Задачи оптимизации и поиска корней функции

Мальковский Н. В.

Санкт-Петербургский Академический Университет

Задача нахождения корня функции

Пусть $f : D \to \mathbb{R}^n$. Корнем функции называется точка $x^* \in D$, такая что

$$f(x^*) = 0_n.$$

Традиционно задачу о нахождении корня функции попросту называют уравнением в случае $n = 1$, и системой уравнений в случае $n > 1$.

Задача оптимизации

Пусть $f : D \to \mathbb{R}^n$. Глобальным минимумом функции f на D называется точка x^*, такая что

$$\forall x \in D : \; f(x) \geq f(x^*).$$

Задача оптимизации

Пусть $f : D \to \mathbb{R}^n$. Глобальным минимумом функции f на D называется точка x^*, такая что

$$\forall x \in D: \; f(x) \geq f(x^*).$$

Обозначается

$$\operatorname*{argmin}_{x \in D} f(x).$$

Задача оптимизации

Пусть $f : D \to \mathbb{R}^n$. Глобальным минимумом функции f на D называется точка x^*, такая что

$$\forall x \in D: \ f(x) \geq f(x^*).$$

Обозначается

$$\operatorname*{argmin}_{x \in D} f(x).$$

Аналогично, *глобальный максимум* f на D – такая точка x^*, что

$$\forall x \in D: \ f(x) \geq f(x^*), \ x^* = \operatorname*{argmax}_{x \in D} f(x).$$

Задача оптимизации

Пусть $f : D \to \mathbb{R}^n$. Глобальным минимумом функции f на D называется точка x^*, такая что
$$\forall x \in D: \ f(x) \geq f(x^*).$$
Обозначается
$$\operatorname*{argmin}_{x \in D} f(x).$$
Аналогично, *глобальный максимум* f на D – такая точка x^*, что
$$\forall x \in D: \ f(x) \geq f(x^*), \ x^* = \operatorname*{argmax}_{x \in D} f(x).$$

Задача оптимизации – нахождение либо глобального минимума, либо глобального максимума.

Задача оптимизации

Пусть $f : D \to \mathbb{R}^n$. Глобальным минимумом функции f на D называется точка x^*, такая что

$$\forall x \in D: \ f(x) \geq f(x^*).$$

Обозначается

$$\operatorname*{argmin}_{x \in D} f(x).$$

Аналогично, *глобальный максимум* f на D – такая точка x^*, что

$$\forall x \in D: \ f(x) \geq f(x^*), \ x^* = \operatorname*{argmax}_{x \in D} f(x).$$

Задача оптимизации – нахождение либо глобального минимума, либо глобального максимума.

Замечание. Задача максимизации f – это задача минимизации “$-f$”. Большинство задач оптимизации формулируются в терминах нахождения минимума. Далее под x^* будет обычно обозначаться точка глобального минимума.

Связь задач нахождения минимума и корня функций

Заметим, что

$$f(x^*) = 0 \;\Rightarrow\; x^* = \underset{x}{\operatorname{argmin}} |f(x)| = \underset{x}{\operatorname{argmin}} f(x)^2.$$

Связь задач нахождения минимума и корня функций

Заметим, что

$$f(x^*) = 0 \;\Rightarrow\; x^* = \operatorname*{argmin}_x |f(x)| = \operatorname*{argmin}_x f(x)^2.$$

Таким нехитрым образом можно свести задачу нахождения корня к задаче нахождения минимума. В обратную сторону сведение можно получить только для дифференцируемых функций.

Теорема Ферма

Теорема

$f : \mathcal{D} \subset \mathbb{R}^n$, x^* – внутренняя точка $\mathcal{D}$ и f дифференцируема в точке x^*. Если x^* – точка минимума f на $\mathcal{D}$, тогда выполняется условие:

$$\nabla f(x^*) = 0_n.$$

Теорема Ферма

Теорема

$f : \mathcal{D} \subset \mathbb{R}^n$, x^* – внутренняя точка $\mathcal{D}$ и f дифференцируема в точке x^*. Если x^* – точка минимума f на $\mathcal{D}$, тогда выполняется условие:

$$\nabla f(x^*) = 0_n.$$

Док-во: Пусть $\nabla f(x_*) \neq 0_n$, $\epsilon > 0$: $B(x, \epsilon) \subset \mathcal{D}$.
В силу дифференцируемости f в точке x^*

