nabla_A \text{tr}(x^T A y) = \nabla_A \text{tr}(A y x^T) = x y^T$$

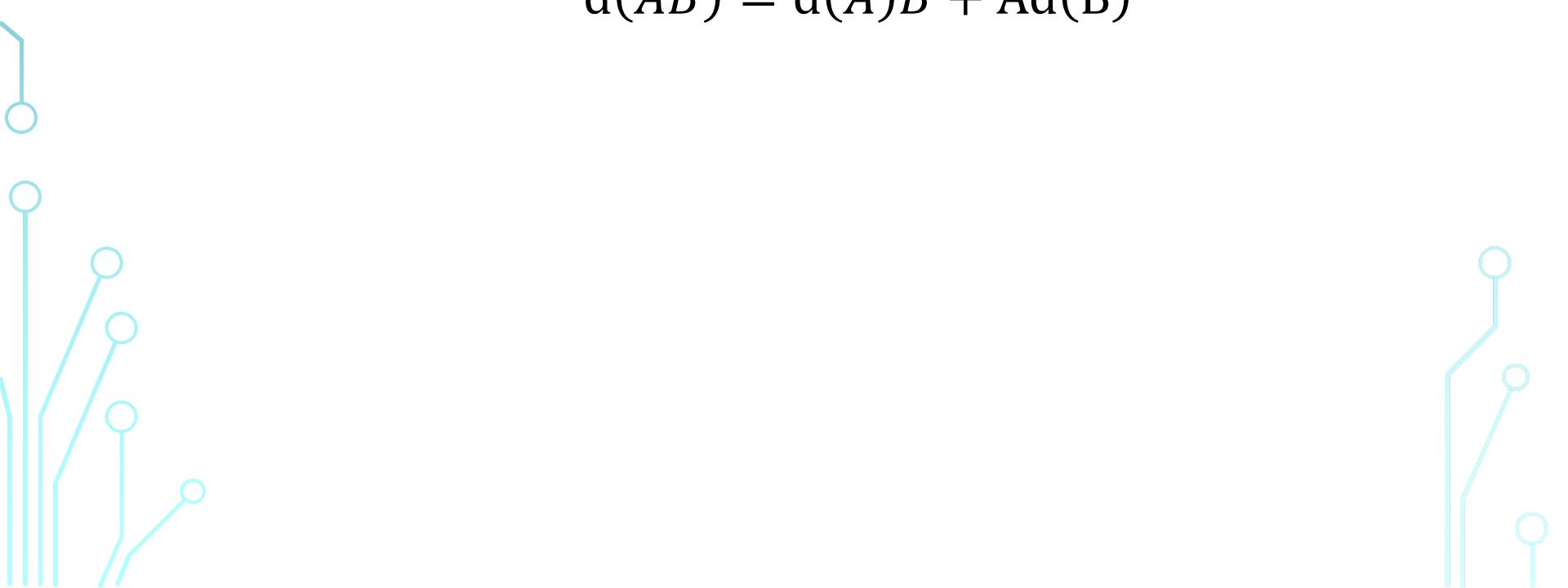


ДИФФЕРЕНЦИАЛ ПРОИЗВЕДЕНИЯ МАТРИЦ


$$A \in \mathbb{R}^{p \times n}, B \in \mathbb{R}^{n \times q}$$

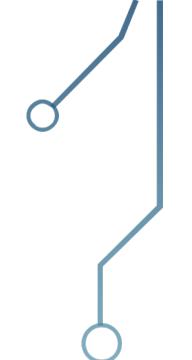
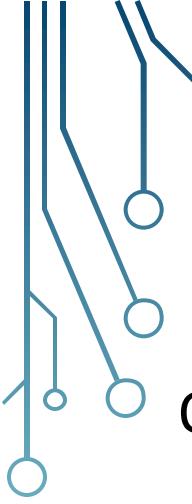
Необходимо доказать:

$$d(AB) = d(A)B + Ad(B)$$



ДИФФЕРЕНЦИАЛ ПРОИЗВЕДЕНИЯ МАТРИЦ

$$d(AB) = \sum_{ij} \frac{\partial(AB)}{\partial a_{ij}} da_{ij} + \sum_{ij} \frac{\partial(AB)}{\partial b_{ij}} db_{ij}$$

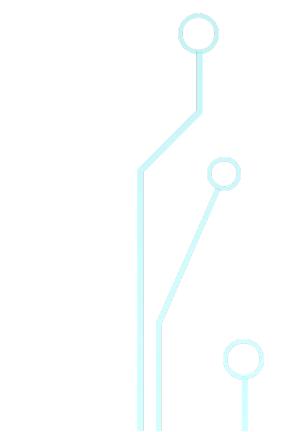
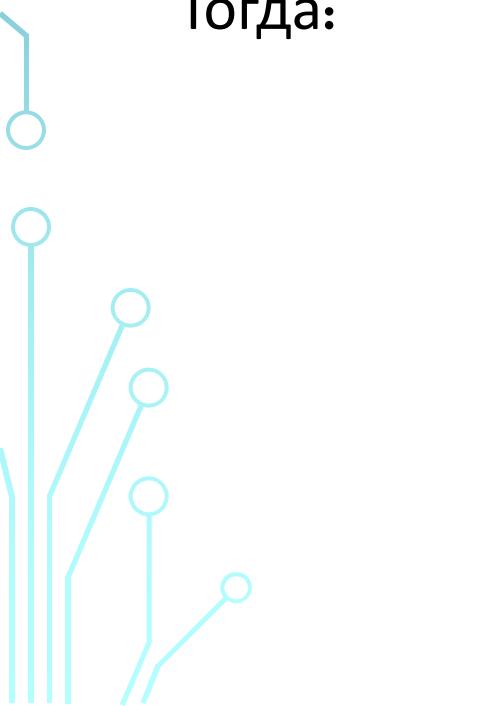


