Тверской
государственный
технический
университет

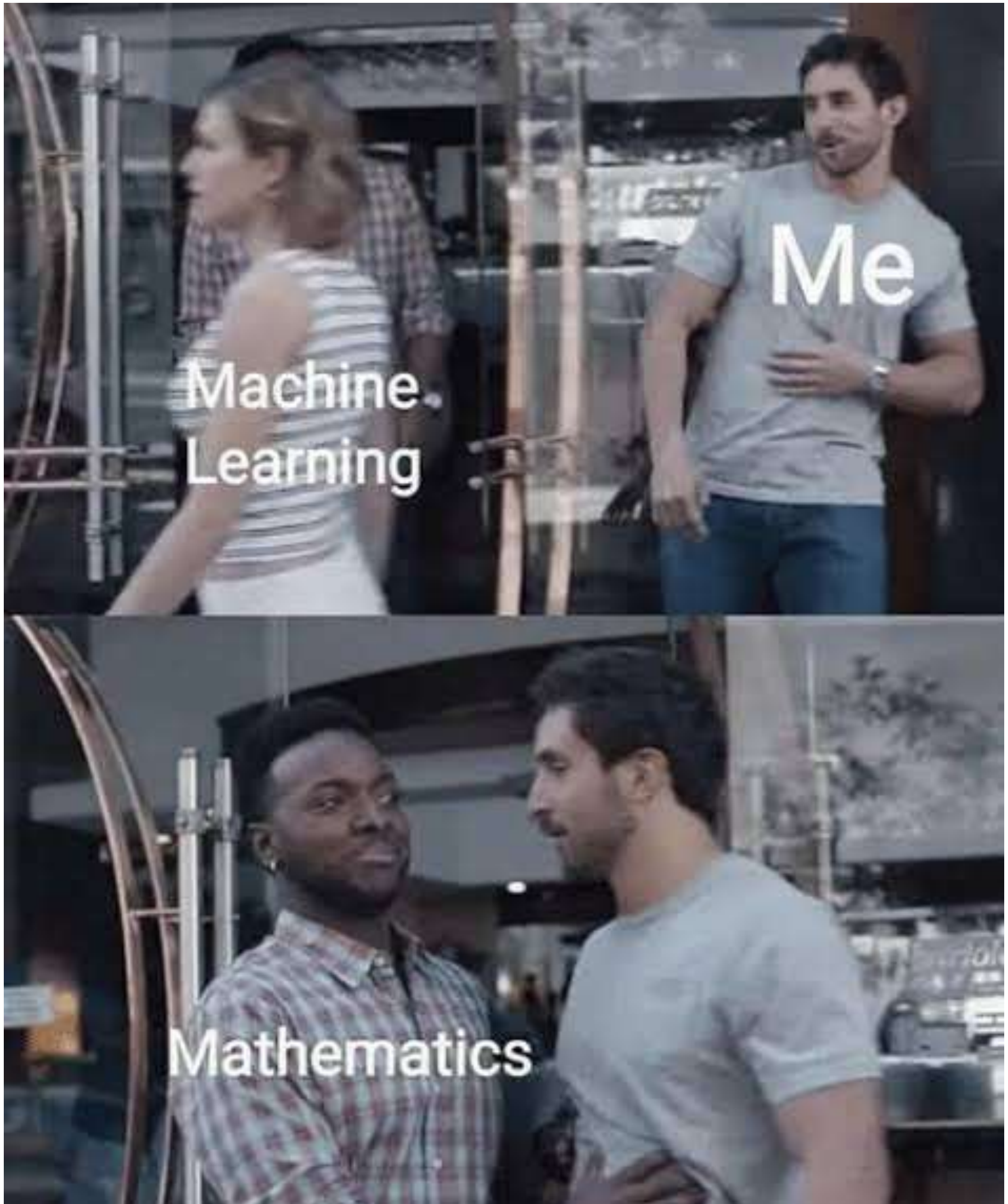
Интеллектуальные информационные системы
Math refresher

Шпигарь Андрей Николаевич

Материалы курса доступны по ссылке:
https://github.com/AndreyShpigar/ML-course

2024 г.

