Введение. Функции нескольких переменных

> Существует функция n переменных $x,\ y,\ z...$, если по некоторому закону каждой системе n чисел $(x,\ y,\ z...)$ из некоторого множества ставится в соответствие число u.

То есть, по сути, относительно функции одной переменной меняется только то, что значение функции зависит от нескольких аргументов, а не от одного:

$$f\underbrace{(x, y)}_{\text{Принимает на вход несколько аргументов}} = x^2 y$$

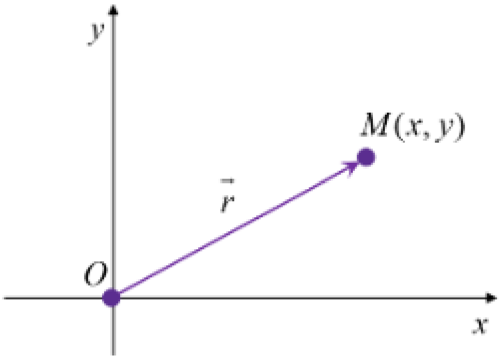
Примечание. Исключением из этого правила являются так называемые **векторнозначные функции**, для которых значением может быть не одно число, а несколько:

$$f(x) = \begin{bmatrix} cos(x) \\ sin(x) \end{bmatrix}$$

Если говорить про евклидово расстояние для двух переменных, то его функция записывается как $f(x)\ =\ \sqrt{x^2\ +\ y^2}$ и вычисляет корень из суммы квадратов x и y. Это не что иное, как расстояние от точки (x,y) до начала координат или длина вектора с координатами (x,y). В качестве области определения для такой функции могут выступать любые вещественные числа, а её областью значений являются все неотрицательные числа.

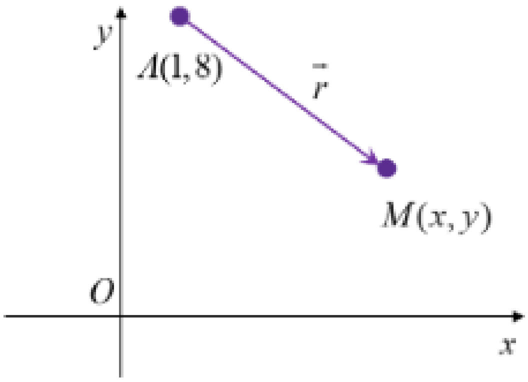
Евклидово расстояние до начала координат:

$$f(x, y) = \sqrt{x^2 + y^2} = \rho(M(x, y), O(0, 0)) = \left\| \overrightarrow{r} \right\|$$



Также евклидово расстояние можно использовать для того, чтобы найти **расстояние между двумя точками:**

$$f(x, y) = \sqrt{(x-1)^2 + (y-8)^2} = \rho(M(x, y), A(1, 8))$$



Также функцией нескольких переменных является квадратичная функция потерь или, как её ещё называют, **L2-loss-функция**:

$$L^2 LossFunction = \sum_{i=1}^{n} (y_{true} - y_{predicted})^2$$

Эта функция находит сумму квадратов всех ошибок, то есть, сумму квадратов разностей между реальным значением и значением, предсказанным моделью. Вы уже сталкивались с ней, так как она используется для построения и оценки качества в линейных регрессиях (в методе наименьших квадратов), а также для метода обратного распространения ошибки, применяемого в нейронных сетях.

→ Так как эта функция выражает величину отклонения от истины, нам очень важно уметь её исследовать и находить её минимальное значение, чтобы достигать максимально точных предсказаний.

В качестве аргументов функция L2-loss принимает любые вещественные числа, а её значения могут быть любыми неотрицательными числами.

Сигмоида является функцией активации нейрона, которая используется при обучении нейронных сетей. Её вариацией является логистическая функция, используемая для решения задачи классификации. Если параметров много, то, разумеется, мы получаем её как функцию многих переменных:

$$E(y) = \frac{\alpha}{\left(1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}\right)}$$

Ровно так же, как и одномерная, она может принимать любые значения от 0 до 1 не включительно, и в качестве её аргументов могут выступать любые вещественные числа.

Частные производные

Необязательно искать производные в определённых точках — мы можем, как и в случае с одномерными функциями, находить значения производных в общем виде.

