Производная функции нескольких аргументов

Функция нескольких аргументов. Производная функции нескольких аргументов. Градиент в задачах оптимизации. Производная по направлению. Касательная плоскость и линейное приближение.

Даниил Корбут
Специалист по Анализу Данных

НЕТОЛОГИЯ

Даниил Корбут
DL Researcher
Insilico Medicine, Inc

Окончил бакалавриат ФИВТ МФТИ (Анализ данных) в 2018г
Учусь на 2-м курсе магистратуры ФИВТ МФТИ
Работал в Statsbot и Яндекс. Алиса.
Сейчас в Insilico Medicine, Inc, занимаюсь генерацией активных молекул и исследованиями старения с помощью DL.

НЕТОЛОГИЯ

Функция нескольких переменных

Из прошлой лекции:
Функция - это некоторое соответствие x -> f(x), причём для каждого x определено единственное значение f(x).

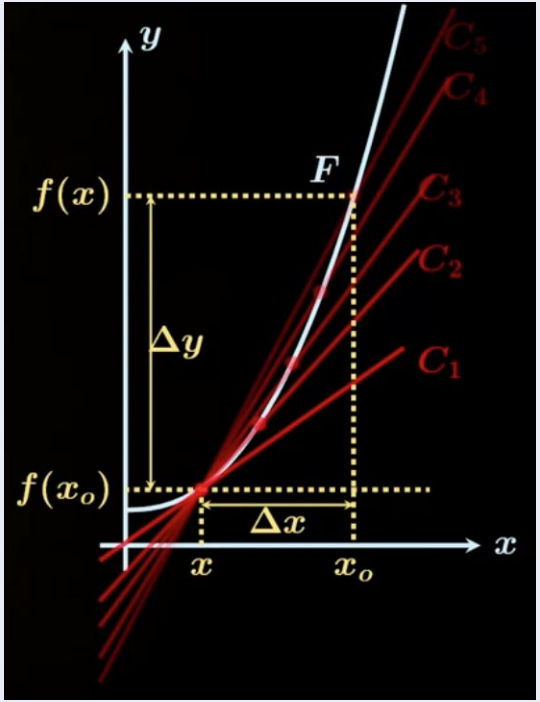
Теперь x - не число из R, а вектор (x1, ..., xn), где каждое xi из R, а весь x из R^n.

D(f) - область определения функции (ничего не изменилось!)
E(f) - область значений функции (ничего не изменилось!)

Снова будем работать только с функциями, у которых D(f) и E(f) - подмножество R^n.

Функция нескольких переменных

Из прошлой лекции мы знаем и геометрический смысл производной - *угловой коэффициент касательной*.



Как посчитать производную функции нескольких переменных?

- $\dfrac{\Delta f}{\Delta x} \xrightarrow{\Delta x \to 0} f'_x$, y — фиксирован
- $\dfrac{\Delta f}{\Delta y} \xrightarrow{\Delta y \to 0} f'_y$, x — фиксирован

https://www.coursera.org/learn/mathematics-and-python

Частная производная

Частная производная — это одно из обобщений понятия производной на случай функции нескольких переменных.

Частная производная функции f(x, y) по x определяется как производная по x, взятая в смысле функции одной переменной, при условии постоянства оставшейся переменной y.

$$f'_x(x,y) = \lim_{h\to 0} \frac{f(x+h,y)-f(x,y)}{h}, \qquad f'_y(x,y) = \lim_{h\to 0} \frac{f(x,y+h)-f(x,y)}{h}.$$

https://www.coursera.org/learn/mathematics-and-python

Касательная плоскость

Пусть дана некоторая функция двух переменных f : R^2 → R. Она, вообще говоря, определяет некоторую поверхность z = f(x, y) в трехмерном пространстве.

Если в некоторой точке (x0, y0) функция дифференцируема как функция многих переменных, то в этой точке можно рассмотреть **касательную плоскость** к данной поверхности.

