Производные и градиентный спуск • Gradient descent

Производная функции

Производная функции характеризует скорость изменения функции в данной точке.

Через предел: производной функции f в точке x_0 называется предел:

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

Производная функции

Производная функции характеризует скорость изменения функции в данной точке.

Через предел: производной функции f в точке x_0 называется предел:

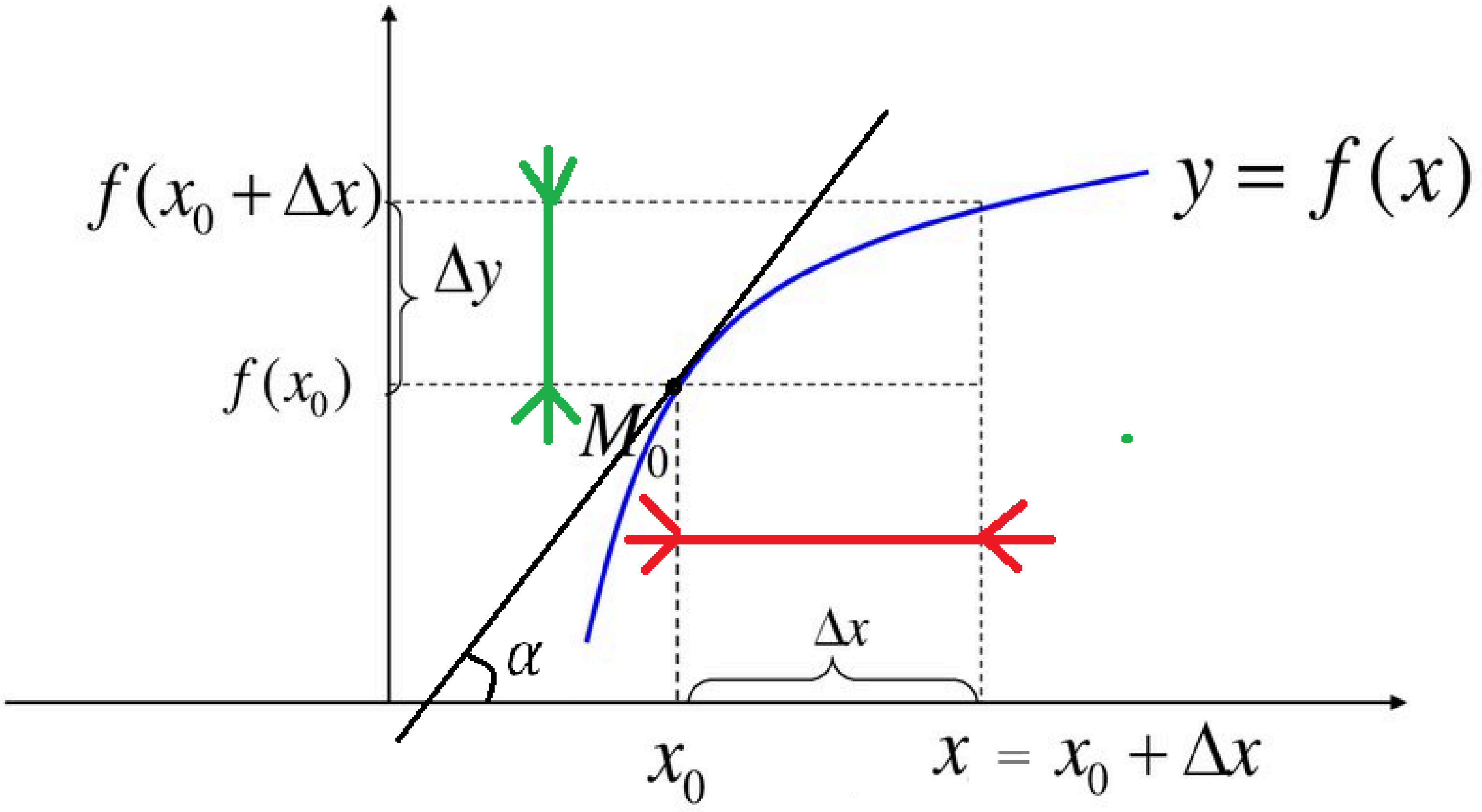
$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f(x)}{\Delta x}$$

Геометрически: если функция f имеет конечную производную в точке x_0, то ее можно приблизить линейной функцией:

$$f_l(x) = f(x_0) + f'(x_0)(x - x_0)$$

Тогда функция f_l называется *касательной* к f в точке x_0, а число $f'(x_0)$ является *тангенсом* угла наклона касательной прямой.

