# Математика для ML

Паточенко Евгений

НИУ ВШЭ

# План занятия

- Функции
- Пределы
- Производные
- Интегралы
- Векторы
- Матрицы

# Функции

Функция — это правило, которое связывает входное значение и возвращает результат некоторых математических действий над ним в виде выходного значения

$f(x)$ – обозначение функции, где $x$ — это входная переменная, а $f$ — это выходное значение

$$f(x) = x^2 + 4$$

Множество входных значений называют *областью определения*, а множество выходных значений — *областью значений*

# Функции

Функция — это правило, которое связывает входное значение и возвращает результат некоторых математических действий над ним в виде выходного значения

$f(x)$ – обозначение функции, где $x$ — это входная переменная, а $f$ — это выходное значение

$$f(x) = x^2 + 4$$

Множество входных значений называют *областью определения*, а множество выходных значений — *областью значений*

***В машинном обучении мы всегда работаем с функциями: выбираем, анализируем, оптимизируем***

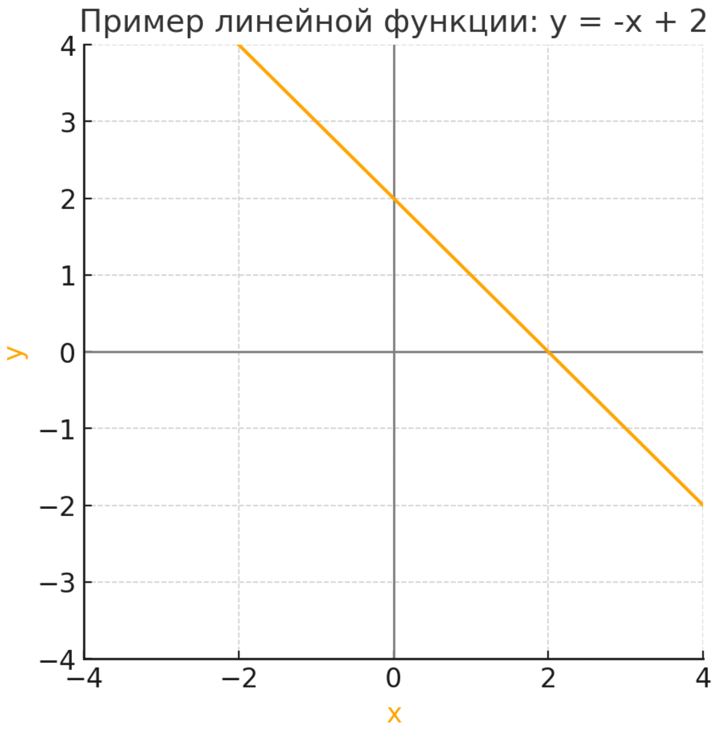
# Функции

| Тип функции | Формула | Пояснения | Для чего нужна | Пример |
|---|---|---|---|---|
| **Линейная функция** | $f(x) = mx + b$ | $m$ — коэффициент наклона, $b$ — сдвиг | Моделирование прямой пропорциональности | Зависимость цены товара от количества. Если $m > 0$, цена растёт, если $m < 0$, падает |
| **Квадратичная функция** | $f(x) = ax^2 + bx + c$ | $a, b, c$ — константы, $a \neq 0$ | Моделирование параболических зависимостей | Затраты на производство товара в зависимости от объёма выпуска |
| **Экспоненциальная функция** | $f(x) = ae^{kx}$ | $a$ — масштаб, $k$ — скорость роста или распада, $e$ — основание натурального логарифма, $k \neq 0$ | Описание процессов роста или распада | Рост популяции бактерий или распад радиоактивных веществ |

# Функции

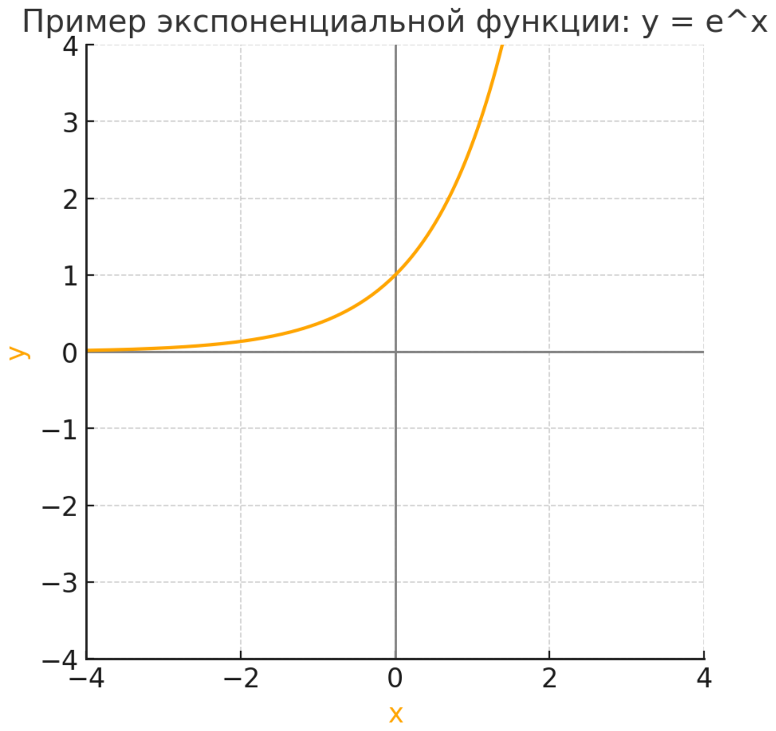
| | | | | |
|---|---|---|---|---|
| **Логарифмическая функция** | $f(x) = a\ln(bx)$ | $a, b$ — константы, $b > 0$ | Анализ процессов с затухающей динамикой | Уровень насыщения в маркетинговых кампаниях: добавление рекламы увеличивает продажи, но с каждым разом эффект от рекламы становится меньше |
| **Гиперболическая функция** | $f(x) = \frac{a}{x}$ | $a$ — масштаб, $x \neq 0$ | Моделирование процессов, где результат убывает обратно пропорционально входным данным | Распределение давления в жидкости по мере удаления от источника |