$$f(x) = f(x^*) + \nabla f(x^*)(x - x^*) + \alpha(x - x^*)(x - x^*),$$

$$\alpha(x) \to 0 \ (x \to 0).$$

Теорема Ферма

Теорема

$f : \mathcal{D} \subset \mathbb{R}^n$, x^* – внутренняя точка $\mathcal{D}$ и f дифференцируема в точке x^*. Если x^* – точка минимума f на $\mathcal{D}$, тогда выполняется условие:

$$\nabla f(x^*) = 0_n.$$

Док-во: Пусть $\nabla f(x_*) \neq 0_n$, $\epsilon > 0$: $B(x, \epsilon) \subset \mathcal{D}$.
В силу дифференцируемости f в точке x^*

$$f(x) = f(x^*) + \nabla f(x^*)(x - x^*) + \alpha(x - x^*)(x - x^*),$$

$$\alpha(x) \to 0 \ (x \to 0).$$

Пусть $x' = x^* - \delta \nabla f(x^*) \neq x^*$, $\delta > 0$ достаточно малое, чтобы выполнялось $\alpha(x') \leq |\frac{\nabla f(x^*)}{2}|$ и $x' \in B(x^*, \epsilon)$, тогда

$$f(x') = f(x^*) - \delta |\nabla f(x^*)|^2 + \alpha(x')(x' - x^*) \leq f(x^*) - \frac{\delta}{2} |\nabla f(x^*)|^2 < f(x^*).$$

Т.е. x^* не точка минимума. ■

Стационарность

Замечания:

- Данную теорему обычно называют условием стационарности или условиями оптимальности первого порядка. Точку, удовлетворяющей этим условием называют стационарной точкой или критической точкой.

Стационарность

Замечания:

- Данную теорему обычно называют <u>условием стационарности</u> или условиями оптимальности первого порядка. Точку, удовлетворяющей этим условием называют <u>стационарной точкой</u> или критической точкой.
- Условия стационарности не являются достаточными, как например для точки 0 функции $f(x) = x^3$.

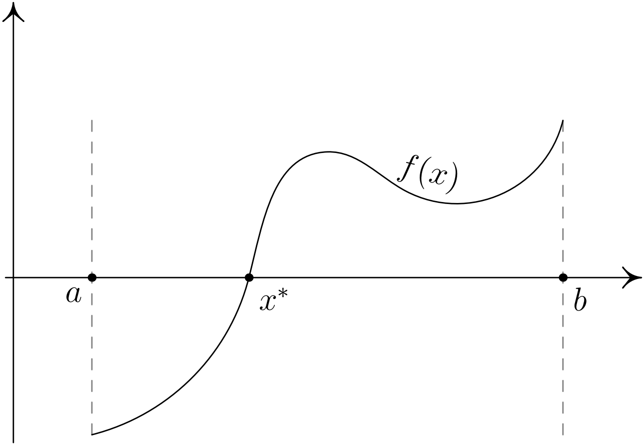
Стационарность

Замечания:

- Данную теорему обычно называют условием стационарности или условиями оптимальности первого порядка. Точку, удовлетворяющей этим условием называют стационарной точкой или критической точкой.
- Условия стационарности не являются достаточными, как например для точки 0 функции $f(x) = x^3$.
- Точка x^* называется *локальным минимумом* f на D, если существует $\epsilon > 0$, что
$$\forall x \in D,\ ||x - x^*|| < \epsilon\ f(x) \geq f(x^*).$$
Если для для f в точке x^* гессиан положительно определен, то x^* гарантированно является точка локального минимума. Это свойство принято называть условием оптимальности второго порядка.

Метод бисекции

Пусть $f : [a, b] \to \mathbb{R}$, f непрерывна и $f(a) < 0 < f(b)$. Требуется найти такую точку $x^* \in (a, b)$, что $f(x^*) = 0$. Из непрерывности f следует, что x^* существует.



