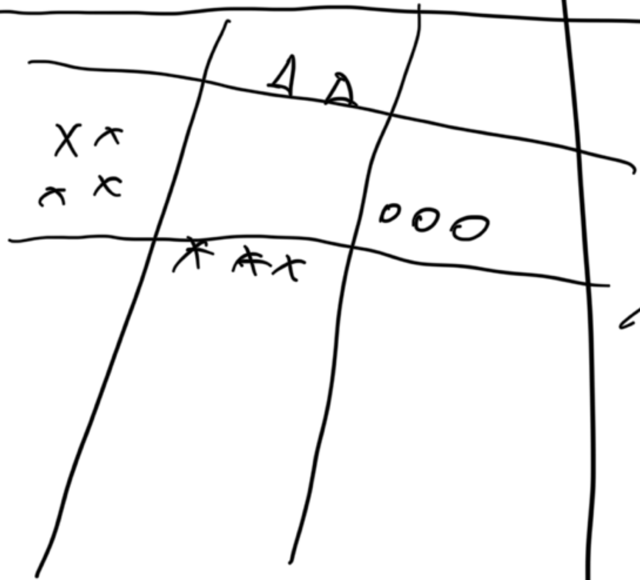
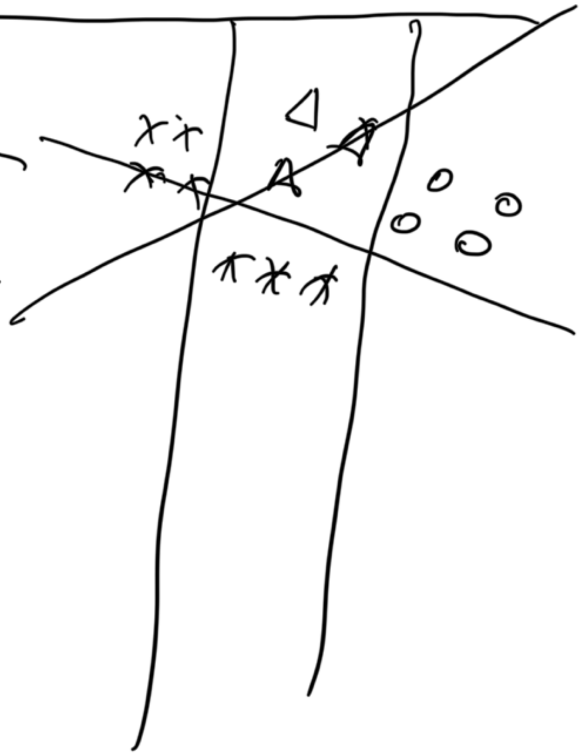
One-vs-All	All-vs-All One-vs-One
K	C_K^2





Accuracy

$$\frac{1}{\ell}\sum_{i=1}^{\ell}[y_i = a(x_i)]$$

Precision

$$\frac{TP}{TP + FP}$$

K задач бин. клас.

1) Микро усредн.

$$\overline{TP} = \frac{1}{K}\sum_{k=1}^{K} TP_k$$

$$y_i^k = [y_i = k]$$

TP_k
FP_k
FN_k
TN_k.

$$\frac{\overline{TP}}{\overline{TP} + \overline{FP}}$$

2) Макро усредн.

$$\text{Precision}_k = \frac{TP_k}{TP_k + FP_k}$$

$$\text{Precision} = \frac{1}{K}\sum_{k=1}^{K} \text{precision}_k$$

Макро

y	1	1	1	1	1	1	2	2	2	2	2	2	3	3
$a(x)$	1	1	1	1	1	3	3	2	2	2	2	2	1	1

$$P_1 = \frac{5}{7} \qquad P_2 = \frac{5}{5} = \frac{TP}{TP+FP} = 1$$

$$P_3 = \frac{0}{2} = 0 \qquad \frac{\frac{5}{7}+1+0}{3} \approx \frac{12}{21}$$

$\frac{4}{7}$

Микро

$$\overline{TP} = \frac{5+5+0}{3} = \frac{10}{3}$$

$$\overline{FP} = \frac{2+0+2}{3} = \frac{4}{3}$$

$$\frac{\frac{10}{3}}{\frac{10}{3} + \frac{4}{3}} = \frac{10 \cdot 3}{3 \cdot 14} = \frac{10}{14} = \frac{5}{7}$$