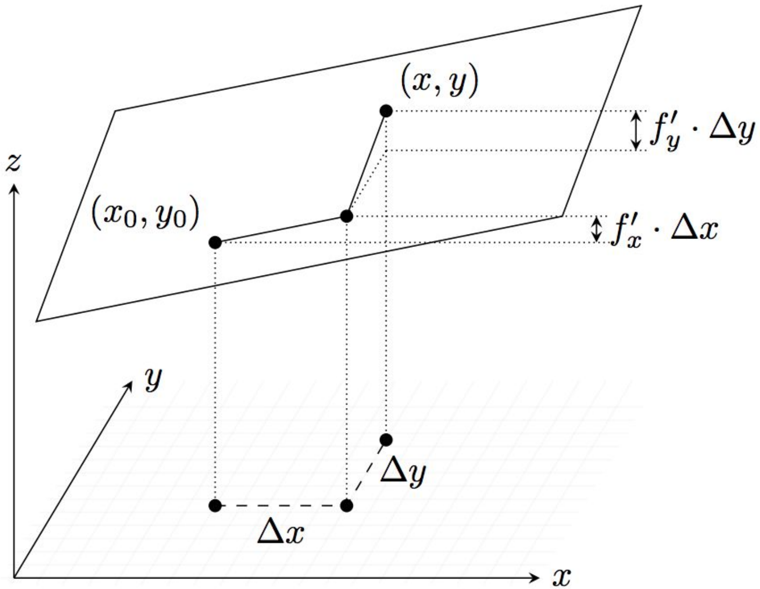


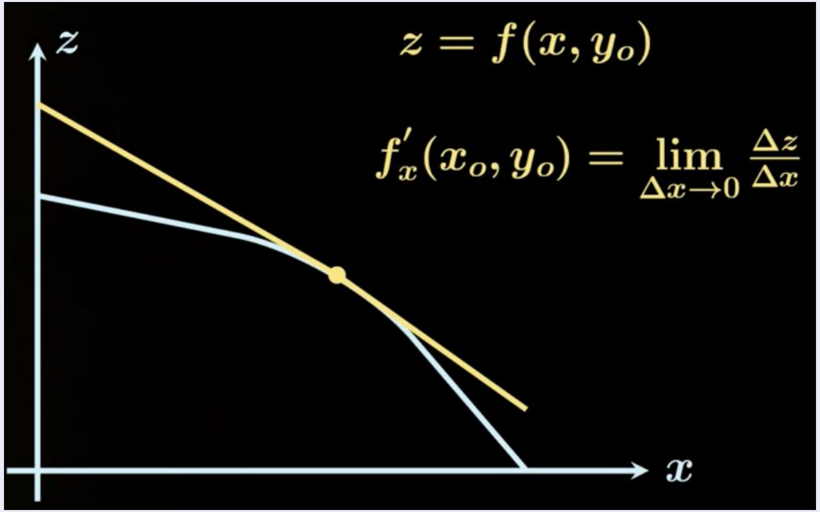
Рис. 1: Геометрический смысл частных производных.

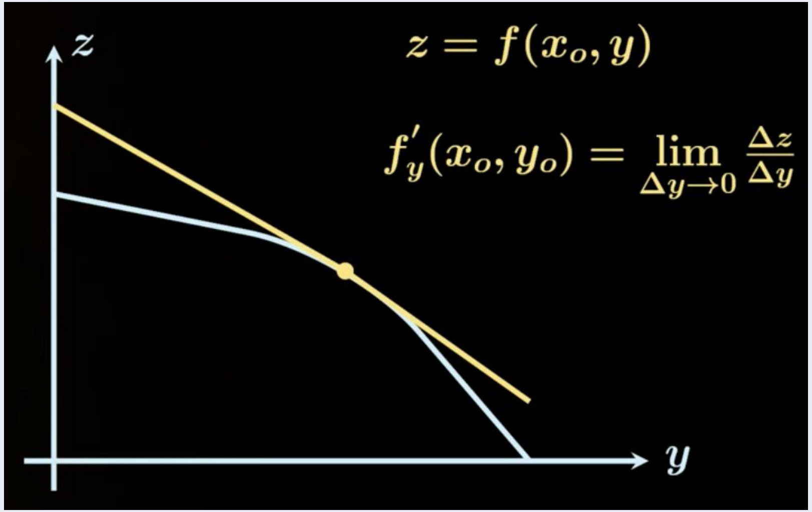
https://www.coursera.org/learn/mathematics-and-python