# Функции



Пример линейной функции: y = -x + 2
y
x

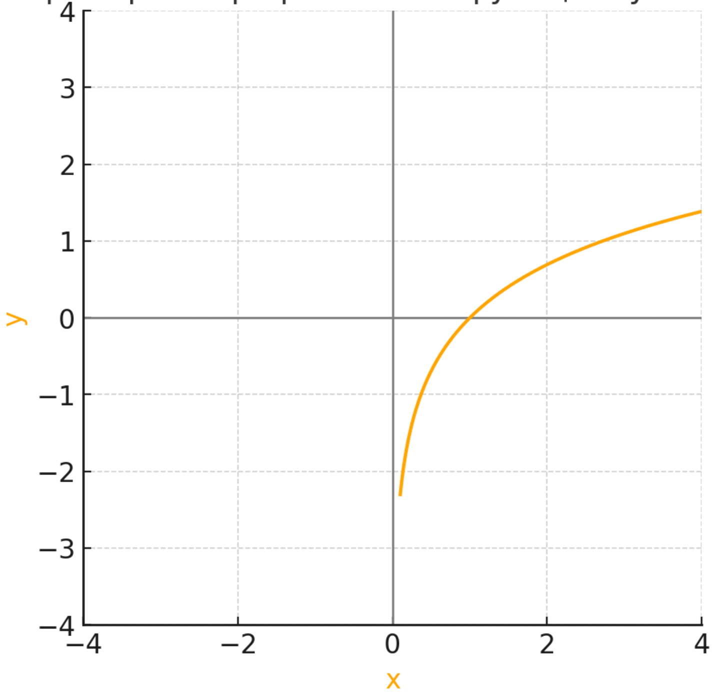

Пример квадратичной функции: y = x² - 1
y
x

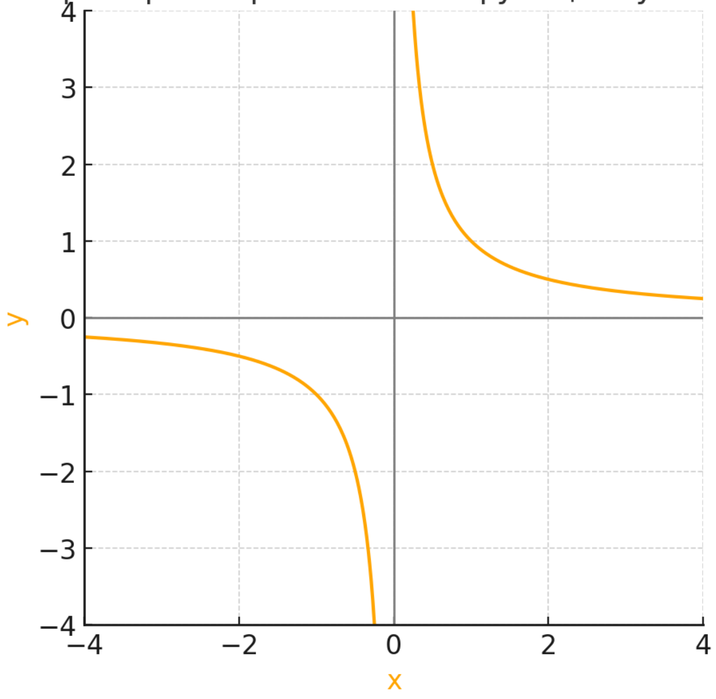
Пример экспоненциальной функции: y = e^x
y
x

# Функции

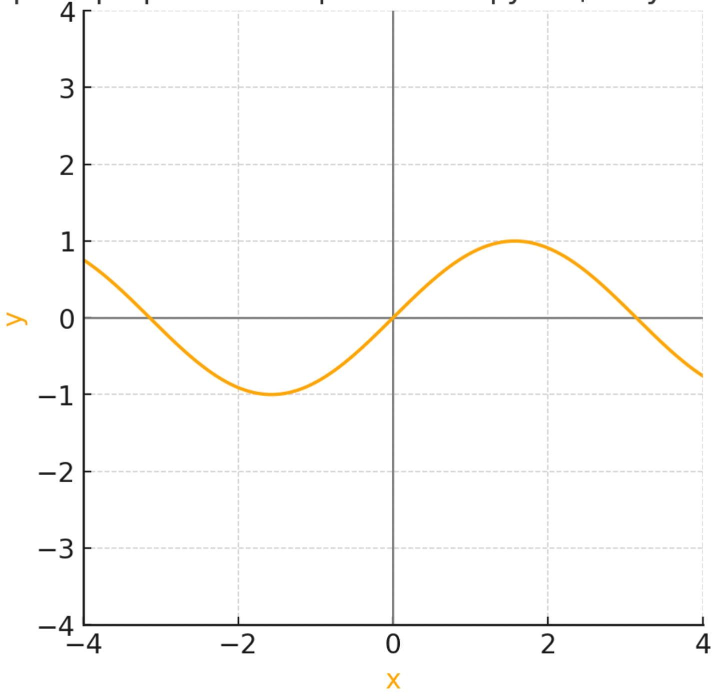
Пример логарифмической функции: y = ln(x)