Производная функции



$$tg(\alpha) = f'(x_0)$$

Элементарные производные

Эльбрус буткемп

$c' = 0, c = const$	$\ln x = \frac{1}{x}$	$(\log_a x)' = \frac{1}{x \ln a}$
$(x^n)' = nx^{n-1}$	$(\sin x)' = \cos x$	$\operatorname{tg} x = \frac{1}{\cos^2 x}$
$(a^x)' = a^x \ln a$	$(\cos x)' = -\sin x$	$(\sqrt{x})' = \frac{1}{2\sqrt{x}}$
$(e^x)' = e^x$		

Свойства производных

1. $(f(x)+g(x))' = f'(x)+g'(x)$
2. $(f(x)-g(x))' = f'(x)-g'(x)$
3. $(f(x)g(x))' = f'(x)g(x)+f(x)g'(x)$
4. $(Cf(x))' = Cf'(x)$
5. $\left(\dfrac{f(x)}{g(x)}\right)' = \dfrac{f'(x)g(x)-f(x)g'(x)}{g^2(x)}$

Эльбрус буткемп

Производная сложной функции

Пусть y - сложная функция, т.е. $y = f(u), u = g(x)$. Если обе функции дифференцируемые, то сложная функция тоже дифференцируема в точке x и производная в точке равна:

$$y'(x) = f'(u)g'(x)$$

Производная сложной функции

Пусть y - сложная функция, т.е. $y = f(u), u = g(x)$. Если обе функции дифференцируемые, то сложная функция тоже дифференцируема в точке x и производная в точке равна:

$$y'(x) = f'(u)g'(x)$$

Пример: $y = \operatorname{tg} x^2$.

Что здесь $f(u)$ и $g(x)$?

Какая производная?

Производная сложной функции

Пусть y - сложная функция, т.е. $y = f(u), u = g(x)$. Если обе функции дифференцируемые, то сложная функция тоже дифференцируема в точке x и производная в точке равна:

$$y'(x) = f'(u)g'(x)$$

Пример: $y = \operatorname{tg} x^2$.

$f(u) = \operatorname{tg} u\,; g(x) = x^2$

$$y' = \frac{1}{\cos^2 x^2} \times (x^2)' = \frac{2x}{\cos^2 x^2}$$

Найти производную сложной функции

Эльбрус буткемп

$$y = \operatorname{arctg}^2 x^3$$

$$y' = ?$$

Найти производную сложной функции

$$y = \ln \cos^2(\sin \sqrt{x})$$

$$y' = ?$$

Найти производную функции

$\sigma(x)$ - Одна из самых часто встречаемых в машинном и глубоком обучении

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\sigma'(x) = ?$$

Найти производную функции

Эльбрус буткемп

- Сигмоида

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\sigma'(x) = \frac{e^{-x}}{(1 + \mathrm{e}^{-x})^2}$$

Еще ее часто выражают таким образом:

$$\frac{d}{dx}\sigma(x) = \sigma(x)(1 - \sigma(x))$$

Найти производную функции

$$\frac{d}{dx}\sigma(x) = \frac{d}{dx}\left[\frac{1}{1+e^{-x}}\right] = \frac{d}{dx}(1+e^{-x})^{-1}$$

$$= -1 * (1+e^{-x})^{-2}(-e^{-x}) = \frac{-e^{-x}}{-(1+e^{-x})^2} = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{1}{1+e^{-x}}\frac{e^{-x}}{1+e^{-x}} = \frac{1}{1+e^{-x}}\frac{e^{-x}+(1-1)}{1+e^{-x}}$$

$$= \frac{1}{1+e^{-x}}\frac{(1+e^{-x})-1}{1+e^{-x}} = \frac{1}{1+e^{-x}}\left[\frac{(1+e^{-x})}{1+e^{-x}} - \frac{1}{1+e^{-x}}\right]$$

$$= \frac{1}{1+e^{-x}}\left[1 - \frac{1}{1+e^{-x}}\right] = \sigma(x)(1-\sigma(x))$$

Если кому-то интересно, почему ее можно выразить по-другому.