Касательная плоскость

Таким образом, график функции $f(x, y)$ в окрестности точки можно приблизить касательной плоскостью:

$$f(x_0 + \Delta x, y_0 + \Delta y) \approx f(x_0, y_0) + f'_x(x_0, y_0)\Delta x + f'_y(x_0, y_0)\Delta y.$$





https://www.coursera.org/learn/mathematics-and-python

Градиент и линии уровня функции

Если f(x1, . . . , xn) — функция n переменных x1, ..., xn, то n-мерный вектор из частных производных:

$$\operatorname{grad} f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)$$

!

называется **градиентом функции**.

!

Линией уровня функции называется множество точек, в которых функция принимает одно и то же фиксированное значение. Оказывается, что **градиент перпендикулярен линии уровня**.

Градиент в задачах оптимизации

Задачей оптимизации называется задача по нахождению экстремума функции, например минимума:

$$f(x_1, ..., x_n) \to \min$$

Такая задача часто встречается в приложениях, например при выборе оптимальных параметров рекламной компании, а также в задачах классификации.

Вспомним необходимые условия экстремума из прошлой лекции!

https://www.coursera.org/learn/mathematics-and-python

Градиент в задачах оптимизации

Но не всегда задачу можно решать аналитически. В таком случае используется численная оптимизация. Наиболее простым в реализации из всех методов численной оптимизации является метод градиентного спуска.

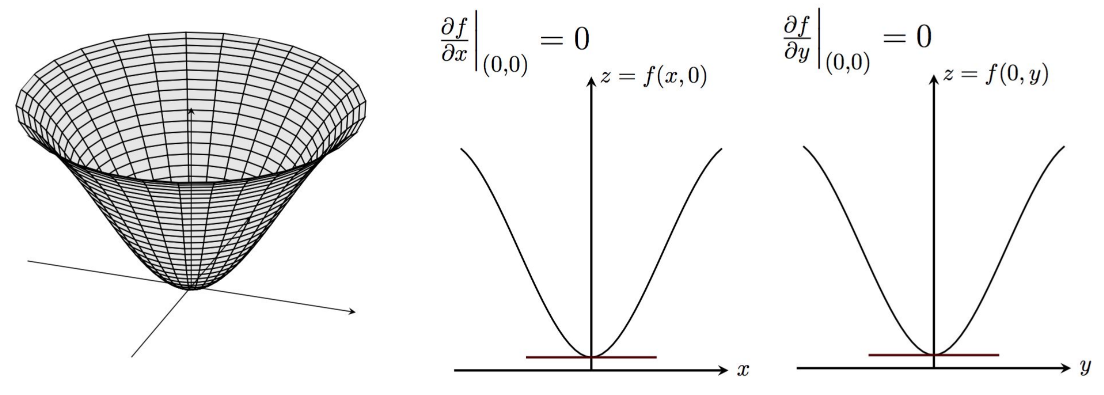


Рис. 2: Функция двух переменных достигает минимума в начале координат.

https://www.coursera.org/learn/mathematics-and-python

Градиентный спуск

Это итерационный метод. Решение задачи начинается с выбора начального приближения $\vec{x}^{[0]}$

После вычисляется приблизительное значение $\vec{x}^1$

Затем $\vec{x}^2$

и так далее…

!

$$\vec{x}^{[j+1]} = \vec{x}^{[j]} - \gamma^{[j]} \nabla F(\vec{x}^{[j]}), \qquad \text{где } \gamma^{[j]} \text{ — шаг градиентного спуска.}$$

Идея: идти в направлении наискорейшего спуска, а это направление задаётся антиградиентом $-\nabla F$.