---

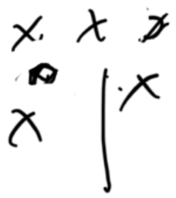
1 0 pr_1

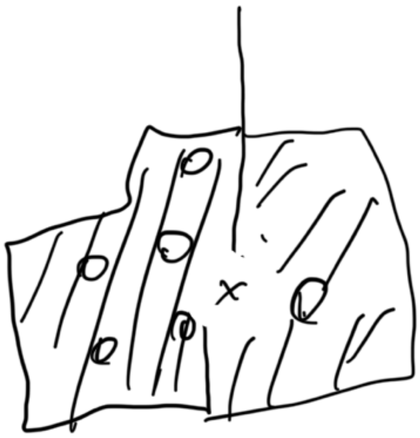
0 1. pr_2.

y	1	1	1	1	1	1	0	0	0	0	0
$a(x)$	0	1	1	1	1	0	1	1	0	0	0

$\frac{4}{6}$ $\frac{3}{5}$

композиции моделей





$$X = (x_i, y_i)_{i=1}^{\ell}$$

бутстрап

$$X \to X_1, X_2, \dots X_n \sim 2/3$$

X_i – подвыборка

$$b_1(x),\ b_2(x) \dots\ b_n(x).$$

$y(x)$ – ответы

$p(x)$ – плотность распред. объектов на $\mathbb{X}$

$\varepsilon_j(x) = b_j(x) - y(x)$ – ошибка модели

$$E_1 = \mathbb{E}\,\varepsilon_j^2(x) = \mathbb{E}\left(b_j(x) - y(x)\right)$$

1) $\mathbb{E}\,\varepsilon_j(x) = 0$

2) $\mathbb{E}\,\varepsilon_j(x)\,\varepsilon_i(x) = 0 \quad i \neq j$

некоррелиров. ошибок.

$$\underline{a(x)} = \frac{1}{N}\sum_{j=1}^{N} b_j(x)$$

$$\mathbb{E}\left(a(x) - y(x)\right)^2 = \mathbb{E}\left(\frac{1}{N}\sum_{j=1}^{N} b_j(x) -\right.$$

$$\left. - y(x)\right)^2 = \mathbb{E}\left(\frac{1}{N}\sum_{j=1}^{N} b_j(x) - \right.$$

$$\left.\frac{1}{N}\sum_{j=1}^{N} y(x)\right)^2 = \frac{1}{N^2}\mathbb{E}\left(\sum_{j=1}^{N}\varepsilon_j(x)\right)^2 =$$

$$= \frac{1}{N^2}\mathbb{E}\left(\sum_{j=1}^{N}\varepsilon_j^2(x) + \underbrace{\sum_{i\neq j}\varepsilon_i(x)\,\varepsilon_j(x)}_{}\right)$$

$$= \frac{1}{N^2}\mathbb{E}\left(\sum_{j=1}^{N}\varepsilon_j^2(x)\right) = \frac{1}{N^2}\sum_{j=1}^{N}\mathbb{E}\,\varepsilon_j^2(x)$$