Алгоритм:

- Разделить отрезок $[a, b]$ пополам и посмотреть значение f посередине.
- Если $f\left(\frac{a+b}{2}\right) > 0$, отбросить правую половину

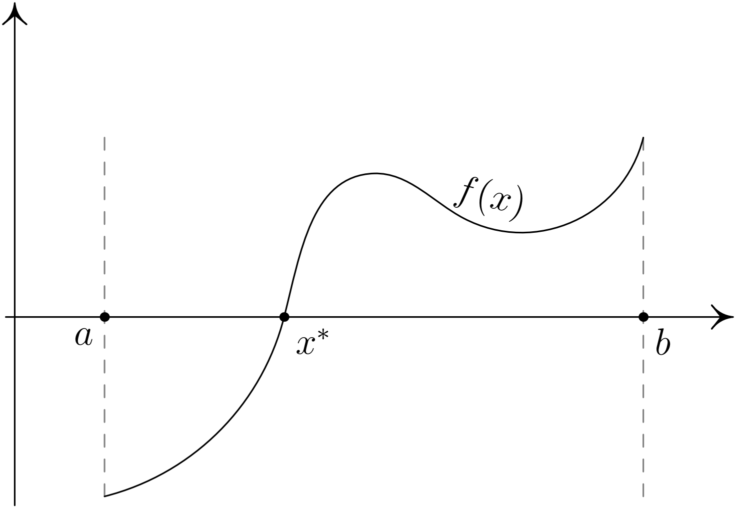
$$b \leftarrow \frac{a+b}{2}.$$

- Если $f\left(\frac{a+b}{2}\right) < 0$, отбросить правую половину

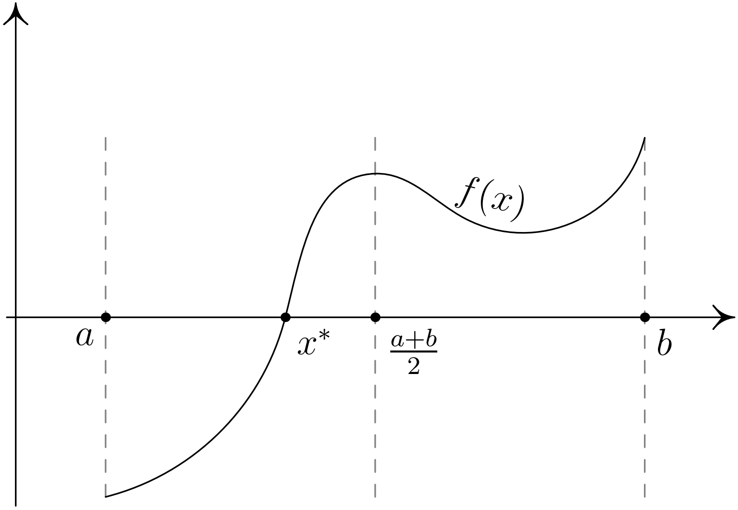
$$a \leftarrow \frac{a+b}{2}.$$

- Если $f\left(\frac{a+b}{2}\right) = 0$, то решение найдено.
- Если $b - a > \epsilon$, повторить сначала.

