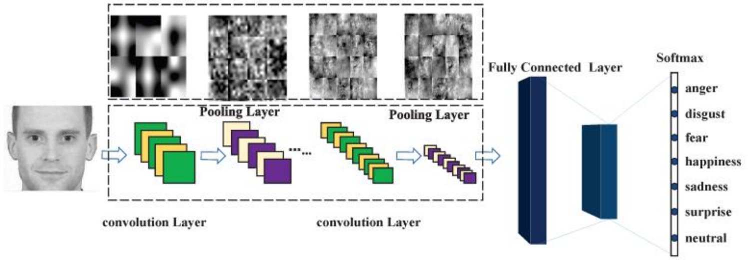
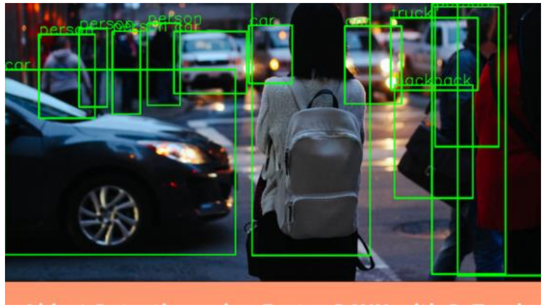
# Основы искусственного интеллекта







## Лекция 2

## Основы прикладной математики для машинного обучения

к.ф.-м.н., доцент кафедры ИСиЦТ
Корнаева Е.П.

# Основы прикладной математики для машинного обучения

## Основные величины в линейной алгебре (ЛА) и операции над ними

**Матрица ($\mathrm{M_a}$)** – прямоугольная таблица чисел с компонентами $\mathrm{a}_{ij}$

$\mathrm{M_a} = \left(\left(\mathrm{a}_{ij}\right)\right)$ или $\mathrm{M_a} = \left(\mathrm{a}_{ij}\right)$

Понятие ***тензорной*** величины в ЛА отождествляется с ее физическими и геометрическими аналогами

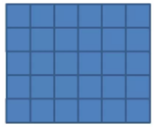
**Тензором нулевого ранга** (скаляром) в N-мерном пространстве называется математическая величина, характеризуемая одной ($N^0=1$) компонентой a, которая инвариантна относительно поворота множества координат.

Физические аналоги: ?

**Тензором первого ранга** (вектором) в N-мерном пространстве называется математическая величина, характеризуемая N ($N^1=N$) компонентами $\mathrm{a}_i$, которая при повороте множества координат с помощью матрицы косинусов $\mathrm{M}_\alpha = \left(\left(\alpha_{ij}\right)\right)$ преобразуется по закону:

$$a_i' = \alpha_{ij} a_j$$

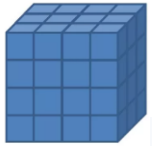
Физические аналоги: ?



**Тензором второго ранга** (тензором) в $N$-мерном пространстве называется математическая величина, характеризуемая $N^2$ компонентами $a_{ij}$, каждая из которых при повороте множества координат с помощью матрицы косинусов преобразуется по закону:

$$\mathrm{a}_{ij}' = \alpha_{ip} \alpha_{jq} a_{pq}$$

Физические аналоги: ?



$$\overset{3}{T}_a = (a_{ijk})$$

*В машинном обучении тензор отождествляется с многомерным массивом**!!!*

* Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с анг. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652с.

# Основы прикладной математики для машинного обучения

## Форма записи. Правило Эйнштейна и исключение Лурье.

$\mathrm{M_a} = (a_{ij})$ – матрица $\mathrm{M_a}$ с компонентами $a_{ij}$
$i$ – номер строки
$j$ – номер столбца

**Правило Эйнштейна:** если в одночлене (например, $a_i b_i$ или $kf_k$, или $c_j d_j$ и т.п.), содержащем индексированные переменные, встречаются повторяющиеся индексы или одинаковые с индексами буквы, то по этим индексам или индексам и буквам производится суммирование.
Например:

**Исключение Лурье:** суммирование в одночлене по повторяющимся индексам или индексам и одинаковым с ними буквам не производится, если такие индексы или буквы в любом виде встречаются с обеих сторон знака равенства (неравенства, тождества и т.п.) в уравнениях или равенствах (неравенствах, тождествах и т.п.).
Например:

# Основы прикладной математики для машинного обучения

## Основные алгебраические операции над матрицами в ЛА

| Операция | Матричная форма | Скалярная форма* | Примечание |
|---|---|---|---|
| Умножение матрицы на скаляр | $\lambda\mathbf{M}_{\mathrm{a}}$ | $\lambda a_{ij}$ | |
| Транспонирование | $\boldsymbol{M}_c = \mathbf{M}_{\mathrm{a}}^T$ | $c_{ij} = a_{ji}$ | |
| Сложение (вычитание) матриц одинаковой размерности | $\mathbf{M}_c = \mathbf{M}_{\mathrm{a}} \pm \mathbf{M}_b$ | $c_{ij} = a_{ij} \pm b_{ij}$ | $\mathbf{M}_{\mathrm{a}[m\times n]}$, $\mathbf{M}_{b[m\times n]}$, $\mathrm{M}_{\mathrm{c}[m\times n]}$ |

# Основы прикладной математики для машинного обучения

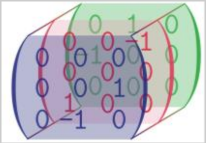
## Основные алгебраические операции над матрицами в ЛА

| Операция | Матричная форма | Скалярная форма* | Примечание |
|---|---|---|---|
| Умножение матриц | $\mathbf{M}_c = \mathbf{M}_\mathrm{a}\mathbf{M}_b$ | $c_{ij} = a_{ik}b_{kj}$ | $\mathbf{M}_{\mathrm{a}[m\times p]}, \mathbf{M}_{b[p\times n]}, \mathbf{M}_{c[m\times n]}$<br>Операция не коммутативна!!! |

Пример: Найти произведение матриц, используя покомпонентную форму записи: $c_{ij} = a_{ik}b_{kj}$

# Основы прикладной математики для машинного обучения

## Основные алгебраические операции над матрицами в ЛА

| Операция | Матричная форма | Скалярная форма* | Примечание |
|---|---|---|---|
| Определитель матрицы | $c = \lvert \mathbf{M}_a \rvert$<br>$\mathrm{c} = \left\lvert \mathbf{M}_{a_{[2\times2]}} \right\rvert$<br>$\mathrm{c} = \left\lvert \mathbf{M}_{a_{[3\times3]}} \right\rvert$ | $c = \epsilon_{ij3}\ a_{1i}a_{2j}$<br>$c = \epsilon_{ijk}\ a_{1i}a_{2j}a_{3k}$ | $c$ – скаляр,<br>$\epsilon_{ijk}$ - символ Леви-Чивиты<br> |

Пример: Найти определитель матрицы [3x3], используя покомпонентную форму записи: $c = \epsilon_{ijk}\ a_{1i}a_{2j}a_{3k}$

# Основы прикладной математики для машинного обучения

## Тензорная и скалярная (покомпонентная) форма записи

| Величина | Тензорная форма | Скалярная форма |
|---|---|---|
| Тензор нулевого ранга (скаляр) | $\overset{0}{\mathbf{T}} = a$ | $a$ |
| Тензор первого ранга (вектор) | $\overset{1}{\mathbf{T}}_{\mathbf{a}} = \vec{a} = [\![a_i]\!]$ | $a_i$<br>$i = 1, .. n$ |
| Тензор второго ранга | $\overset{2}{\mathbf{T}}_{\mathbf{a}} = [\![a_{ij}]\!]$ | $a_{ij}$ |
| Тензор третьего ранга | $\overset{3}{\mathbf{T}}_{a} = [\![a_{ijk}]\!]$ | $a_{ijk}$ |
| … | | |