Актуализация знаний



- Скаляр – число, в общем случае элемент поля, над которым определено векторное пространство

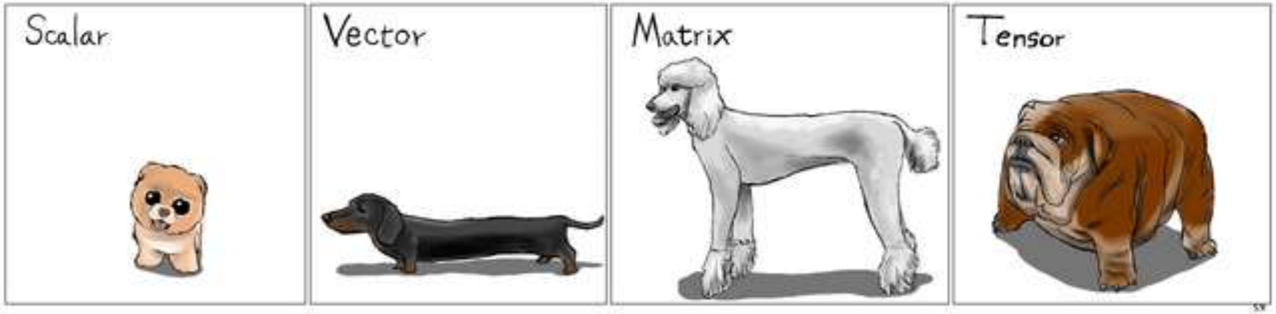
- Вектор – массив *n* чисел (*n* – элемент или скаляр вектора). Отдельный элемент вектора определяется своим индексом x_i. Количество элементов вектора определяют его порядок (длину). Может быть явно указан тип хранящихся в векторе чисел, например $\mathbb{R}^m$, означает, что каждый элемент вектора принадлежит множеству вещественных чисел, вектор порядка *m*.

$$\boldsymbol{x} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \ldots, \boldsymbol{x_n}]$$

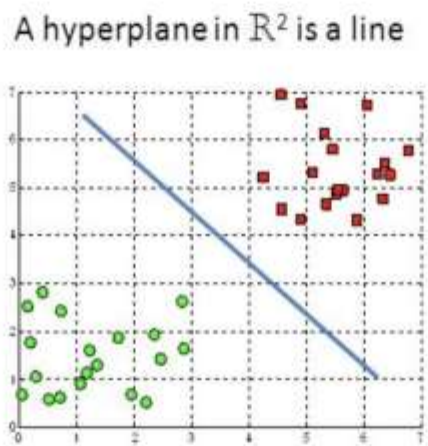
- Матрица $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{m \times n}}$ двумерный массив чисел с m строками и n столбцами. Каждый элемента матрицы индексируется двумя индексами $\boldsymbol{A_{mn}}$.

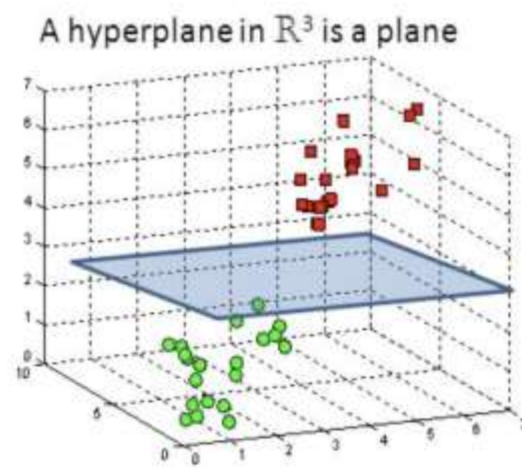
$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a_{11}} & \cdots & \boldsymbol{a_{1n}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{a_{m1}} & \cdots & \boldsymbol{a_{mn}} \end{bmatrix}$$

- <u>Тензор</u> – обобщение двумерного массива на большее число измерений. Характеризуется рангом. Скаляр имеет ранг 0, вектор ранг 1, матрица ранг 2. Тензоры имеют ранг 3 и выше.



- <u>Гиперплоскость</u> – подпространство, размерность которого на единицу меньше размерности объемлющего пространства. Разделяет *n*-мерное пространство на две части.