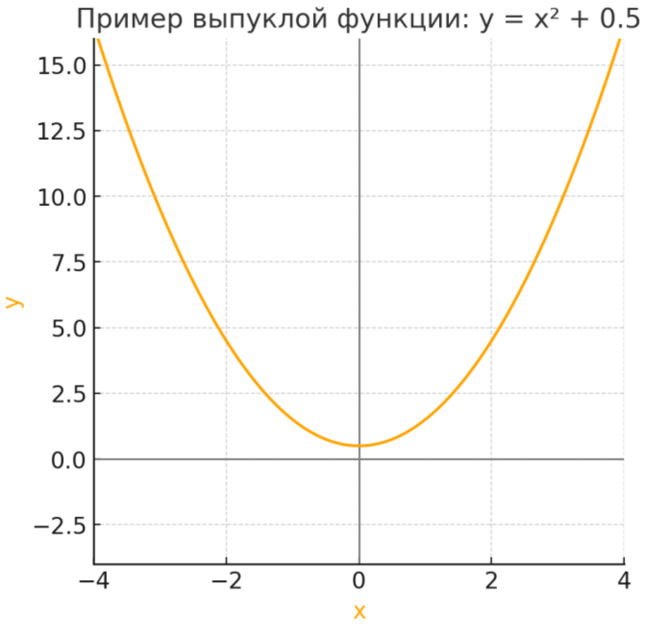

Пример гиперболической функции: y = 1/x




Пример тригонометрической функции: y = sin(x)

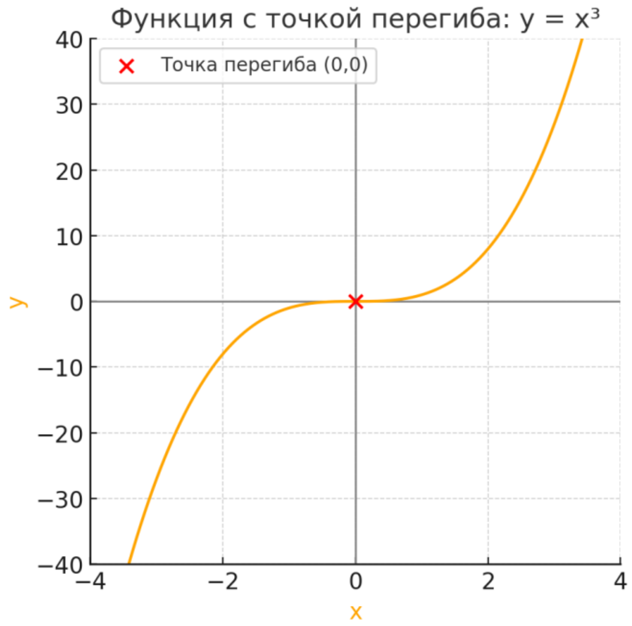


# Функции



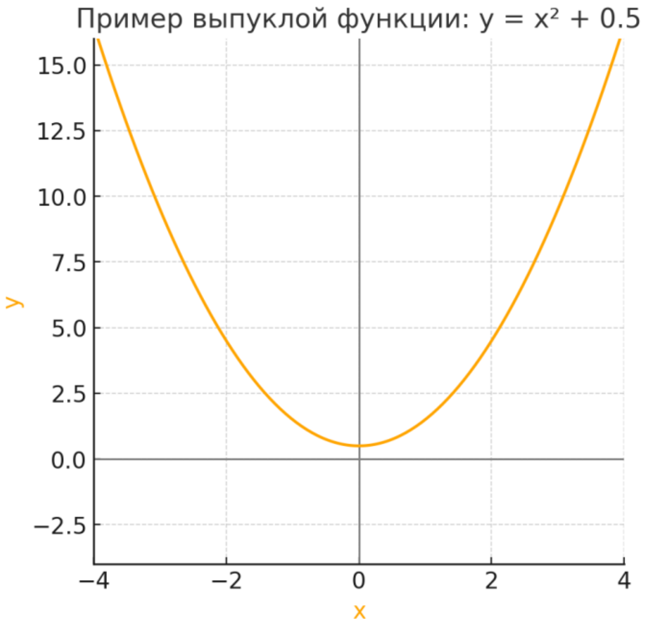

Функция везде выпуклая


Функция везде вогнутая



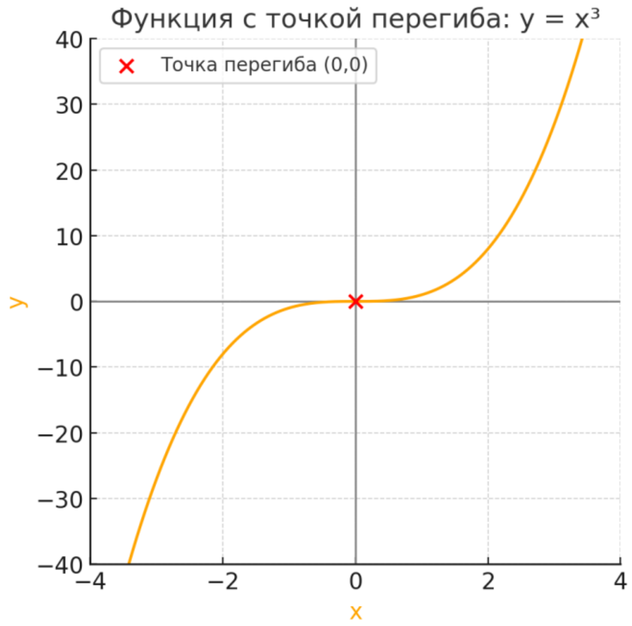

Точка, в которой выпуклость меняется на вогнутость

# Функции




Функция везде выпуклая


Функция везде вогнутая




Точка, в которой выпуклость меняется на вогнутость

**Почему это важно?**

В машинном обучении мы оптимизируем функцию потерь, то есть ищем такую точку, в которой ее значение минимально (в большинстве случаев) или максимально (реже). Если функция **выпуклая**, мы гарантированно ее оптимизируем. Классический пример — линейная регрессия из прошлого занятия.