# Основы прикладной математики для машинного обучения

## Основные операции над тензорами в ЛА

Пример: Найти скалярное произведение векторов

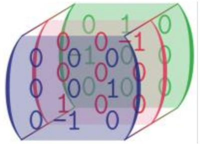
Пример: Найти скалярное произведение тензоров

| Операция | Тензорная форма | Скалярная форма* | Примечание |
|---|---|---|---|
| Умножение тензора на скаляр | $\lambda \mathbf{T}_{\mathrm{a}}$ | $\lambda a$,<br>$\lambda a_i$,<br>$\lambda a_{ij}$,<br>$\ldots$ | для тензора нулевого ранга,<br>для тензора первого ранга,<br>для тензора второго ранга,<br>$\ldots$ |
| Транспонирование | $\boldsymbol{T}_c = \mathbf{T}_{\mathrm{a}}^T$ | $c_{ij} = a_{ji}$ | |
| Сложение (вычитание) тензоров одного ранга одинаковой размерности | $\mathbf{T}_c = \mathbf{T}_{\mathrm{a}} \pm \mathbf{T}_b$ | $c_{ij} = a_{ij} \pm b_{ij}$ | $\mathbf{T}_{\mathrm{a}[m\times n]}$, $\mathbf{T}_{b[m\times n]}$, $\mathbf{T}_{\mathrm{c}[m\times n]}$ аналогично для тензоров любого ранга одинаковой размерности |
| Скалярное произведение тензоров одного ранга | $c = \vec{\mathbf{a}} \cdot \vec{\mathbf{b}}$<br>$c = \mathbf{T}_{\mathrm{a}} \cdot \mathbf{T}_b$<br>$\ldots$ | $c = a_i b_i$<br>$c = a_{ij} b_{ij}$<br>$\ldots$ | $c$ - скаляр |
| Скалярное произведение тензоров различного ранга | $\mathbf{T}_c = \mathbf{T}_{\mathrm{a}} \cdot \mathbf{T}_b$<br><br>$\mathbf{T}_d = \mathbf{T}_b \cdot \mathbf{T}_a$ | $r_a = 2, r_b = 3: c_k = a_{ij} b_{ijk}$<br><br>$r_a = 2, r_b = 3: \mathrm{d}_k = b_{ijk} a_{jk}$ | $r = \begin{cases} r_b - r_a \; if \; r_a < r_b \\ r_a - r_b \; if \; r_b < r_a \end{cases}$<br><br>Операция не коммутативна!!! |

# Основы прикладной математики для машинного обучения

## Основные операции над тензорами в ЛА

| Операция | Тензорная форма | Скалярная форма* | Примечание |
|---|---|---|---|
| Тензорное произведение тензоров | $\overset{r}{\mathbf{T}}_{\mathrm{c}} = \overset{r_a}{\mathbf{T}}_{\mathrm{a}} \otimes \overset{r_b}{\mathbf{T}}_{b}$ | $c_{\mathrm{ij}} = a_i b_{\mathrm{j}}$<br>$c_{\mathrm{ijk}} = a_{ij} b_{\mathrm{k}}$<br>$\ldots$ | $r = r_a + r_b$ |
| Векторное произведение тензоров первого | $\vec{\boldsymbol{c}} = \vec{\boldsymbol{a}} \times \vec{\boldsymbol{b}}$ | $c_{\mathrm{i}} = \epsilon_{ijk}\, a_j a_k$ | $\epsilon_{ijk}$ - символ Леви-Чивиты |



Пример: Найти тензорное произведение тензоров

# Основы прикладной математики для машинного обучения

## Понятие градиента скалярной и векторной функции

$f = f(\boldsymbol{X})$ - скалярная функция, $\boldsymbol{X} = (x_i), i = 1,..n$

$\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{X})$ - векторная функция, $\boldsymbol{X} = (x_i), i = 1,..n$