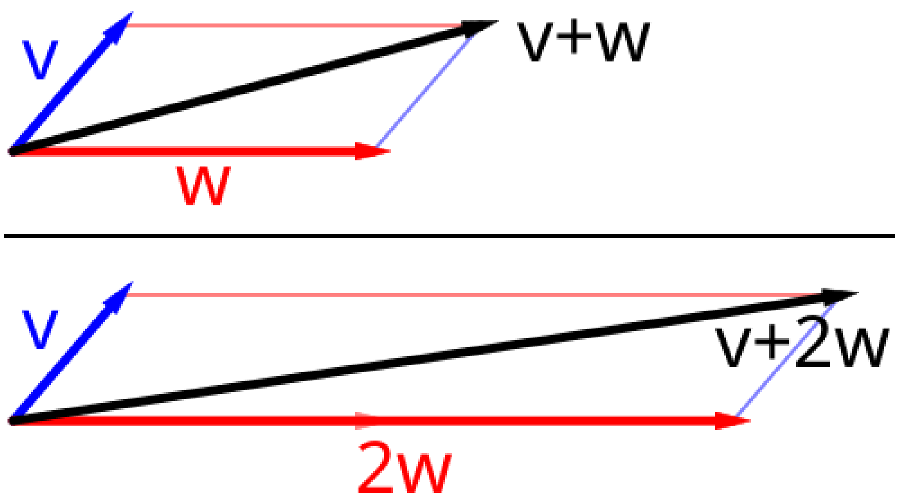




Векторные пространства

Векторное пространство – множество таких векторов, на котором определены операции сложения и умножения на скаляр.

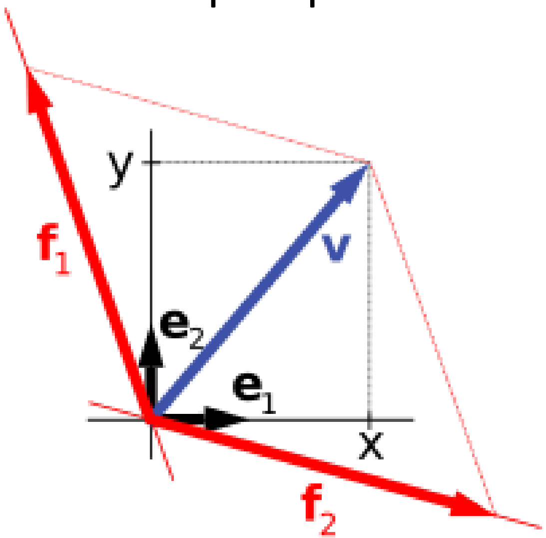
Множество векторов $\{x_1, x_2, \dots, x_n\}$ называется линейно независимым, если никакой вектор нельзя представить в виде линейной комбинации остальных.



Линейная оболочка множества векторов $\{x_1, x_2, \dots, x_n\}$ называется множество всех векторов, которые можно представить линейными комбинациями

$$\boldsymbol{span}(\{\boldsymbol{x_1}, \dots, \boldsymbol{x_n}\}) \triangleq \left\{\boldsymbol{v} : \boldsymbol{v} = \sum_{i=1}^{n} \boldsymbol{\alpha_i x_i},\ \boldsymbol{\alpha_i} \in \mathbb{R}\right\}$$

Базис $\mathcal{B}$ - множество линейно независимых векторов, линейной оболочкой которых является все пространство.



Линейное отображение – функция $f : \mathcal{V} \to \mathcal{W}$, такая, что

$$\boldsymbol{f}(\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{f}(\boldsymbol{v}) + \boldsymbol{f}(\boldsymbol{w}) \text{ и } \boldsymbol{f}(\boldsymbol{a}\boldsymbol{v}) = \boldsymbol{a}\boldsymbol{f}(\boldsymbol{v}) \text{для любых } \boldsymbol{v}, \boldsymbol{w} \in \boldsymbol{\mathcal{V}}$$

Нормы вектора

Норма вектора $\|\boldsymbol{x}\|$ - любая функция $\boldsymbol{f} : \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}$, обладающая свойствами:

- Неотрицательность: для любого $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$ $\boldsymbol{f}(\boldsymbol{x}) \geq \mathbf{0}$
- Определенность: $\boldsymbol{f}(\boldsymbol{x}) = \mathbf{0}$ только при $\boldsymbol{x} = \mathbf{0}$
- Однородности по абсолютной величине:

  $$\text{для любых } \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}, \mathbf{t} \in \mathbb{R}\ \boldsymbol{f}(\boldsymbol{t}\boldsymbol{x}) = |\boldsymbol{t}|\boldsymbol{f}(\boldsymbol{x})$$

- Неравенство треугольника: для любых $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{\boldsymbol{n}}$

  $$\boldsymbol{f}(\boldsymbol{x} + \boldsymbol{y}) \leq \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{f}(\boldsymbol{y})$$

Первая норма: $\|\boldsymbol{x}\|_{\mathbf{1}} = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} |\boldsymbol{x}_{\boldsymbol{i}}|$

Вторая норма: $\|\boldsymbol{x}\|_{\mathbf{2}} = \sqrt{\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{x}_{\boldsymbol{i}}^{\mathbf{2}}}$

Р-норма: $\|\boldsymbol{x}\|_{\boldsymbol{p}} = \left(\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} |\boldsymbol{x}_{\boldsymbol{i}}|^{\boldsymbol{p}}\right)^{\mathbf{1}/\boldsymbol{p}}$