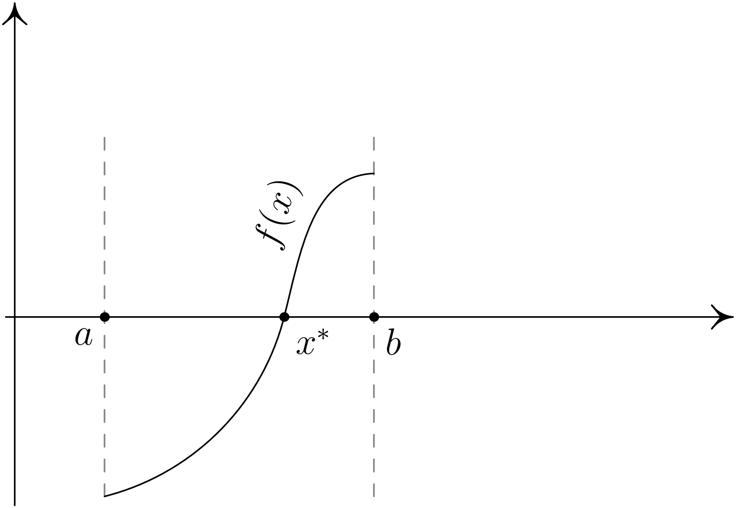
Метод бисекции



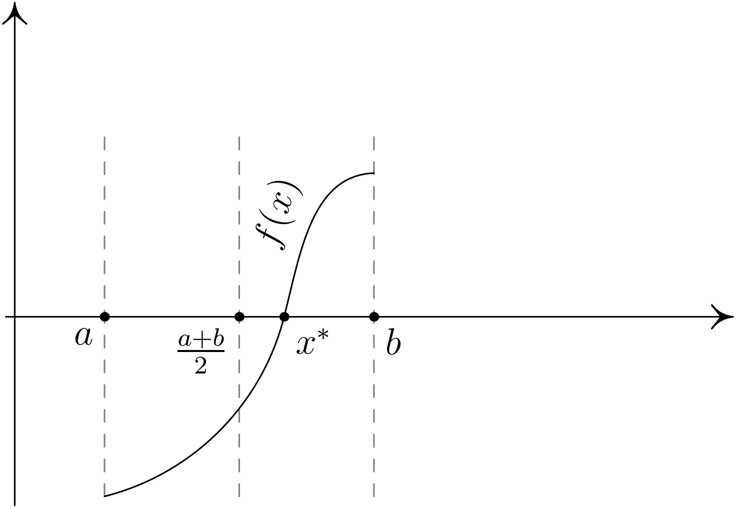
Метод бисекции



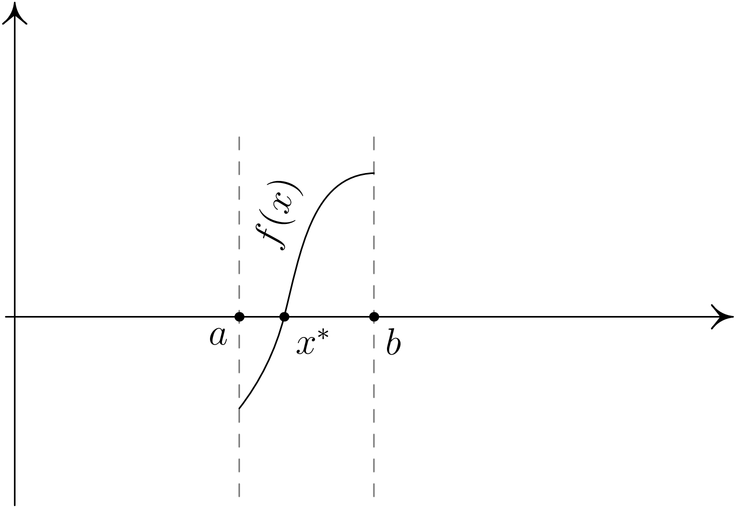
Метод бисекции



Метод бисекции



Метод бисекции



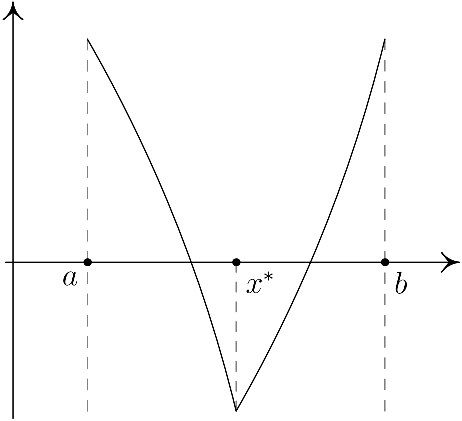
Метод бисекции

Свойства:

- На каждом этапе алгоритма выполняется свойство $f(a) < 0 < f(b)$, а значит алгоритм всегда работает с отрезком, который содержит корень.
- Не находит точное значение корня, но может найти приближенное значение с любой заданной точностью.
- Для нахождения корня с точностью $\epsilon > 0$ необходимо сделать $\log_2\left(\frac{b-a}{\epsilon}\right)$ итераций.
- * Иногда этот метод называют *методом дихотомии*, в дискретном случае этом метод обычно называют *двоичным* или *бинарным поиском*.

Тернарный поиск

Пусть $f : [a, b] \to \mathbb{R}$, $\exists x^* \in (a, b) : f$ строго убывает на $[a, x^*]$, f строго возрастает на $[x^*, b]$. Нужно найти минимум функции f на $[a, b]$, т.е. точку x^*. Функцию, обладающую подобным свойством принято называть унимодальной.