$$= \frac{1}{N^2} \cdot N \cdot E_1 = \underline{\underline{\frac{1}{N} E_1}}$$

Bais - Variance - decomposition

BVD

$$X = (x_i, y_i)_{i=1}^{\ell} \quad y_i \in \mathbb{R}$$

$$\mathbb{Y} = \mathbb{R}$$

$\underline{p(x,y)}$ - плотность распред. на $\mathbb{X} \times \mathbb{Y}$

$$L(y, a) = (y - a)^2$$

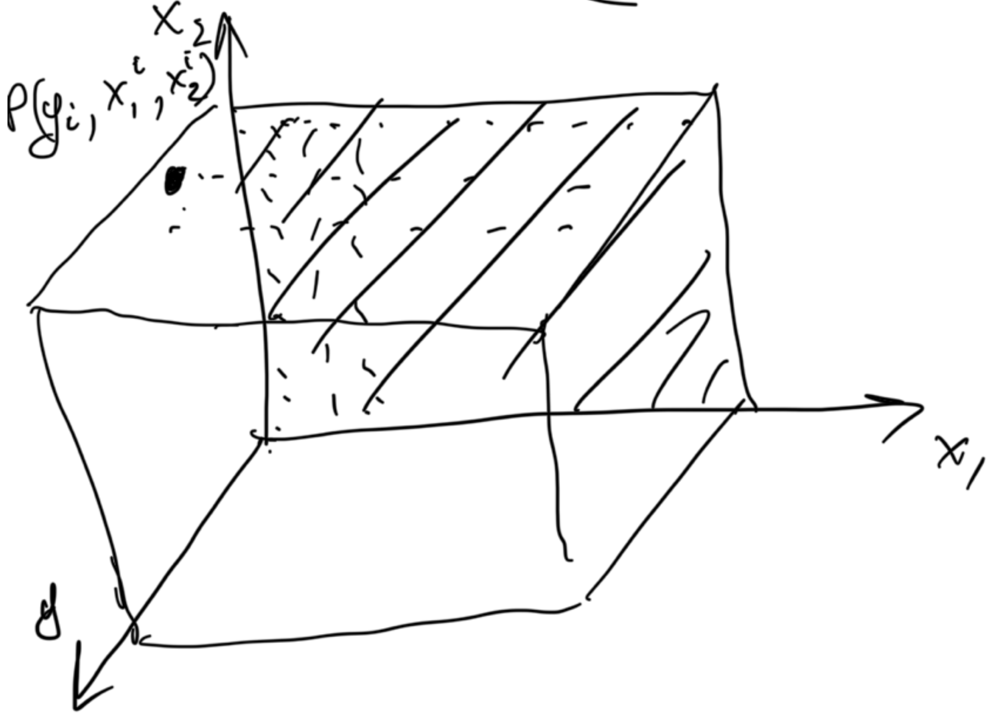
среднеквадратичный риск.

$$\boxed{\underline{R(a)}} = \mathbb{E}_{x,y}\left((y - a(x))^2\right) =$$

$$= \int \int p(x, y) \, (y - a(x))^2 \, dx \, dy.$$

$$\int_X \int_Y \ldots$$

$$\left(a(x_1^i, x_2^i) - y_i\right)^2 \cdot p$$



$$\boxed{Q} = \frac{1}{l}\sum_{j=1}^{l}\left(a(x_j) - y_j\right)^2 \qquad a(x)$$

X

$$a_*(x) = \mathbb{E}[y|x] = \int_Y y \cdot p(y|x)\,dy$$

$$\arg\min_a R(a)$$

$\mu: (X \times Y)^l \Rightarrow A$ → семейство

$\mu: (X \to y)$ — алгоритм

пр-во обуч. выборок.

$\mu(X) = a$

$$L(\mu) = \mathbb{E}_X \mathbb{E}_{x,y} \big(y - \underbrace{\mu(X)(x)}_{a}\big)^2$$