Свойства матриц

<u>След матрицы</u> – сумма диагональных элементов

$$\boldsymbol{tr(A) \triangleq \sum_{i=1}^{n} A_{ij}}$$

<u>Определитель матрицы</u> *det*(A) или |A|, измеряет, как изменяется единичный объем под воздействием линейного преобразования, описываемого данной матрицей

<u>Ранг матрицы</u> – максимальное число линейно независимых строк или столбцов в этой матрице

Произведение двух матриц $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{m}\times\boldsymbol{n}}$ и $\boldsymbol{B} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{p}}$ - матрица $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} \in \mathbb{R}^{\boldsymbol{m}\times\boldsymbol{p}}$, где:

$$\mathrm{C}_{\boldsymbol{ij}} = \sum_{\boldsymbol{k}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{A}_{\boldsymbol{ik}}\boldsymbol{B}_{\boldsymbol{kj}}$$

Скалярное произведение векторов $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{\boldsymbol{n}}$:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \triangleq \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{x}_{\boldsymbol{i}}\boldsymbol{y}_{\boldsymbol{i}}$$

- Норма Фробениуса:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2} = \sqrt{tr(A^T A)} = \|vec(A)\|_2$$

- Число обусловленности матрицы: - мера *численной устойчивости* при вычислениях

$$k(A) = \|A\| \left\|A^{-1}\right\|$$

<u>Обратная матрица</u> для $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{n}}$**:**

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I} = \boldsymbol{A}\boldsymbol{A}^{-1}$$

существует только при det(A) ≠ 0

Если $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{n}}$*:*

- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^{-1})^{T} = (A^{T})^{-1}$

Собственные значения $\lambda \in \mathbb{R}$ матрицы $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{n}}$:

$$\boldsymbol{Au} = \boldsymbol{\lambda u}, \boldsymbol{u} \neq \boldsymbol{0}$$

где $\mathrm{u} \in \mathbb{R}^{\boldsymbol{n}}$ –собственный вектор

Спектральное разложение матрицы A:

$$\boldsymbol{A} = \boldsymbol{U diag}(\boldsymbol{\lambda})\boldsymbol{U}^{-\boldsymbol{1}}$$

$\boldsymbol{U} = \left[\boldsymbol{u}^{(\boldsymbol{1})}, \dots, \boldsymbol{u}^{(\boldsymbol{n})}\right]$ - матрица из $\boldsymbol{n}$ линейно-независимых собственных векторов

$\boldsymbol{\lambda} = [\boldsymbol{\lambda_1}, \dots, \boldsymbol{\lambda_n}]^{\boldsymbol{T}}$- вектор собственных значений

Сингулярное разложение (SVD) матрицы $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{m}\times\boldsymbol{n}}$:

$$\boldsymbol{A} = \boldsymbol{UDV}^{\boldsymbol{T}}$$

D – диагональная матрица сингулярных значений матрицы A

U,V – матрицы левых и правых сингулярных векторов

Производная функции $f : \mathbb{R} \to \mathbb{R}$ в точке a определяется как величина:

$$\boldsymbol{f'(x) \triangleq \lim_{n\to 0}\frac{f(x+h)-f(x)}{h}}$$

измеряет скорость изменения выхода при перемещении на малое расстояние от x в пространстве входов

Частная производная функции векторного аргумента $f : \mathbb{R}^n \to \mathbb{R}$ по x_i:

$$\boldsymbol{\frac{\partial f}{\partial x_i} = \lim_{h\to 0}\frac{f(x+he_i)-f(x)}{h}}$$

где $\boldsymbol{e_i}$ - i-й единичный вектор

Градиент функции в точке $\boldsymbol{x}$ – вектор, составленный из ее частных производных:

$$\boldsymbol{g = \frac{\partial f}{\partial x} = \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}}$$

Вероятность событий

- Вероятность <u>конъюнкции</u> двух событий A и B:

$$\boldsymbol{P(A \wedge B) = P(A, B)}$$

- События A и B <u>независимы</u>, если:

$$\boldsymbol{P(A, B) = P(A)P(B)}$$

- Вероятность <u>объединения</u> двух событий:

$$\boldsymbol{P(A \vee B) = P(A) + P(B) - P(A \wedge B)}$$

- Если события взаимно <u>исключающие</u>:

$$\boldsymbol{P(A \vee B) = P(A) + P(B)}$$

- <u>Условная вероятность</u>:

$$\boldsymbol{P(B|A) = \frac{P(A, B)}{P(A)}}$$

- <u>Условная независимость</u> событий $\boldsymbol{A \perp B|C}$:

$$\boldsymbol{P(A, B|C) = P(A|C)P(B|C)}$$

- Математическое ожидание (μ):

$$\mathbb{E}_{\boldsymbol{x}\sim\boldsymbol{P}}[\boldsymbol{f}(\boldsymbol{x})] = \sum_{\boldsymbol{x}} \boldsymbol{P}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})$$

