Анализ данных

Семинар 4. Производные и градиентный спуск.

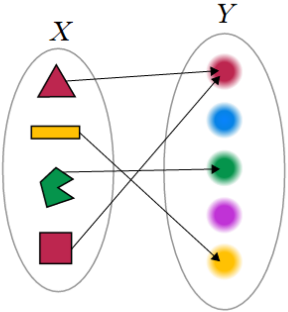
2 февраля 2016

Основы мат.анализа

Говорят, что на множестве X **задана функция** f, принимающая значения из Y, если каждому элементу x из множества X поставлен в соответствие некоторый элемент y из множества Y.

Функцию обозначают также следующими способами:

$y = f(x); \quad f : X \to Y.$



Основы мат.анализа

Примеры математических функций:

- $f(x) = x^3, X = \mathbb{R}$ (вещ. числа), $Y = \mathbb{R}$;

Основы мат.анализа

Примеры математических функций:

- $f(x) = x^3, X = \mathbb{R}$ (вещ. числа), $Y = \mathbb{R}$;
- $f(x) = sin(x), X = \mathbb{R},\ Y = [-1; 1]$;

Основы мат.анализа

Примеры математических функций:

- $f(x) = x^3, X = \mathbb{R}$ (вещ. числа), $Y = \mathbb{R}$;
- $f(x) = sin(x), X = \mathbb{R}$, $Y = [-1; 1]$;
- $f(x, y) = sin(xy), X = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ (пары вещ. чисел), $Y = [-1; 1]$;

Основы мат.анализа

Примеры математических функций:

- $f(x) = x^3, X = \mathbb{R}$ (вещ. числа), $Y = \mathbb{R}$;
- $f(x) = sin(x), X = \mathbb{R},\ Y = [-1; 1]$;
- $f(x, y) = sin(xy), X = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ (пары вещ. чисел), $Y = [-1; 1]$;
- моном n-ой степени:
  $f(x) = C * x^n, X = \mathbb{R}, Y = \mathbb{R}, C$ — константа;

Основы мат.анализа

Примеры математических функций:

- $f(x) = x^3, X = \mathbb{R}$ (вещ. числа), $Y = \mathbb{R}$;
- $f(x) = sin(x), X = \mathbb{R}, \ Y = [-1; 1]$;
- $f(x, y) = sin(xy), X = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ (пары вещ. чисел), $Y = [-1; 1]$;
- моном n-ой степени:
  $f(x) = C * x^n, X = \mathbb{R}, Y = \mathbb{R}, C$ — константа;
- полином n-ой степени:
  $f(x) = \alpha_n * x^n + \ldots, +\alpha_1 * x + \alpha_0, X = \mathbb{R}, Y = \mathbb{R}$,
  $\alpha_i, i = \overline{0, n}$, — константы.

Основы мат.анализа

Примеры математических функций:

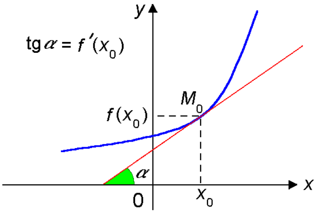
- $f(x) = x^3, X = \mathbb{R}$ (вещ. числа), $Y = \mathbb{R}$;
- $f(x) = sin(x), X = \mathbb{R},\ Y = [-1; 1]$;
- $f(x, y) = sin(xy), X = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ (пары вещ. чисел), $Y = [-1; 1]$;
- моном n-ой степени:
  $f(x) = C * x^n, X = \mathbb{R}, Y = \mathbb{R}, C$ — константа;
- полином n-ой степени:
  $f(x) = \alpha_n * x^n + \dots, +\alpha_1 * x + \alpha_0, X = \mathbb{R}, Y = \mathbb{R}$,
  $\alpha_i, i = \overline{0, n},$ — константы.
- и т.д.

Основы мат.анализа

Пусть в некоторой окрестности точки $x_0 \in \mathbb{R}$ определена функция f. **Производной** функции f в точке x_0 называется предел (если он существует)

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

Смысл производной — наклон касательной к графику функции в данной точке.



Основы мат.анализа

Производная функции — это функция!

Если функция получена при помощи арифметических операций и композиции из более простых, её производная может быть найдена при помощи следующих правил:

1. $(c * f(x))' = c * f'(x), c$ — константа;
2. $(f(x) \pm g(x))' = f'(x) \pm g'(x)$;
3. $(f(x) * g(x))' = f'(x)g(x) + f(x)g'(x)$;
4. $(\frac{f(x)}{g(x)})' = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$;
5. $f(g(x)) = g'(x) * f'(g(x))$.