Градиентный спуск

Эльбрус буткемп

- Дифференцирование используется для минимизации функции по параметрам
- Рассмотрим пример:

$$L(w) = w^2 + 12$$

- Производная функции:

$$\frac{\partial L}{\partial w} = 2w$$

- Эта функция очень простая, мы можем найти ее минимум аналитически
- Но мы воспользуемся процедурой градиентного спуска для поиска минимума по параметру w

Градиентный спуск

Эльбрус буткемп

Общий алгоритм

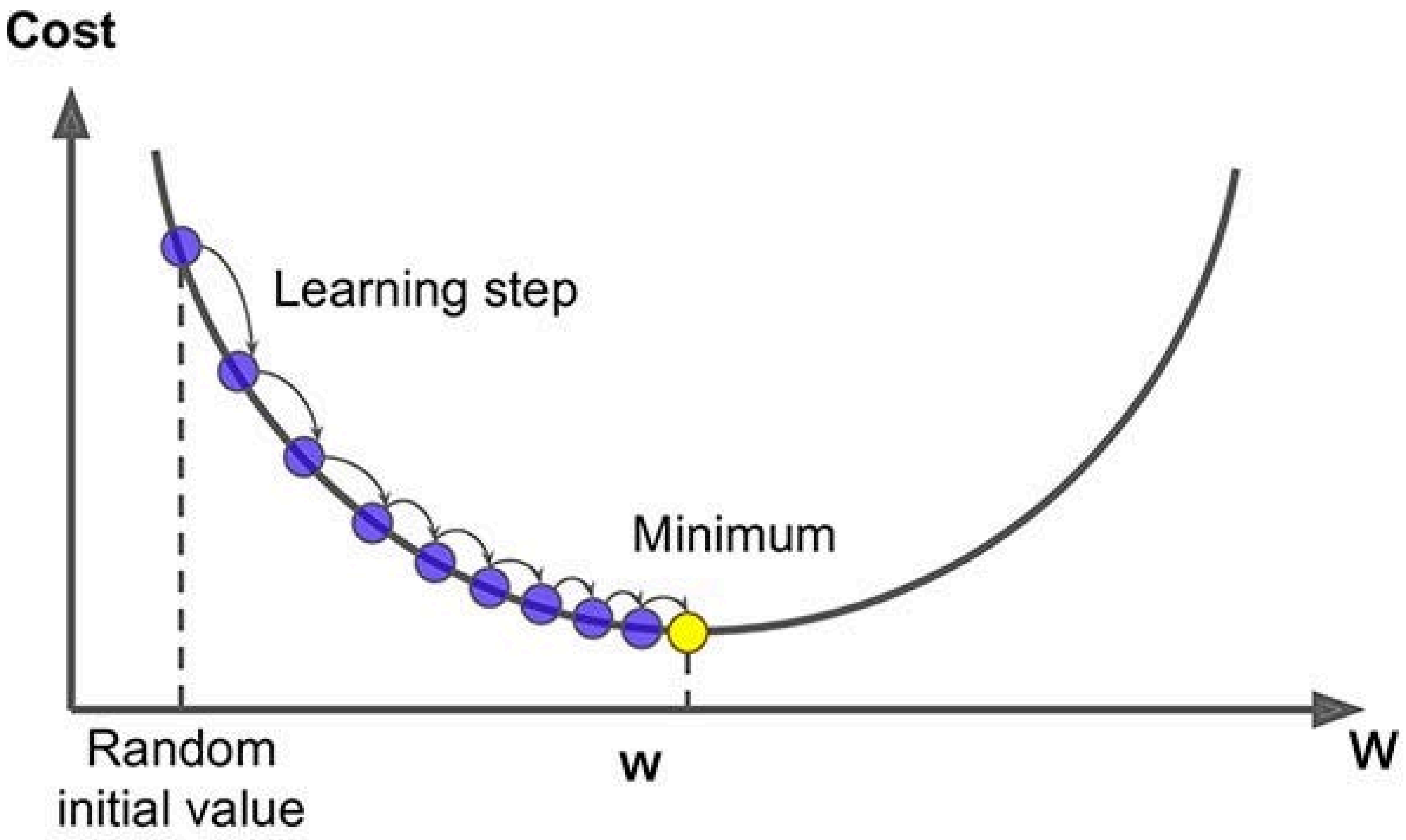
- Дана функция $L(w)$, надо найти значение параметра w, при котором значение функции будет минимальным: $L(w) \to \min_w$, w_0 = random
- Производная функции по параметру w равна $\dfrac{\partial L}{\partial w}$
- На каждой итерации обновляем значение параметра w по правилу:

$$w_{i+1} = w_i - \lambda \frac{\partial L}{\partial w}$$

- λ - скорость обучения (`learning_rate`)
- Вычисляем значение $L(w_{i+1})$
- Останавливаемся, если $|L(w_{i+1}) - L(w_i)| < \epsilon$, ϵ - параметр (напр., 0.01)

📊Let's play with gradient descent

Градиентный спуск



Итоги

- на основе итеративной процедуры корректировки параметров работает много алгоритмов машинного обучения
- градиент - вектор частных производных по каждому параметру
- у обычного градиентного спуска есть проблемы, для алгоритмов в машинном обучении используются более сложные варианты