- Дисперсия (σ^2):

$$\boldsymbol{V}\big(\boldsymbol{f}(\boldsymbol{x})\big) = \mathbb{E}[(\boldsymbol{f}(\boldsymbol{x}) - \mathbb{E}[\boldsymbol{f}(\boldsymbol{x})])^{\boldsymbol{2}}]$$

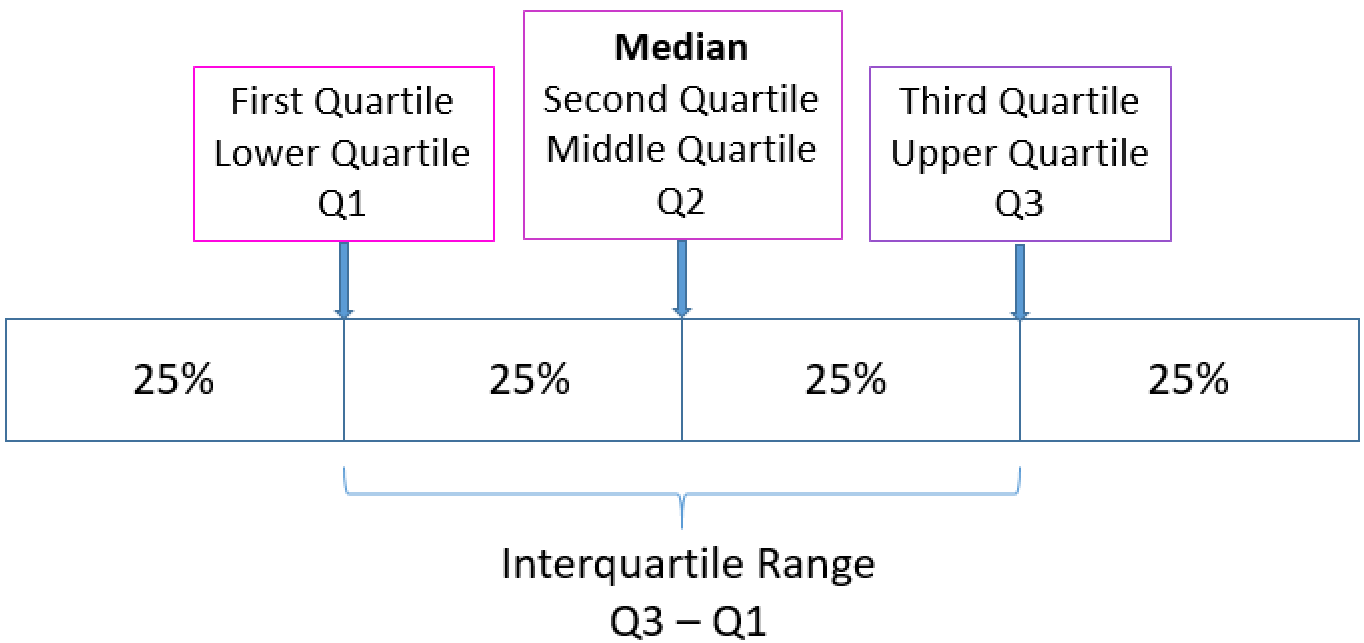
- Ковариация:

$$\boldsymbol{Cov}[\boldsymbol{X},\boldsymbol{Y}] = \mathbb{E}[\boldsymbol{XY}] - \mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{Y}]$$

- Коэффициент корреляции Пирсона:

$$\boldsymbol{\rho} = \frac{\boldsymbol{Cov}[\boldsymbol{X},\boldsymbol{Y}]}{\sqrt{\boldsymbol{V}[\boldsymbol{X}]\boldsymbol{V}[\boldsymbol{Y}]}}$$

- Мода – значение во множестве наблюдений, которое встречается наиболее часто
- Медиана – значение, которое в упорядоченной выборке находится ровно посередине
- Квантиль – значение, которое заданная случайная величина не превышает с фиксированной вероятностью. Если вероятность задана в процентах, то квантиль называется перцентилем.



<u>Нормальное распределение</u> – распределение вероятностей, которое в одномерном случае задается функцией плотности вероятности, совпадающей с функцией Гаусса

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$