# Пределы

**Предел** — это значение, к которому стремится функция $f(x)$, в том числе тогда, когда значение $x$ по правилам математики подставить в функцию нельзя

**Обозначение (пример)**

$$\lim_{\boldsymbol{x}\to\infty}\left(\mathbf{1}+\frac{\mathbf{1}}{\boldsymbol{x}}\right)^{\boldsymbol{x}}$$

$\lim$ — предел

$\left(1+\frac{1}{x}\right)^{x}$ — функция

$x \to \infty$ — $x$ на графике будет бесконечно увеличиваться

# Пределы

Самый простой способ узнать предел — подставить число, к которому он стремится, в функцию и посчитать выходное значение

# Пределы

Самый простой способ узнать предел — подставить число, к которому он стремится, в функцию и посчитать выходное значение

Найдем $\lim_{x \to 2}(x + 3)$

# Пределы

Самый простой способ узнать предел — подставить число, к которому он стремится, в функцию и посчитать выходное значение

Найдем $\lim_{x \to 2}(x + 3)$

Подставим $2$ в $(x + 3)$

# Пределы

Самый простой способ узнать предел — подставить число, к которому он стремится, в функцию и посчитать выходное значение

Найдем $\lim_{x \to 2}(x + 3)$

Подставим $2$ в $(x + 3)$

$$\lim_{x \to 2}(x + 3) = 5$$

# Пределы

Подставить число в функцию не всегда возможно — но предел можем посчитать

# Пределы

Подставить число в функцию не всегда возможно — но предел можем посчитать

Найдем $\lim_{x \to 0} \frac{\sin x}{x}$

# Пределы

Подставить число в функцию не всегда возможно — но предел можем посчитать

Найдем $\lim\limits_{x \to 0} \frac{\sin x}{x}$

0 подставить в знаменатель не можем. Но если «подбираться» ближе и ближе — результат стабилизируется около 1

# Пределы

Подставить число в функцию не всегда возможно — но предел можем посчитать

Найдем $\lim\limits_{x \to 0} \frac{\sin x}{x}$

0 подставить в знаменатель не можем. Но если «подбираться» ближе и ближе — результат стабилизируется около 1

$$\lim_{x \to 0} \frac{\sin x}{x} = 1$$

# Пределы

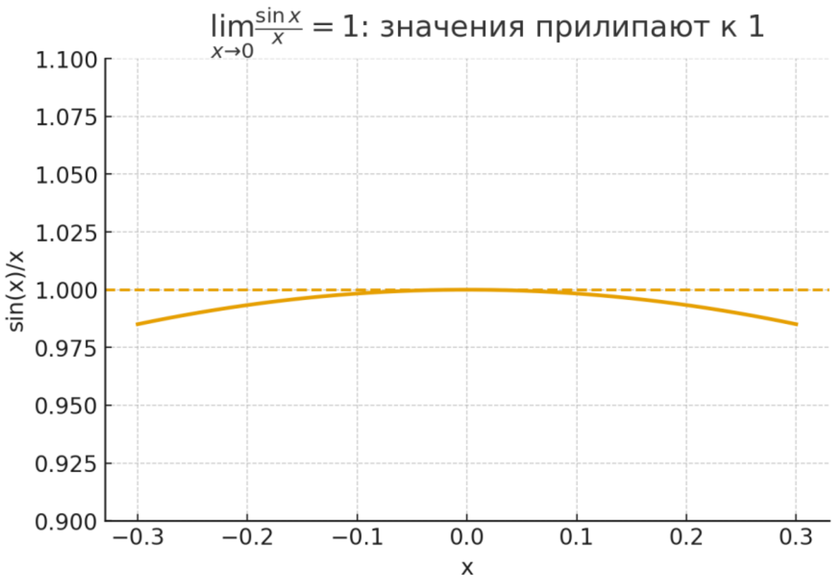
Переменные могут стремиться к бесконечности (или минус бесконечности)

Найдем $\lim\limits_{x \to \infty} \frac{1}{x}$

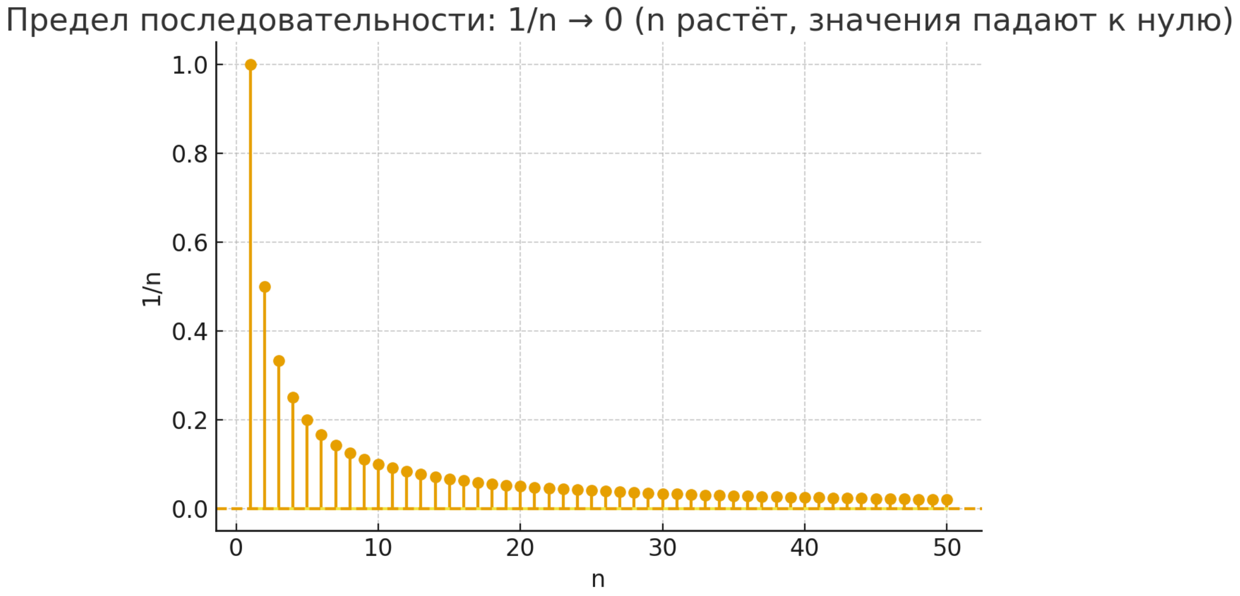
Чем больше $x$, тем сильнее уменьшается значение функции $\frac{1}{x}$, то есть стремится к нулю, хотя нуля никогда не достигнет, так как в $x = 0$ значение функции не определено

$$\lim_{x \to \infty} \frac{1}{x} = 0$$

# Пределы



$\lim_{x\to 0}\frac{\sin x}{x} = 1$: значения прилипают к 1
1.100
1.075
1.050
1.025
1.000
0.975
0.950
0.925
0.900
sin(x)/x
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
x