Можно сказать, что мы ищем производную для функции $f(x, y) = x^2 y^3$ именно по x. Таким образом, значение функции будет меняться вследствие изменения значения x, а значение переменной y будет оставаться прежним. Раз значение y не будет изменяться, мы можем воспринимать его как константу.

$$\frac{d}{dx} f(x, y) = \underbrace{\frac{d}{dx}(x^2 y^3)}_{y \text{ воспринимаем как константу}} = 2xy^3$$

Иногда для того, чтобы подчеркнуть, что это именно многомерная функция, используют следующую запись:

$$\frac{\partial}{\partial x} f(x, y) = \frac{\partial}{\partial x}(x^2 y^3) = 2xy^3$$

Есть ли разница в записи между $\frac{d}{dx}$ и $\frac{\partial}{\partial x}$? На самом деле нет, хотя если быть очень педантичными, то можно сказать, что первый вариант используется скорее для функции одной переменной, а второй — для функций нескольких переменных. Но в литературе вы можете встретить оба варианта, и их стоит воспринимать одинаково.

Также вы можете встретить (в том числе и в наших юнитах) следующие общепринятые обозначения:

$$f_x' \leftrightarrow \frac{\partial f}{\partial x} \qquad f_y' \leftrightarrow \frac{\partial f}{\partial y}$$

Это можно прочитать как «производная от функции f по x» или «производная от функции f по y» (в зависимости от переменной). Вне зависимости от формы записи, читаются они одинаково, и обе являются верными.

Алгоритм поиска частной производной аналогичен алгоритму поиска обычной производной для одномерного случая.

Пусть дана функция. Чтобы найти частную производную по переменной x_i:

1. Фиксируем все переменные, кроме x_i.
2. Считаем приращение функции при изменении только этой переменной.
3. Делим приращение из пункта 2 на приращение нашей переменной.
4. Уменьшаем Δx_i и получаем $\frac{\partial f}{\partial x_i}$.

Пример № 1

Найти обе частные производные для функции $f(x,y) = cos(x^2y) + y^3$.

Для начала найдём производную по x — f'_x.
Фиксируем y как константу и дифференцируем:

$$f'(x) = (cos(x^2y))'_x + (y^3)'_x = -sin(x^2y) \cdot (x^2y)'_x + 0 =$$

$$= -sin(x^2y) \cdot 2xy = -2xy \cdot sin(x^2y)$$

Повторим то же самое для y:

$$f'(y) = (cos(x^2y))'_y + (y^3)'_y = -sin(x^2y) \cdot (x^2y)'_y + 3y^2 =$$

$$= -sin(x^2y) \cdot x^2 + 3y^2 = -x^2 \cdot sin(x^2y) + 3y^2$$

Так же, по аналогии с одномерными функциями, мы можем найти и вторые частные производные.

Для вторых производных используются следующие обозначения:

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)=\frac{\partial^2 f}{\partial x^2}$$ → Первая производная взята по **x**, и вторая тоже по **x**

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)=\frac{\partial^2 f}{\partial x \partial y}$$ → Первая производная взята по **y**, а вторая по **x**

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)=\frac{\partial^2 f}{\partial y \partial x}$$ → Первая производная взята по **x**, а вторая по **y**

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)=\frac{\partial^2 f}{\partial y^2}$$ → Первая производная взята по **y**, и вторая тоже по **y**

Также могут использоваться такие вариации:

$$(f_x)_x = f_{xx}$$ → Первая производная взята по **x**, и вторая тоже по **x**

$$(f_y)_x = f_{yx}$$ → Первая производная взята по **y**, а вторая по **x**

$$(f_x)_y = f_{xy}$$ → Первая производная взята по **x**, а вторая по **y**

$$(f_y)_y = f_{yy}$$ → Первая производная взята по **y**, и вторая тоже по **y**

Рассмотрим пример вычисления всех вторых частных производных для функции $f(x, y) = sin(x)y^2$.

Для начала вычислим первые частные производные по x и по y:

$$\frac{\partial}{\partial x}(\sin(x)y^2) = \cos(x)y^2$$

$$\frac{\partial}{\partial y}(\sin(x)y^2) = 2\sin(x)y$$

Теперь для каждой из двух частных производных возьмём ещё по две частных производных, чтобы получить все четыре возможных частных производных:

$$\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}(\sin(x)y^2)\right) = \frac{\partial}{\partial x}(\cos(x)y^2) = -\sin(x)y^2$$

$$\frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}(\sin(x)y^2)\right) = \frac{\partial}{\partial x}(2\sin(x)y) = 2\cos(x)y$$

$$\frac{\partial}{\partial y}\left(\frac{\partial}{\partial x}(\sin(x)y^2)\right) = \frac{\partial}{\partial y}(\cos(x)y^2) = 2\cos(x)y$$

$$\frac{\partial}{\partial y}\left(\frac{\partial}{\partial y}(\sin(x)y^2)\right) = \frac{\partial}{\partial y}(2\sin(x)y) = 2\sin(x)$$