Тернарный поиск

Алгоритм:

- Выбрать любые две точки $a < m_1 < m_2 < b$.
- Если $f(m_1) < f(m_2)$, то $x^* \notin [m_2, b]$,

$$b \leftarrow m_2.$$

- Если $f(m_1) > f(m_2)$, то $x^* \notin [a, m_1]$.

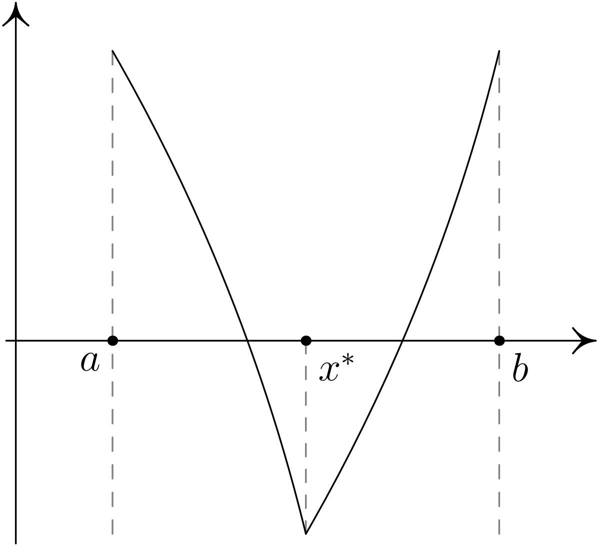
$$a \leftarrow m_1.$$

- Если $f(m_1) = f(m_2)$, то $x^* \notin [m_2, b] \cup [a, m_1]$,

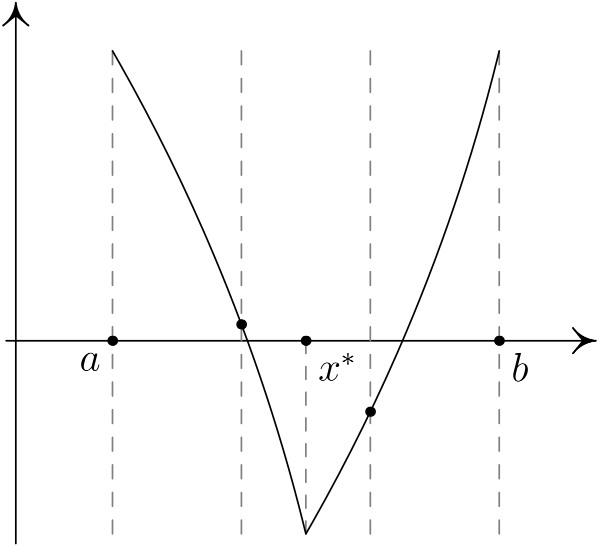
$$a \leftarrow m_1,\ b \leftarrow m_2.$$

- Если $b - a > \epsilon$, повторить сначала.