НЕТОЛОГИЯ

https://www.coursera.org/learn/mathematics-and-python

Градиентный спуск

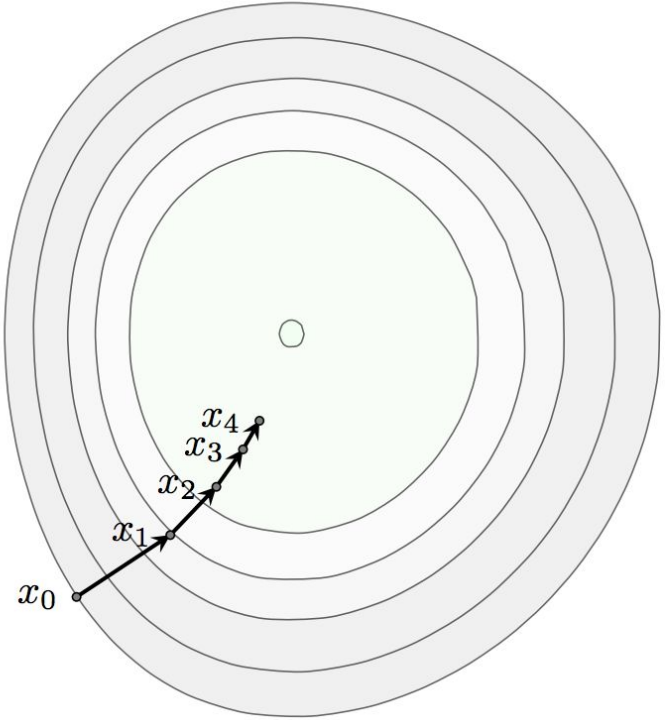


Рис. 3: Градиентный спуск

Аналогия: домик в низине

Если заблудились, то верным решением будет двигаться в направлении наискорейшего спуска

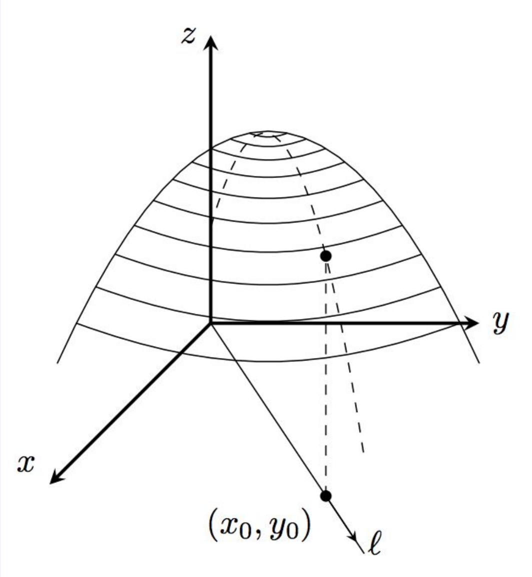
https://www.coursera.org/learn/mathematics-and-python

Производная по направлению

Пусть $f(\vec{x}) = f(x_1, x_2, \ldots, x_n)$ — функция n переменных, $\vec{\ell} \in \mathbb{R}^n$, $|\vec{\ell}| = 1$, тогда частной производной в точке x_0 по направлению $\vec{\ell}$ называется

$$\frac{\partial f}{\partial \vec{\ell}}(\vec{x}_0) = \lim_{t \to 0} \frac{f(\vec{x}_0 + t \cdot \vec{\ell}) - f(\vec{x}_0)}{t}.$$

!



Производная по направлению показывает, насколько быстро функция изменяется при движении вдоль заданного направления.

https://www.coursera.org/learn/mathematics-and-python

Связь градиента и производной по направлению

Производную по направлению дифференцируемой по совокупности переменных функции можно рассматривать как проекцию градиента функции на это направление, или иначе, как скалярное произведение градиента на орт направления:

$$\frac{\partial f}{\partial e} = \nabla f \cdot \vec{e}$$

Отсюда следует, что максимальное значение в точке производная по направлению принимает, если направление совпадает с направлением градиента функции в данной точке.