ДИФФЕРЕНЦИАЛ ПРОИЗВЕДЕНИЯ МАТРИЦ

Обозначим:

$$W = AB$$

Тогда:

$$w_{kl} = \sum_{r=1}^n a_{kr} b_{rl}$$


ДИФФЕРЕНЦИАЛ ПРОИЗВЕДЕНИЯ МАТРИЦ

$$\frac{\partial w_{kl}}{\partial a_{ij}} = \begin{cases} 0, & k \neq i \\ b_{jl}, & k = i \end{cases}$$

ДИФФЕРЕНЦИАЛ ПРОИЗВЕДЕНИЯ МАТРИЦ

$$\frac{\partial(AB)}{\partial a_{ij}} = \left(\frac{\partial w_{kl}}{\partial a_{ij}} \right)_{pq} = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ b_{j1} & \ddots & b_{jq} \\ \vdots & \dots & \vdots \\ 0 & \dots & 0 \end{pmatrix}_{\begin{matrix} 1 \\ i \\ \vdots \\ p \end{matrix}}$$

ДИФФЕРЕНЦИАЛ ПРОИЗВЕДЕНИЯ МАТРИЦ

Тогда:

$$\sum_{ij} \frac{\partial(AB)}{\partial a_{ij}} da_{ij} = \sum_{ij} \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ b_{j1}da_{ij} & \ddots & b_{jq}da_{ij} \\ \vdots & \cdots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}_{\begin{matrix} 1 \\ i \\ p \end{matrix}} =$$

$$= \sum_j \begin{pmatrix} b_{j1}da_{1j} & \cdots & b_{jq}da_{1j} \\ \vdots & & \vdots \\ b_{j1}da_{ij} & \ddots & b_{jq}da_{ij} \\ \vdots & & \vdots \\ b_{j1}da_{pj} & \cdots & b_{jq}da_{pj} \end{pmatrix}_{\begin{matrix} 1 \\ i \\ p \end{matrix}}$$

ДИФФЕРЕНЦИАЛ ПРОИЗВЕДЕНИЯ МАТРИЦ

$$\sum_{ij} \frac{\partial(AB)}{\partial a_{ij}} da_{ij} = \begin{pmatrix} \sum_j b_{j1} da_{1j} & \cdots & \sum_j b_{jq} da_{1j} \\ \vdots & & \vdots \\ \sum_j b_{j1} da_{ij} & \ddots & \sum_j b_{jq} da_{ij} \\ \vdots & & \vdots \\ \sum_j b_{j1} da_{pj} & \cdots & \sum_j b_{jq} da_{pj} \end{pmatrix}_{\begin{matrix} 1 \\ \vdots \\ i \\ \vdots \\ p \end{matrix}}$$

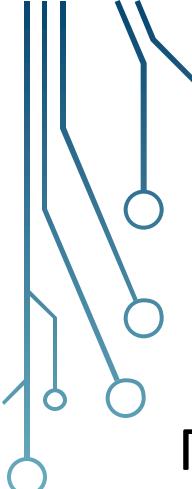
ДИФФЕРЕНЦИАЛ ПРОИЗВЕДЕНИЯ МАТРИЦ

Посмотрим, чему должно быть равно $d(A)B$:

$$d(A)B = \begin{pmatrix} da_{11} & \cdots & da_{1n} \\ \vdots & \ddots & \vdots \\ da_{p1} & \cdots & da_{pn} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1q} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nq} \end{pmatrix}$$

ДИФФЕРЕНЦИАЛ ПРОИЗВЕДЕНИЯ МАТРИЦ

$$d(A)B = \begin{pmatrix} \sum_j da_{1j} b_{j1} & \cdots & \sum_j da_{1j} b_{jq} \\ \vdots & & \vdots \\ \sum_j da_{ij} b_{j1} & \ddots & \sum_j da_{ij} b_{jq} \\ \vdots & & \vdots \\ \sum_j da_{pj} b_{j1} & \cdots & \sum_j da_{pj} b_{jq} \end{pmatrix} \begin{matrix} 1 \\ \vdots \\ i \\ \vdots \\ p \end{matrix}$$



ДИФФЕРЕНЦИАЛ ПРОИЗВЕДЕНИЯ МАТРИЦ

Получилось, что $d(A)B = \sum_{ij} \frac{\partial(AB)}{\partial a_{ij}} da_{ij}$.

Аналогично можно доказать, что $Ad(B) = \sum_{ij} \frac{\partial(AB)}{\partial b_{ij}} db_{ij}$.