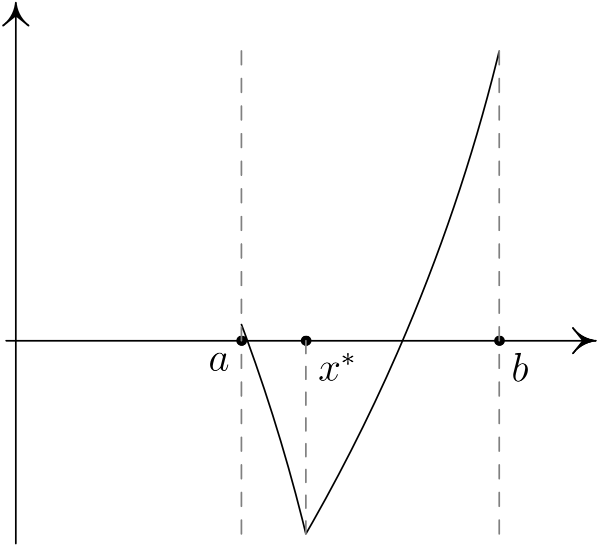
Тернарный поиск



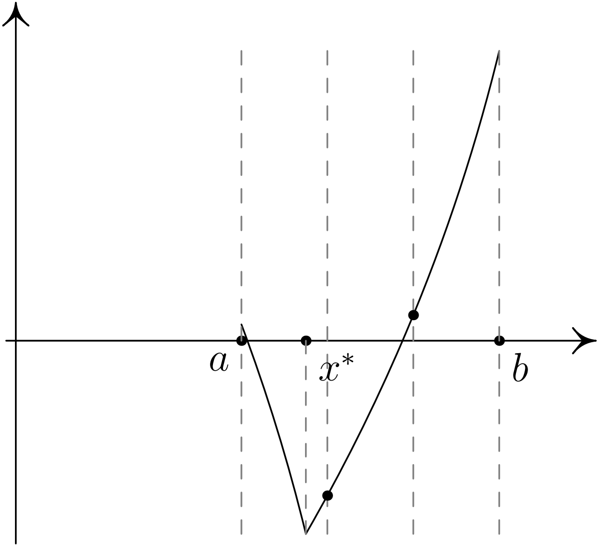
Тернарный поиск



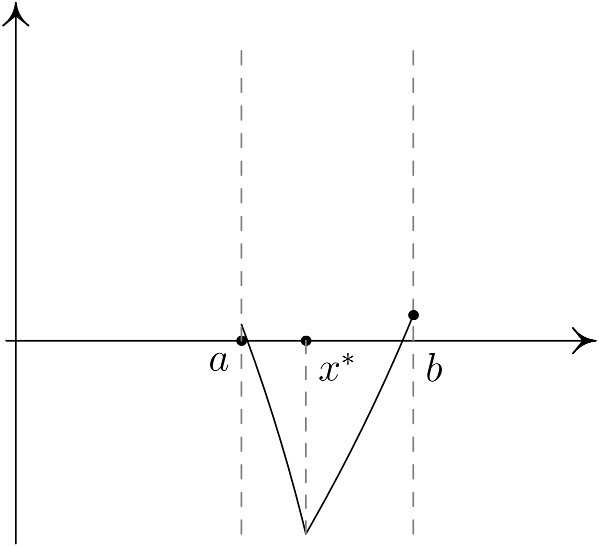
Тернарный поиск



Тернарный поиск



Тернарный поиск



Тернарный поиск

Свойства:

- На каждом этапе алгоритма выполняется свойство $a < x^* < b$.
- Опять же, не получится найти минимум точно, только с любой заданной точностью.
- Если точки делят отрезок на 3 равные части, то для нахождения минимума с точностью $\epsilon > 0$ необходимо сделать $\log_{\frac{3}{2}} \frac{b-a}{\epsilon}$ итераций.

Оптимальность

Пусть D – некоторое множество, есть некоторый оракул, который загадывает элемент x из этого множества. Нам нужно отгадать этот элемент, при этом для любого подмножества $X \subset D$ мы можем спрашивать, принадлежит x множеству X? Требуется отгадать элемент x или же найти множество заданного размер, содержащее его?

Оптимальность

Пусть D – некоторое множество, есть некоторый оракул, который загадывает элемент x из этого множества. Нам нужно отгадать этот элемент, при этом для любого подмножества $X \subset D$ мы можем спрашивать, принадлежит x множеству X? Требуется отгадать элемент x или же найти множество заданного размер, содержащее его?

Метод бисекции и тернарный поиск как раз представляют собой подобную процедуру. Они последовательно задают вопрос, принадлежит ли элемент некоторому отрезку или нет? Нетрудно показать, что оценка $\log_2 \frac{b-a}{\epsilon}$ является оптимальной для таких процедур.

Оптимальность

Пусть D – некоторое множество, есть некоторый оракул, который загадывает элемент x из этого множества. Нам нужно отгадать этот элемент, при этом для любого подмножества $X \subset D$ мы можем спрашивать, принадлежит x множеству X? Требуется отгадать элемент x или же найти множество заданного размер, содержащее его?

Метод бисекции и тернарный поиск как раз представляют собой подобную процедуру. Они последовательно задают вопрос, принадлежит ли элемент некоторому отрезку или нет? Нетрудно показать, что оценка $\log_2 \frac{b-a}{\epsilon}$ является оптимальной для таких процедур.

Теорема

Пусть $\mathcal{F}$ – σ-алгебра на множестве D и μ – мера на $\mathcal{F}$, при этом $\forall A \in \mathcal{F}\ \exists B \in \mathcal{F}:\ B \subset A,\ \mu(B) = \frac{1}{2}\mu(A)$. Тогда для нахождения множества $D_x:\ x \in D_x \subset D,\ \mu(D_x) \le \epsilon$ необходимо и достаточно $\lceil \log_2 \frac{\mu(D)}{\epsilon} \rceil$ обращений к оракулу в худшем случае.