Предел последовательности: 1/n → 0 (n растёт, значения падают к нулю)
1.0
0.8
0.6
0.4
0.2
0.0
1/n
0
10
20
30
40
50
n

# Пределы

**Почему это важно?**

Многие алгоритмы в машинном обучении строятся на плавных изменениях — например, в градиентном спуске (метод оптимизации функции потерь, о котором будем говорить далее на занятии) шаги становятся все меньше по мере того как мы стремимся к минимуму функции

Чтобы описывать такие процессы, нужно описывать куда стремится функция, если шаг сделать бесконечно маленьким

# Производные

**Производная** — это скорость изменения функции в точке

Насколько выросла функция

$$f'(x_0) = \lim_{x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

Насколько сдвинули по оси x

Строгое определение:

Производная — предел отношения приращения функции в данной точке к приращению аргумента, когда последнее стремится к нулю.

# Производные

**Производная** — это скорость изменения функции в точке

Насколько выросла функция

$$f'(x_0) = \lim_{x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

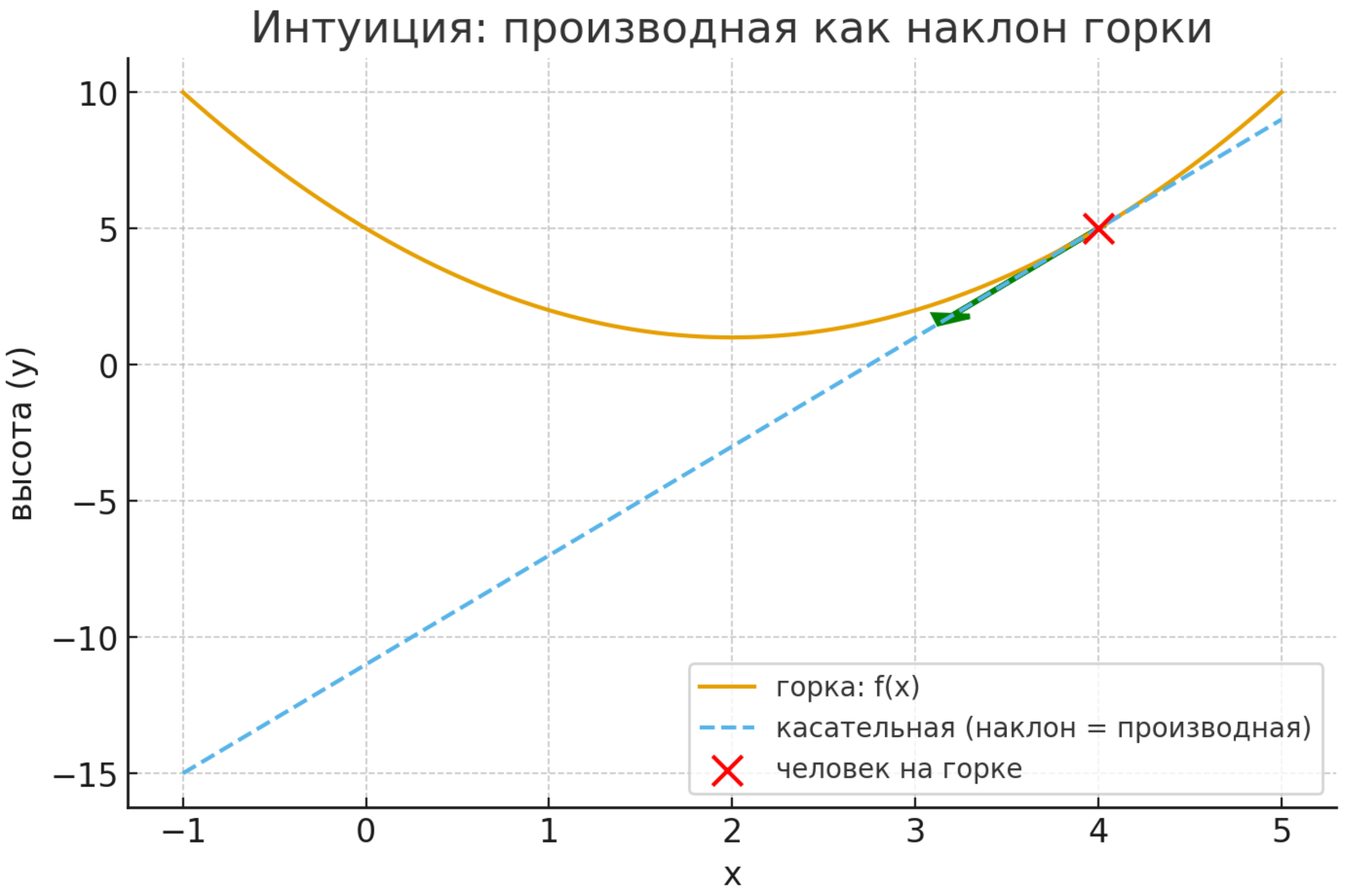
Насколько сдвинули по оси x

Строгое определение:

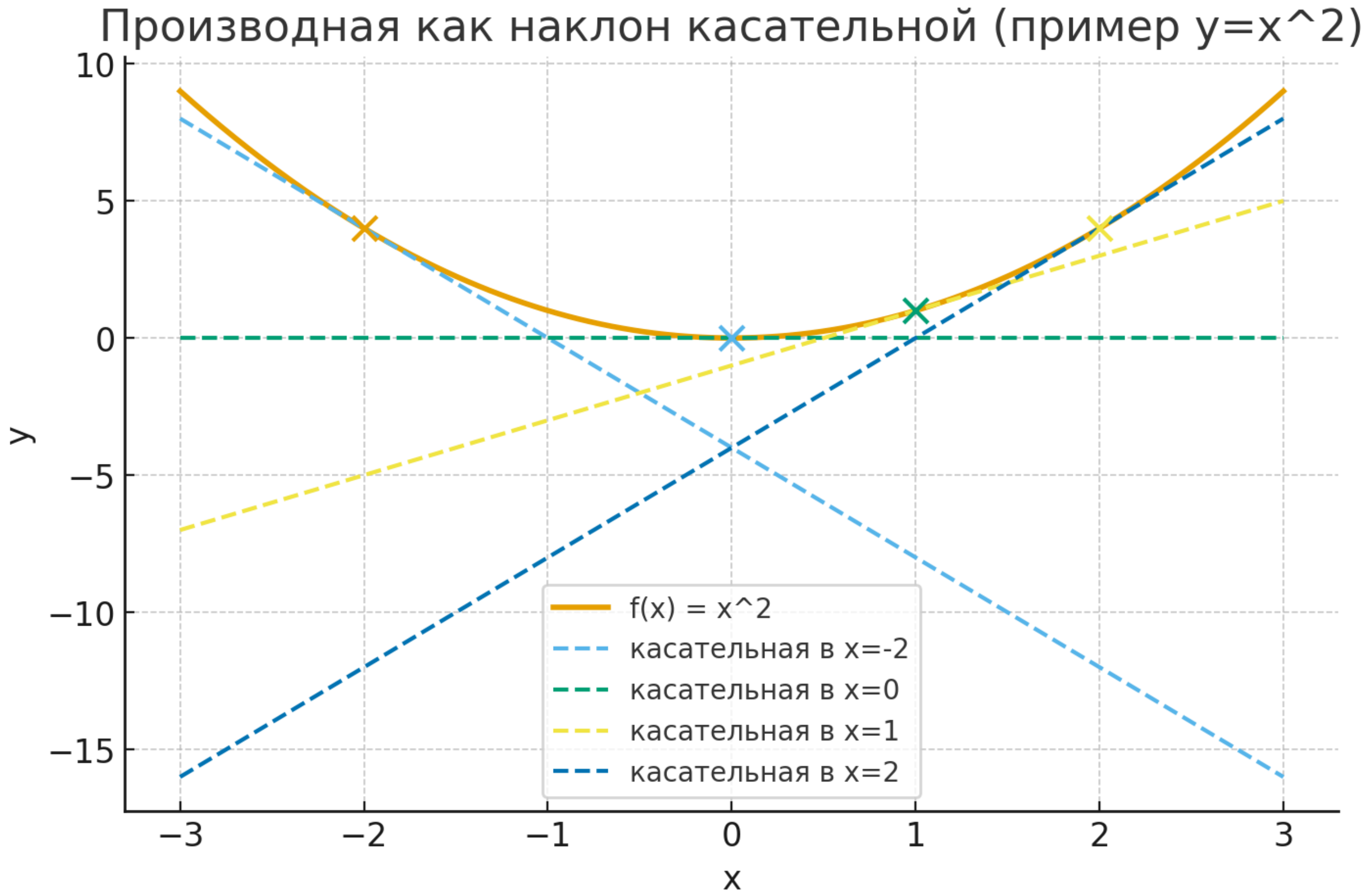
Производная — предел отношения приращения функции в данной точке к приращению аргумента, когда последнее стремится к нулю.