Можно заметить, что **значения смешанных производных** (когда ищем производную сначала по одной переменной, а затем — по другой) **совпадают**. Это свойство очень поможет нам при поиске минимумов и максимумов функции, так как сильно упростит расчёты: получается, что для функции с двумя переменными нужно вычислить не четыре производных, а только три. Уже в следующем юните вы сможете по достоинству оценить, как это упростит вам жизнь и сократит длительность вычислений.

$$\sin(x)y^2$$

$$\frac{\partial}{\partial x}\swarrow \qquad \searrow\frac{\partial}{\partial y}$$

$$\cos(x)y^2 \qquad\qquad 2\sin(x)y$$

$$\frac{\partial}{\partial x}\swarrow \qquad \searrow\frac{\partial}{\partial y} \qquad \frac{\partial}{\partial x}\swarrow \qquad \searrow\frac{\partial}{\partial y}$$

$$\sin(x)y^2 \qquad \underbrace{2\cos(x)y \qquad 2\cos(x)y}_{\text{смешанные частные производные совпадают!}} \qquad 2\sin(x)$$

Безусловные экстремумы. Функции нескольких переменных

Рассмотрим алгоритм поиска минимумов и максимумов для функции нескольких переменных $f(x, y) = x^2 - 10x + y^2 - 12y + 71$.

На первом этапе необходимо найти частные производные этой функции:

$$f_x(x, y) = 2x - 10$$

$$f_y(x, y) = 2y - 12$$

На втором этапе мы находим точки, в которых эти производные принимают нулевые значения:

$$2x - 10 = 0 \qquad 2y - 12 = 0$$

$$2x = 10 \qquad 2y = 12$$

$$x = 5 \qquad y = 6$$

Итак, мы получили точку (5;6) — это стационарная точка.

Рассмотрим пример составления матрицы Гессе для нашей функции:

$$f'x = 2x - 10 \; f'y = 2y - 12$$

$$f''_{xx} = (2x - 10)'_x = 2$$

$$f''_{xy} = (2x - 10)'_y = 0$$

$$f''_{yx} = (2y - 12)'_x = 0$$

$$f''_{yy} = (2y - 12)'_y = 2$$

$$Hesse = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Теперь нам необходимо вспомнить линейную алгебру, так как критерии для максимума и минимума следующие:

- ➔ Если матрица Гессе положительно определена в этой точке, то в ней находится минимум.
- ➔ Если матрица Гессе отрицательно определена в этой точке, то в ней находится строгий максимум.
- ➔ В иных случаях точка является седловой.

В нашем случае знак первого углового минора положителен (его значение равно 2), как и знак второго углового минора (его значение равно 4). Оба собственных значения также положительны (оба равны 2). Это значит, что в данной точке находится минимум.

Условные экстремумы. Метод Лагранжа

Рассмотрим, как формулируется и решается **задача оптимизации, в которой есть ограничения**.

Пусть у нас есть функция $f(x)$, для которой мы хотим найти минимум при ограничении $\varphi_i(x) = 0$. Можно записать это следующим образом:

$$min_{\phi_i(x)=0} \; f(x)$$

Чтобы решить такую задачу, в первую очередь записываем функцию Лагранжа, т. е. объединяем с её помощью саму целевую функцию и функцию ограничений внутри одного выражения:

$$L(x, \lambda) = f(x) + \sum \lambda_i \phi_i(x)$$

Дополнительные переменные λ_i, которые появляются при формулировании функции Лагранжа, называются **множителями Лагранжа**.

Разберём алгоритм применения метода множителей Лагранжа на конкретном примере. Представим, что мы хотим максимизировать следующую функцию:

$$f(x, y) = 2x + y$$

Однако мы хотим не просто найти её максимально возможное значение (на самом деле оно равно бесконечности, если на и не стоит никаких ограничений), а сделать это при условии выполнения следующего равенства:

$$x^2 + y^2 = 1$$

Мы хотим найти такую точку этой окружности, для которой значение функции $f(x, y) = 2x + y$ будет максимально возможным.

Для того чтобы решить данную задачу, нам необходимо написать функцию Лагранжа. В общем виде она выглядит следующим образом:

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda(g(x, y) - c)$$

Здесь $f(x, y)$ — это сама функция, $g(x, y) - c$ — функция ограничений, λ — множитель Лагранжа.