прогноз $a(x)$

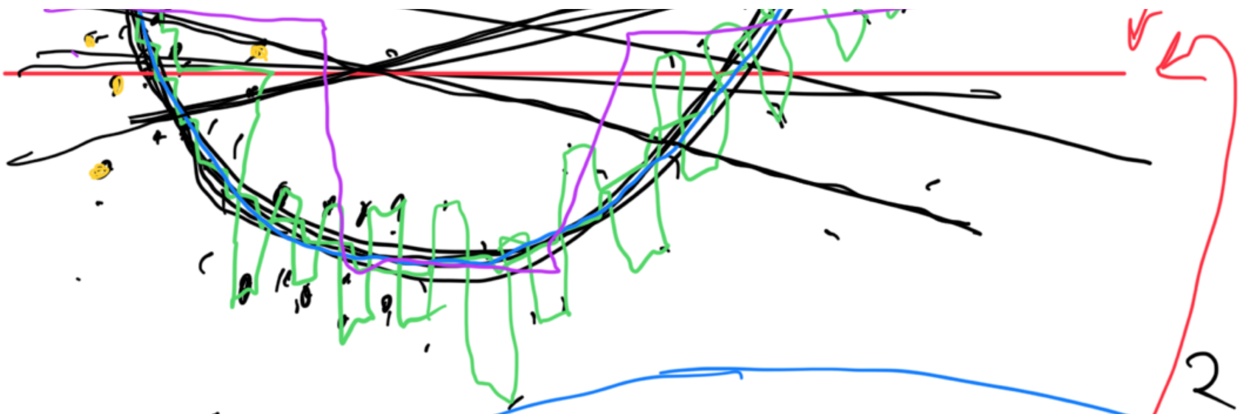
$X \quad a$

$$L(\mu) = \mathbb{E}_{x,y} \big(y - \mathbb{E}(y|x)\big)^2 +$$

шум

$$+\, \mathbb{E}_x \big(\mathbb{E}_X \mu(X) - \mathbb{E}(y|x)\big)^2 \text{ — bias}$$

$p(x,y)$



$$+ \mathbb{E}_x \mathbb{E}_X \left(\mu(X) - \underbrace{\mathbb{E}_X \mu(X)}_{1} \right)^2$$

красная прямая

Bagging

$a(X)$ – метод обучения

$\tilde{X}$ – случайная бутстрапированная подвыборка

$$b_1(x) = \mu(\tilde{X}_1)$$

$$b_2(x) = \mu(\tilde{X}_2)$$

$$\vdots$$

$$b_n(x) = \mu(\tilde{X}_n).$$

$$a_N(x) = \frac{1}{N} \sum_{n=1}^{N} b_n(x) = \frac{1}{N} \sum_{n=1}^{N} \mu(\tilde{X}_n).$$

Оказывается!

$$\text{bias}(a_N(x)) = \text{bias}(b_n(x))$$

$$\text{var}(a_N(x)) = \frac{1}{N}\text{var}(b_n(x)) + \frac{N(N-1)}{N^2}\text{cov}(b_n(x), b_m(x))$$

---

1) Глубокие решающие деревья

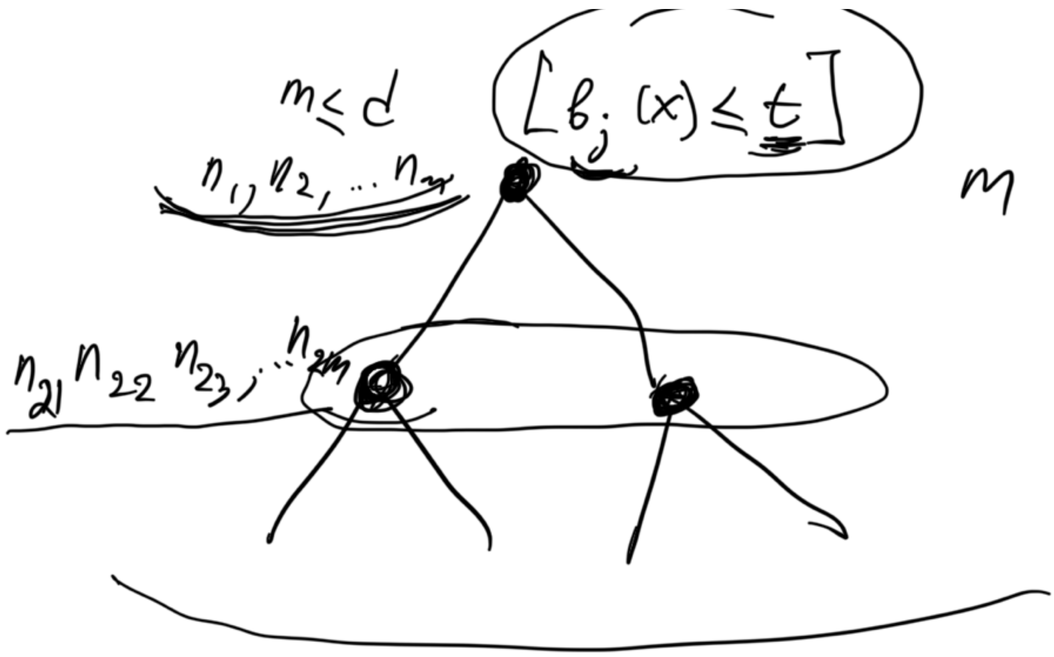
Random Forest (RF)