Иначе можно написать так: $y'(x_0) = \lim_{x \to 0} \frac{\Delta y}{\Delta x}$

# Производные



# Производные



# Производные

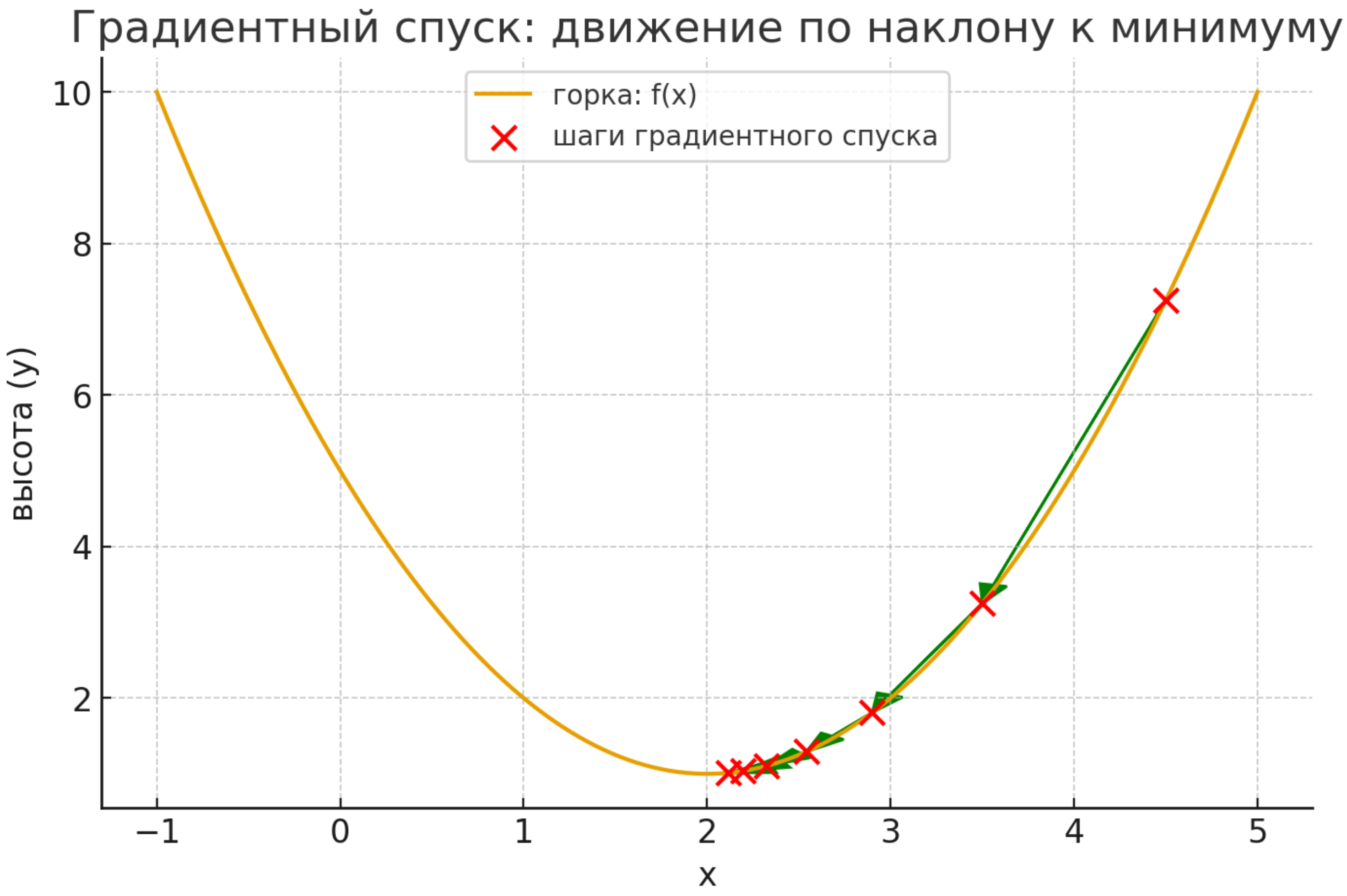
**Почему это важно?**

Снова градиентный спуск!

Алгоритмы (в т. ч. градиентный спуск) ищут минимум функции ошибки

Производная показывает **направление**, куда двигаться, чтобы эту ошибку уменьшить

# Производные



# Производные

Процесс нахождения производной называется дифференцированием. Функция, которая имеет производную в данной точке, называется дифференцируемой

Если функция дифференцируема, это значит, что:

- мы можем посчитать наклон (градиент) в любой точке
- можем корректно двигаться к минимуму
- алгоритм будет сходиться предсказуемо

# Производные

## Основные правила дифференцирования

1. **Константа**

$$(c)' = 0$$

2. **Степенная функция**

$$(x^n)' = n \cdot x^{n-1}, \quad (n \in \mathbb{R})$$

3. **Константа на функцию**

$$(c \cdot f(x))' = c \cdot f'(x)$$

4. **Сумма / разность**

$$(f(x) \pm g(x))' = f'(x) \pm g'(x)$$

5. **Произведение**

$$(f(x) \cdot g(x))' = f'(x)g(x) + f(x)g'(x)$$

6. **Частное**

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$$

7. **Сложная функция (правило цепочки)**

$$(f(g(x)))' = f'(g(x)) \cdot g'(x)$$

# Интеграл

Математически интегрирование — это операция, обратная дифференцированию

$$\int f(x)dx$$

$\int$ — знак интеграла

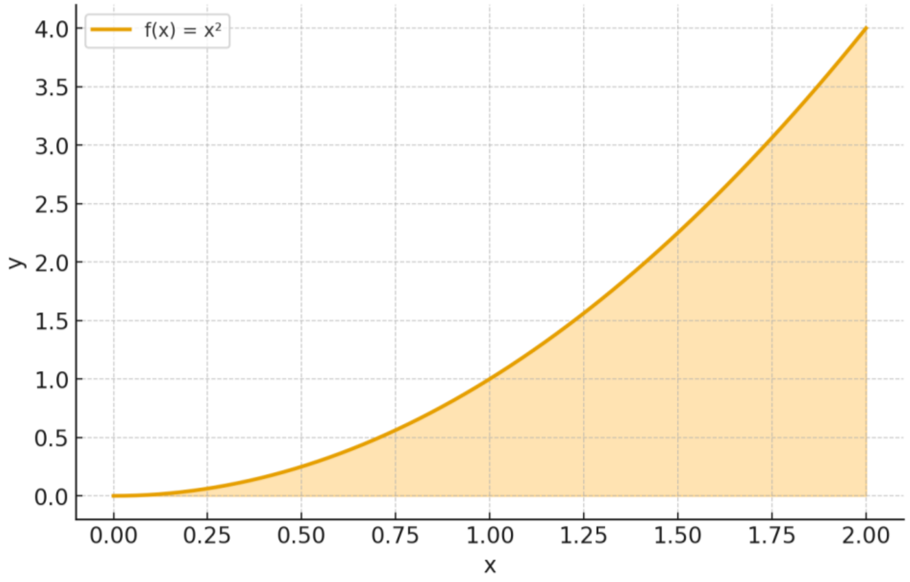
$f(x)$ — функция, которую интегрируем

$dx$ — указывает на то, что переменная интегрирования — $x$

# Интеграл

Определенный интеграл — это площадь под графиком между точками $a$ и $b$ на оси $x$