В данном случае все эти элементы функции Лагранжа принимают следующий вид:

$$f(x,y) = 2x + y$$
$$g(x,y) = x^2 + y^2$$
$$c = 1$$

Сама функция записывается следующим образом:

$$\mathcal{L}(x,y,\lambda) = 2x + y - \lambda(x^2 + y^2 - 1)$$

Первым делом мы должны найти частные производные по каждой переменной:

$$L'_x = (2x + y - \lambda x^2 - \lambda y^2 + \lambda)'_x = 2 - 2\lambda x$$

$$L'_y = (2x + y - \lambda x^2 - \lambda y^2 + \lambda)'_y = 1 - 2\lambda y$$

$$L'_\lambda = (2x + y - \lambda x^2 - \lambda y^2 + \lambda)'_\lambda = -x^2 - y^2 + 1$$

Теперь приравняем их к нулю и решим систему уравнений:

$$\begin{cases} 2 - 2\lambda x = 0 \\ 1 - 2\lambda y = 0 \\ -x^2 - y^2 + 1 = 0 \end{cases}$$

$$x = \frac{2}{2\lambda} = \frac{1}{\lambda},\ y = \frac{1}{2\lambda}$$

$$-(\frac{1}{\lambda})^2 - (\frac{1}{2\lambda})^2 + 1 = 0$$

$$-\frac{1}{\lambda^2} - \frac{1}{4\lambda^2} + 1 = 0$$

$$\lambda_1 = \frac{\sqrt{5}}{2},\ \lambda_2 = -\frac{\sqrt{5}}{2}$$

$$x_1 = \frac{2}{\sqrt{5}},\ x_2 = -\frac{2}{\sqrt{5}}$$

$$y_1 = \frac{1}{\sqrt{5}},\ y_2 = -\frac{1}{\sqrt{5}}$$

Для определения точек максимума и минимума найдём значения целевой функции в найденных точках:

$$(\frac{2}{\sqrt{5}}; \frac{1}{\sqrt{5}}) : 2x + y = \frac{4}{\sqrt{5}} + \frac{1}{\sqrt{5}} = \frac{5}{\sqrt{5}} = \sqrt{5} \approx 2.24$$

$$(-\frac{2}{\sqrt{5}}; -\frac{1}{\sqrt{5}}) : 2x + y = -\frac{4}{\sqrt{5}} - \frac{1}{\sqrt{5}} = -\frac{5}{\sqrt{5}} = -\sqrt{5} \approx -2.24$$

Получаем, что в первой точке — максимум, а во второй — минимум.

Градиент и антиградиент

Градиент функции f, обозначаемый как ∇f, представляет собой совокупность всех её частных производных, представленных в виде вектора:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \vdots \end{bmatrix}$$

Пусть у нас есть функция $f(x, y) = x^2 + y^2$, и мы хотим найти её градиент в точках M (0;0), N (1; -1), P (1;1).

Вычислим частные производные:

$$\frac{\partial f}{\partial x} = 2x \quad \frac{\partial f}{\partial y} = 2y$$

Составим вектор градиента:

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (2x, 2y)$$

Либо, кроме уже известного обозначения, можно использовать такое:

$$\overrightarrow{grad} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (2x, 2y)$$

Теперь подставим интересующие нас точки:

$$\overrightarrow{grad} f|_M = (2x, 2y)|_{x=0, y=0} = (0, 0)$$

$$\overrightarrow{grad} f|_N = (2x, 2y)|_{x=1, y=-1} = (2, -2)$$

$$\overrightarrow{grad} f|_P = (2x, 2y)|_{x=1, y=1} = (2, 2)$$

Получаем, что в точке N градиент равен (2, -2), то есть при увеличении x функция будет возрастать, а при увеличении y — убывать. В точке P значение градиента равно (2, 2) — это значит, что в окрестности P сумма квадратов возрастает по обеим переменным. В точке M градиент нулевой, то есть все частные производные равны нулю, а значит возможно, что в этой точке находится экстремум.

Градиент и градиентный спуск

- ➔ Градиентный спуск минимизирует дифференцируемые функции с любым количество переменных. Он делает это, стартуя из случайной точки и далее двигаясь с заданным шагом в направлении антиградиента.
- ➔ Градиентный спуск не может определить, является найденный минимум локальным или глобальным.
- ➔ Размер шага определяет, быстро или медленно алгоритм сходится к минимуму или расходится.