НЕТОЛОГИЯ

https://ru.wikipedia.org

Градиент - направление наискорейшего роста функции

Если функция дифференцируема в (x0,y0), то в окрестности её можно приблизить линейно

$$\Delta f \approx \left\langle \nabla f(x_0, y_0), \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right\rangle$$

Пусть $\vec{\ell} \in \mathbb{R}^2$, $|\vec{\ell}| = 1$, тогда приращения можно задать вдоль вектора $\vec{\ell}$:

$$\Delta x = t \cdot \ell_x, \qquad \Delta y = t \cdot \ell_y.$$

Подставим в первое выражение:

$$\Delta f \approx \left\langle \nabla f(x_0, y_0), \begin{pmatrix} t \cdot \ell_x \\ t \cdot \ell_y \end{pmatrix} \right\rangle = t \cdot \left\langle \nabla f(x_0, y_0), \vec{\ell} \right\rangle$$

https://www.coursera.org/learn/mathematics-and-python

Градиент - направление наискорейшего роста функции

$$\frac{\partial f}{\partial \vec{\ell}}(x_0, y_0) = \lim_{t \to 0} \frac{\Delta f}{t} = \left\langle \nabla f(x_0, y_0), \vec{\ell} \right\rangle$$

Таким образом, производная по направлению может быть вычислена как скалярное произведение градиента на соответствующий единичный вектор:

$$\frac{\partial f}{\partial \vec{\ell}}(x_0, y_0) = \left\langle \nabla f(x_0, y_0), \vec{\ell} \right\rangle$$

Согласно этой формуле направление максимального роста - направление задаваемое градиентом, а оно максимально при сонаправленности векторов.

$$\left\langle \nabla f(x_0, y_0), \vec{\ell} \right\rangle$$

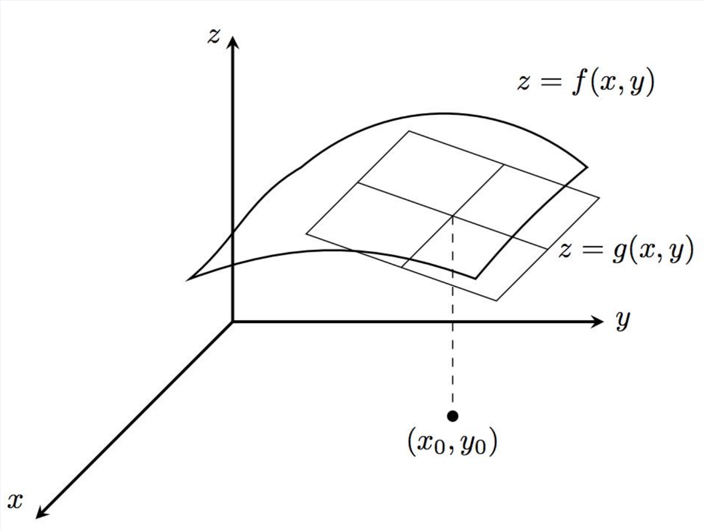
НЕТОЛОГИЯ

https://www.coursera.org/learn/mathematics-and-python

Касательная плоскость и линейное приближение

Пусть функция $f(x, y)$ дифференцируема в точке (x_0, y_0), тогда в окрестности этой точки можно записать:

$$f(x_0 + \Delta x, y_0 + \Delta y) \approx f(x_0, y_0) + f'_x(x_0, y_0)\Delta x + f'_y(x_0, y_0)\Delta y.$$



$$\Delta f \approx \left\langle \nabla f(x_0, y_0), \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right\rangle$$

Выражение для delta f линейно по delta x и delta y

$$f(x, y) \approx f(x_o, y_o) + \left\langle \nabla f(x_o, y_o), \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right\rangle$$

https://www.coursera.org/learn/mathematics-and-python

Спасибо за внимание!

НЕТОЛОГИЯ