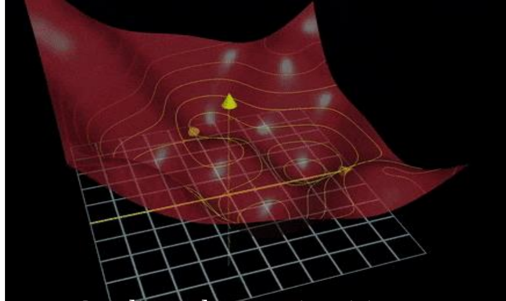
$\boldsymbol{\nabla} = [\frac{\partial}{\partial x_i}]$ – оператор Гамильтона, $i = 1,..n$

| Операция | Тензорная форма | Скалярная форма* | Примечание |
|---|---|---|---|
| Градиент скалярной функции | $\boldsymbol{\nabla} f(x)$ | $\left[\!\left[\frac{\partial f}{\partial x_i}\right]\!\right]$ | $i = 1,..n$<br>результат - вектор |
| Градиент векторной функции | $\boldsymbol{\nabla} \otimes \boldsymbol{F}$ | $\left[\!\left[\frac{\partial F_j}{\partial x_i}\right]\!\right]$ | $i = 1,..n, j = 1,..n$<br>результат – тензор 2го ранга |
| Производная векторной функции по векторному аргументу | $\boldsymbol{F} \otimes \boldsymbol{\nabla}$ | $\left[\!\left[\frac{\partial F_i}{\partial x_j}\right]\!\right]$ | $i = 1,..n, j = 1,..n$<br>результат – тензор 2го ранга |

Пример: градиент функции $f(\boldsymbol{X}) = 2x_1 + x_1 x_2^2$

Пример: градиент функции $\boldsymbol{F}(\boldsymbol{X}) = \left[\!\left[\begin{matrix} 2x_1 + x_2^2 \\ x_1 x_2 \end{matrix}\right]\!\right]$

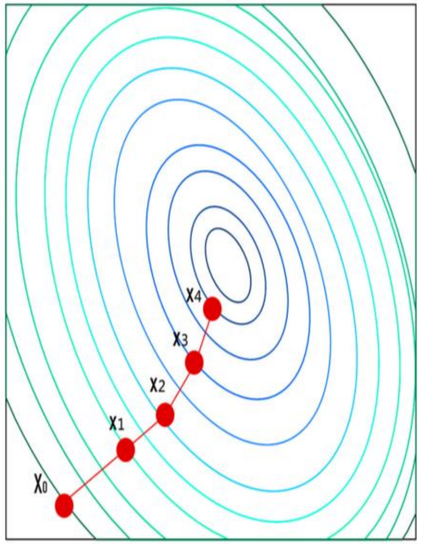
# Основы прикладной математики для машинного обучения


Gradient descent intuition

## Метод градиентного спуска

$f = f(\boldsymbol{X})$ - скалярная функция, $\boldsymbol{X} = \left(x_j\right), i = 1, .. n$

$f_{min} = f(\boldsymbol{X}^*)$ - минимум (экстремум) функции, $\boldsymbol{X}^*$ - точка минимума (экстремума)




**Инициализация:**

- $\alpha$ - шаг градиентного спуска (скорость обучения);
- $\boldsymbol{X}^0 = (x_j^0)$ $-$ начальное приближение точки минимума;
- $\boldsymbol{f}^0 = f(\boldsymbol{X}^0\ )$ $-$ начальное значение экстремума;
- $\boldsymbol{\nabla} f^0 = \left[\!\left[\frac{\partial f(\boldsymbol{X})}{\partial x_j}\right]\!\right]^0$ - значение градиента в начальной точке.

---

**Градиентный спуск:**

- Шаг градиентного спуска:

$$x_j^{l+1} = x_j^l - \alpha \frac{\partial f}{\partial x_j}|_l \text{ или } \boldsymbol{X}^{l+1} = \boldsymbol{X}^l - \alpha\, \boldsymbol{\nabla} f|_l$$

- Компоненты градиента и значение функции на текущем шаге $l$:

$$\boldsymbol{\nabla} f^{l+1} = \left[\!\left[\frac{\partial f(\boldsymbol{X})}{\partial x_j}\right]\!\right]^{l+1}, \qquad \boldsymbol{f}^{l+1} = f(\boldsymbol{X}_j^{l+1})$$

---

**Условие останова:**

$$opt = [\boldsymbol{X}^{l+1}, f^{l+1}]\ if\ \left\|\boldsymbol{\nabla} f^{l+1}\right\| < \varepsilon$$