Основы мат.анализа

Для большинства широко используемых функций производные известны и могут быть найдены в таблицах производных.

1. $C' = 0$;
2. $x' = 1$;
3. $(x^2)' = 2x$;
4. $(x^n)' = n \cdot x^{n-1}$;
5. $(a^x)' = a^x \cdot \ln a$;

5* $(e^x)' = e^x$;

6. $(\log_a x)' = \dfrac{1}{x \ln a}$;

6*. $(\ln x)' = \dfrac{1}{x}$;

7. $(\sin x)' = \cos x$;
8. $(\cos x)' = -\sin x$;
9. $(tg\, x)' = \dfrac{1}{\cos^2 x}$;
10. $(ctg\, x)' = -\dfrac{1}{\sin^2 x}$;
11. $(\arcsin x)' = \dfrac{1}{\sqrt{1-x^2}}$;
12. $(\arccos x)' = -\dfrac{1}{\sqrt{1-x^2}}$;
13. $(arctg\, x)' = \dfrac{1}{1+x^2}$;
14. $(arcctg\, x)' = -\dfrac{1}{1+x^2}$.

Основы мат.анализа

Частная производная — обощение понятия производная на случай функции нескольких переменных.
Частная производная функции $f(x_1, \ldots, x_n)$ по переменной x_k в точке $(a_1, \ldots, a_n)$ — это предел

$$\frac{\partial f}{\partial x_k}(a_1, \cdots, a_n) =$$

$$= \lim_{\Delta x \to 0} \frac{f(a_1, \ldots, a_k + \Delta x, \ldots, a_n) - f(a_1, \ldots, a_k, \ldots, a_n)}{\Delta x}$$

$\frac{\partial f}{\partial x_k}$ **— это обозначение, а не деление!**

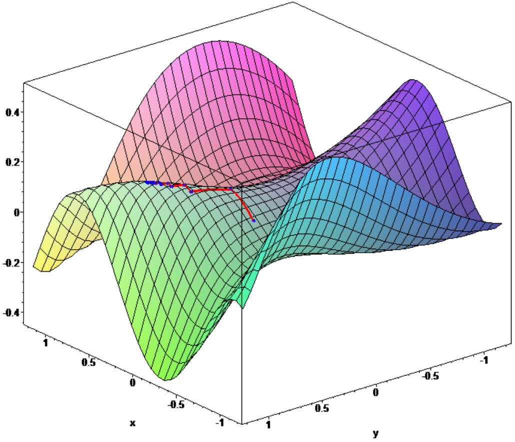
Градиентом функции $f(x_1, \ldots, x_n)$ называется n-мерный вектор, составленный из частных производных

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \; \ldots, \; \frac{\partial f}{\partial x_n} \right).$$

В каждой точке $(a_1, \ldots, a_n)$ градиент функции принимает конкретное значение.

Основы мат.анализа

Кроме того, градиент — вектор, своим направлением указывающий **направление наибольшего возрастания функции многих переменных**.



Метод градиентного спуска

Пусть у нас есть некоторая функция $f(x_1, \ldots, x_n)$.
Наша цель — найти точку $x^* = (x_1^*, \ldots, x_n^*)$, в которой достигается её минимальное значение.

- Рассмотрим произвольную точку $a^{(0)}$. Значение градиента $\nabla f(a^{(0)})$ в этой точке — направление наибольшего роста функции.
- Поэтому попробуем сдвинуться в ровно противоположном направлении (по вектору **антиградиента**) — в новую точку $a^{(1)}$.
- ...И так до тех пор, пока не найдём точку, в которой любое направление движения будет бессмысленным.

Метод градиентного спуска

Как далеко нужно «сдвигаться» на каждой итерации алгоритма?

- Пусть $a^{(i)}$ — полученное на i-ой итерации приближение.
- Тогда положим

$$a^{(i+1)} = a^{(i)} - \alpha \nabla f(a^{(i)}).$$

Параметр α называется **темпом обучения**.

Метод градиентного спуска

Когда нужно остановиться?

- после определенного количества итераций;

Метод градиентного спуска

Когда нужно остановиться?

- после определенного количества итераций;
- когда значение точки на соседних итерациях практически не меняется;

Метод градиентного спуска

Когда нужно остановиться?