$$\int_a^b f(x)dx$$



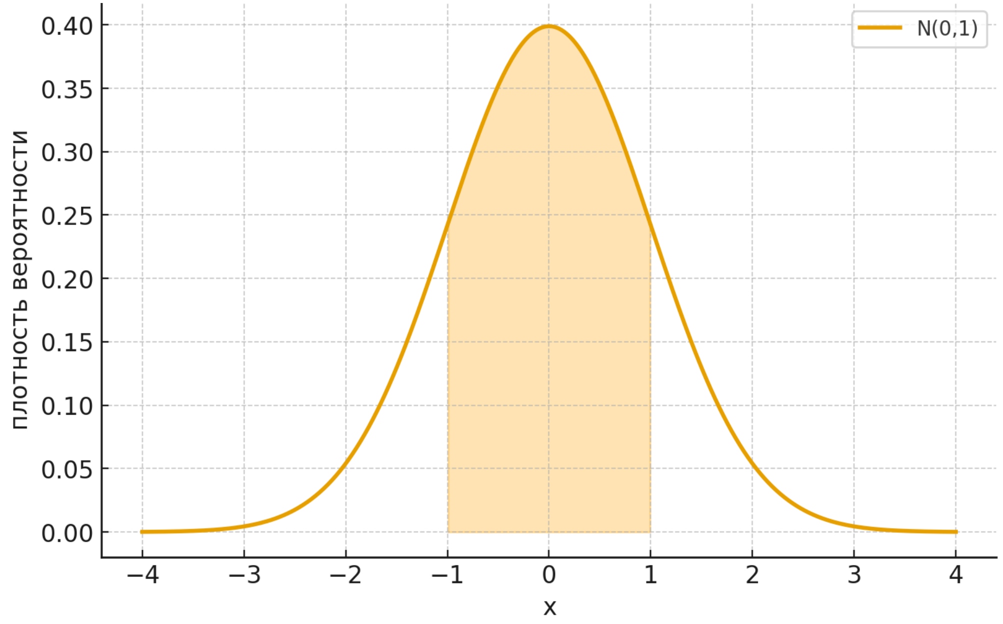

- Делим заштрихованную область на узкие прямоугольники
- Складываем площади прямоугольников
- Делаем ширину прямоугольников все меньше

В пределе (тут снова нужен предел!) мы получаем точную площадь. Это и есть интеграл.

# Интеграл

Интеграл в вероятностях: под «колоколом» нормального распределения

заштрихованная область показывает вероятность попасть в диапазон [-1, 1]



# Интеграл

**Почему это важно?**

- Интеграл позволяет найти среднее значение функции на отрезке
- В теории вероятностей интегралы описывают площадь под кривой распределения
- Интеграл может пригодиться для нормализации: при работе с непрерывными распределениями интеграл нужен, чтобы сумма вероятностей была равна 1:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

# Интеграл

Если **производная** отвечает на вопрос «как быстро меняется функция в точке», то **интеграл** отвечает на вопрос «какова накопленная сумма изменения/площадь под функцией»

# Векторы

Вектор — это упорядоченный набор чисел.

$$\vec{x} = (x_1, x_3, x_3)$$

Может быть вектор-строкой: $\vec{x} = (x_1, x_3, x_3)$ или вектор-столбцом: $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$

# Векторы

**Некоторые операции над векторами**

Сложение $(1, 2)\ +\ (3, 4)\ =\ (4, 6)$

Скалярное произведение $(1, 2)\ +\ (3, 4)\ =\ (3, 8)$

(мера «сходства» векторов)

Умножение на число (на скаляр) $2 \times (1, 2) = (2, 4)$

Длина (норма): $\|\vec{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$

# Векторы

**Почему это важно?**

В машинном обучении в виде векторов представлены данные (признаки), веса модели, градиенты. Мы все время имеем дело с векторными представлениями каких-либо данных или преобразуем данные в вектора для удобства или вообще обеспечения возможности работы с ними (например, модель не может обработать слова — для работы с естественным языком каждое слово нужно преобразовать в вектор с цифрами)

# Матрицы

Матрица — это по сути набор векторов:

$$\begin{matrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{matrix}$$

Единичная матрица:

$$\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}$$

# Матрицы

**Некоторые операции над матрицами**

- Транспонирование (меняем местами строки и столбцы)

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \longrightarrow \mathbf{A}^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

- Умножение матрицы на вектор (линейное преобразование: например, поворот, растяжение)

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 7 \\ 15 \end{bmatrix}$$

# Матрицы

**Некоторые операции над матрицами**

- Умножение матрицы на матрицу (последовательные преобразования)

$$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\times\begin{bmatrix}2 & 0\\1 & 2\end{bmatrix}=\begin{bmatrix}(1\cdot 2+2\cdot 1) & (1\cdot 0+2\cdot 2)\\(3\cdot 2+4\cdot 1) & (3\cdot 0+4\cdot 2)\end{bmatrix}=\begin{bmatrix}4 & 4\\10 & 8\end{bmatrix}$$

# Матрицы

**Обратная матрица**

Матрица $\mathbf{A}^{-1}$ называется обратной к матрице $\mathbf{A}$, если $\mathbf{A} \cdot \mathbf{A^{-1}} = \boldsymbol{I}$ (единичной матрице)

В линейной регрессии аналитическое решение считает веса через обратную матрицу

$w = (X^T X)^{-1} X^T \mathrm{y}$

**Почему это важно?** Если матрица признаков плохо обратима, веса становятся нестабильными и модель плохо обобщает —> понимание этого позволяет выявить мультиколлинеарность (когда признаки линейно связаны между собой) и применять регуляризацию (будем проходить на следующем занятии)

# Матрицы

Почему это важно?

- **В машинном обучении мы имеем дело с матрицами признаков, где объекты — это вектор-строки, а признаки — это вектор-столбцы**

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

# Матрицы

Почему это важно?

- **Линейные модели — по сути и есть матричные операции**

Линейная регрессия, логистическая регрессия сводятся к

$$y = X \cdot \beta$$

где

$X$ — матрица данных, $\beta$ — вектор весов модели, $y$ — предсказания модели

Это позволяет делать предсказания сразу для всех объектов одновременно с помощью одного умножения матриц

# Матрицы

Почему это важно?

- **Нейросети — матричные преобразования**

Каждый слой нейросети — это преобразование:

$$h = f(W \cdot x + b)$$

где

$W$ — матрица весов слоя, $x$ — входной вектор, $b$ — вектор смещений, $f$ — функция, которая добавляет нелинейности