Оптимальность

Док-во.

- Предположим, что мы обратились к оракулу m раз и спрашивали про множества $D_1, \ldots, D_m$.

Оптимальность

Док-во.

- Предположим, что мы обратились к оракулу m раз и спрашивали про множества $D_1, \ldots, D_m$.
- Обозначим

$$B_j = \bigcap_{i=1}^{m} \begin{cases} D_i & \text{если } i\text{-ый бит } j \text{ равен } 1 \\ D \setminus D_i & \text{если } i\text{-ый бит } j \text{ равен } 0. \end{cases}$$

Оптимальность

Док-во.

- Предположим, что мы обратились к оракулу m раз и спрашивали про множества $D_1, \ldots, D_m$.
- Обозначим

$$B_j = \bigcap_{i=1}^{m} \begin{cases} D_i & \text{если } i\text{-ый бит } j \text{ равен } 1 \\ D \setminus D_i & \text{если } i\text{-ый бит } j \text{ равен } 0. \end{cases}$$

- Очевидно, что $\{B_i\}_{i=0}^{2^m}$ – разбиение D и при этом мы точно знаем, к какому из этих множеств принадлежит x. В худшем случае оно принадлежит подмножеству наибольшей меры.

Оптимальность

Док-во.

- Предположим, что мы обратились к оракулу m раз и спрашивали про множества $D_1, \ldots, D_m$.
- Обозначим

$$B_j = \bigcap_{i=1}^{m} \begin{cases} D_i & \text{если } i\text{-ый бит } j \text{ равен } 1 \\ D \setminus D_i & \text{если } i\text{-ый бит } j \text{ равен } 0. \end{cases}$$

- Очевидно, что $\{B_i\}_{i=0}^{2^m}$ – разбиение D и при этом мы точно знаем, к какому из этих множеств принадлежит x. В худшем случае оно принадлежит подмножеству наибольшей меры.
- Очевидно, что $\max_i \mu(B_i) \geq \frac{\mu(D)}{2^m}$, а значит $\epsilon \geq max_i \mu(B_i) \geq \frac{\mu(D)}{2^m}$ следовательно

$$m \geq \log_2 \frac{\mu(D)}{\epsilon}.$$

Оптимальность

Док-во.

- Предположим, что мы обратились к оракулу m раз и спрашивали про множества $D_1, \ldots, D_m$.
- Обозначим
$$B_j = \bigcap_{i=1}^{m} \begin{cases} D_i & \text{если } i\text{-ый бит } j \text{ равен } 1 \\ D \setminus D_i & \text{если } i\text{-ый бит } j \text{ равен } 0. \end{cases}$$
- Очевидно, что $\{B_i\}_{i=0}^{2^m}$ – разбиение D и при этом мы точно знаем, к какому из этих множеств принадлежит x. В худшем случае оно принадлежит подмножеству наибольшей меры.
- Очевидно, что $\max_i \mu(B_i) \geq \frac{\mu(D)}{2^m}$, а значит $\epsilon \geq max_i \mu(B_i) \geq \frac{\mu(D)}{2^m}$ следовательно
$$m \geq \log_2 \frac{\mu(D)}{\epsilon}.$$
- Обозначим через $h(A)$ такое подмножество A, что $\mu(h(A)) = \frac{1}{2}\mu(A)$. Для того, чтобы за m итераций получить множество размера $\frac{\mu(D)}{2^m}$ достаточно брать $D_i = h(D_{i-1})$, $D_0 = D$. ■

Оптимальность

Замечания:

- Стоит отметить, что если множество имеет малый диаметр, то оно имеет и малую меру. В обратную сторону это неверно: диаметр сколь угодного малого по мере множества может быть сколь угодно большим.
- Теорем выполняется и для дискретного случая. Отличие состоит в том, что в дискретном случае не получится делить множество ровно пополам.
- Метод бисекции полностью повторяет описанную процедуру для случая $D = [a, b]$, μ – мера Лебега. При этом для отрезков очевидно, что мера Лебега совпадает с диаметром.