Задача восстановления зависимости $y\colon X \to Y$ по точкам *обучающей выборки* (x_i, y_i), $y_i = y(x_i)$, $i = 1, \ldots, \ell$.

Определение

Линейной композицией базовых алгоритмов $a_t(x) = C\bigl(b_t(x)\bigr)$, $t = 1, \ldots, T$, называется суперпозиция функций

$$a(x) = C\left(\sum_{t=1}^{T} \alpha_t b_t(x)\right),$$

где $C\colon \mathbb{R} \to Y$ — *решающее правило*, $\alpha_t \geqslant 0$.

- **Пример 1:** классификация на 2 класса, $Y = \{-1, +1\}$;
  $C(b) = \operatorname{sign}(b)$, $a(x) = \operatorname{sign}\bigl(b(x)\bigr)$,
  $b\colon X \to \mathbb{R}$ — дискриминантная функция.
- **Пример 2:** регрессия, $Y = \mathbb{R}$,
  $C(b) = b$, $a(x) = b(x)$, решающее правило не используется.

Линейная композиция базовых алгоритмов:

$$b(x) = \sum_{t=1}^{T} \alpha_t b_t(x), \quad x \in X, \quad \alpha_t \in \mathbb{R}_+.$$

Функционал качества с произвольной функцией потерь $\mathscr{L}(b, y)$:

$$Q(\alpha, b) = \sum_{i=1}^{\ell} \mathscr{L}\Biggl(\underbrace{\underbrace{\sum_{t=1}^{T-1} \alpha_t b_t(x_i)}_{u_{T-1,i}} + \alpha b(x_i)}_{u_{T,i}}, y_i \Biggr) \to \min_{\alpha, b}.$$

Ищем вектор $u = \bigl(b(x_i)\bigr)_{i=1}^{\ell}$ из R^{ℓ}, минимизирующий $Q(\alpha, b)$.
$u_{T-1} = \bigl(u_{T-1,i}\bigr)_{i=1}^{\ell}$ — текущее приближение вектора u
$u_T = \bigl(u_{T,i}\bigr)_{i=1}^{\ell}$ — следующее приближение вектора u

---

Friedman G. Greedy Function Approximation: A Gradient Boosting Machine. 1999.

- Градиентный метод минимизации $Q(u) \to \min$, $u \in \mathbb{R}^\ell$:

$$u_0 := \text{начальное приближение;}$$
$$u_{T,i} := u_{T-1,i} - \alpha g_i, \quad i = 1, \dots, \ell;$$

$g_i = \mathscr{L}'(u_{T-1,i}, y_i)$ — компоненты вектора градиента,
α — градиентный шаг.

- Добавление базового алгоритма b_T:

$$u_{T,i} := u_{T-1,i} + \alpha b_T(x_i), \quad i = 1, \dots, \ell$$

Будем искать такой базовый алгоритм b_T, чтобы
вектор $\bigl(b_T(x_i)\bigr)_{i=1}^\ell$ приближал вектор антиградиента $(-g_i)_{i=1}^\ell$:

$$b_T := \arg\max_b \sum_{i=1}^{\ell} \bigl(b(x_i) + g_i\bigr)^2$$

Вход: обучающая выборка X^ℓ; параметр T;
Выход: базовые алгоритмы и их веса $\alpha_t b_t$, $t = 1, \dots, T$;

1: инициализация: $u_i := 0$, $i = 1, \dots, \ell$;
2: **для всех** $t = 1, \dots, T$
3: найти базовый алгоритм, приближающий градиент:
$$b_t := \arg\min_b \sum_{i=1}^{\ell} \bigl(b(x_i) + \mathscr{L}'(u_i, y_i)\bigr)^2;$$
4: решить задачу одномерной минимизации:
$$\alpha_t := \arg\min_{\alpha>0} \sum_{i=1}^{\ell} \mathscr{L}\bigl(u_i + \alpha b_t(x_i), y_i\bigr);$$
5: обновить значения композиции на объектах выборки:
$$u_i := u_i + \alpha_t b_t(x_i); \quad i = 1, \dots, \ell;$$