- после определенного количества итераций;
- когда значение точки на соседних итерациях практически не меняется;
- когда значение оптимизируемой функции на соседних итерациях практически не меняется;
- и т.д.

Метод градиентного спуска

Тем не менее:

1. При неудачном выборе начального приближения метод градиентного спуска может найти локальный, а не глобальный оптимум.

Метод градиентного спуска

Тем не менее:

1. При неудачном выборе начального приближения метод градиентного спуска может найти локальный, а не глобальный оптимум.
2. При неудачном выборе темпа обучения α метод может разойтись.

Метод градиентного спуска

Зачем это нужно?

Напоминание:

- X — пространство объектов,
  Y — пространство ответов,
  X^ℓ — обучающая выборка,
  $a : X \to Y$ — восстановленная по обучающей выборке зависимость.

Метод градиентного спуска

Зачем это нужно?

Напоминание:

- X — пространство объектов,
  Y — пространство ответов,
  X^ℓ — обучающая выборка,
  $a : X \to Y$ — восстановленная по обучающей выборке зависимость.
- $Q(a, X^\ell)$ — оптимизируемый функционал качества.
  Алгоритм a, как правило, характеризуется набором параметров $\alpha = (\alpha_1, \ldots, \alpha_s)$.
  Т.о., необходимо решить задачу: $Q(\alpha, X^\ell) \to min_\alpha$.

Метод градиентного спуска

Зачем это нужно?

Напоминание:

- X — пространство объектов,
  Y — пространство ответов,
  X^ℓ — обучающая выборка,
  $a : X \to Y$ — восстановленная по обучающей выборке зависимость.
- $Q(a, X^\ell)$ — оптимизируемый функционал качества.
  Алгоритм a, как правило, характеризуется набором параметров $\alpha = (\alpha_1, \ldots, \alpha_s)$.
  Т.о., необходимо решить задачу: $Q(\alpha, X^\ell) \to min_\alpha$.
- В качестве метода оптимизации можно (иногда нет другого выхода) использовать итерационные методы оптимизации (в частности, метод градиентного спуска).

Метод градиентного спуска

Пример: решаем задачу предсказания стоимости квартиры в зависимости от её площади.

- Задача регрессии

Метод градиентного спуска

Пример: решаем задачу предсказания стоимости квартиры в зависимости от её площади.

- Задача регрессии
- $X^\ell = (x_i, y_i)_{i=1}^{\ell}$ — пары (площадь, стоимость) для ℓ квартир.

Метод градиентного спуска

Пример: решаем задачу предсказания стоимости квартиры в зависимости от её площади.

- Задача регрессии
- $X^\ell = (x_i, y_i)_{i=1}^{\ell}$ — пары (площадь, стоимость) для ℓ квартир.
- Будем восстанавливать линейную зависимость, т.е. использовать модель: $a(x) = \alpha_1 x + \alpha_0$.

Метод градиентного спуска

Пример: решаем задачу предсказания стоимости квартиры в зависимости от её площади.

- Задача регрессии
- $X^\ell = (x_i, y_i)_{i=1}^{\ell}$ — пары (площадь, стоимость) для ℓ квартир.
- Будем восстанавливать линейную зависимость, т.е. использовать модель: $a(x) = \alpha_1 x + \alpha_0$.
- Будем оптимизировать среднеквадратичную ошибку:

$$Q(\alpha, X^\ell) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i)-y_i)^2 = \frac{1}{\ell}\sum_{i=1}^{\ell}(\alpha_1 x_i+\alpha_0-y_i)^2 \to min_{\alpha_0,\alpha_1}$$

Метод градиентного спуска

Пример: решаем задачу предсказания стоимости квартиры в зависимости от её площади.

- Задача регрессии
- $X^\ell = (x_i, y_i)_{i=1}^{\ell}$ — пары (площадь, стоимость) для ℓ квартир.
- Будем восстанавливать линейную зависимость, т.е. использовать модель: $a(x) = \alpha_1 x + \alpha_0$.
- Будем оптимизировать среднеквадратичную ошибку:

$$Q(\alpha, X^\ell) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2 = \frac{1}{\ell}\sum_{i=1}^{\ell}(\alpha_1 x_i + \alpha_0 - y_i)^2 \to min_{\alpha_0, \alpha_1}$$

- Можем выписать градиент функции Q по переменным α и применить метод градиентного спуска.