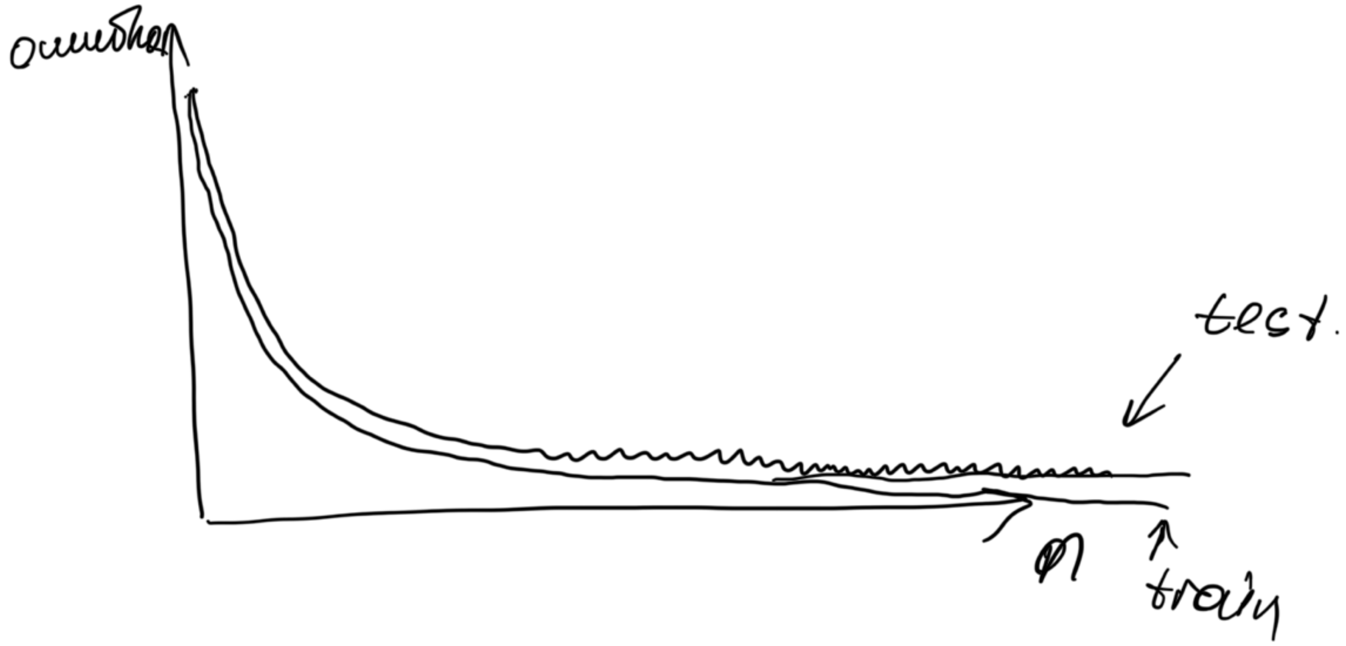
1) бэггинг над решающими деревьями
2) деревья строим глубоким
3) Метод случайных подпространств

$$\tilde{X}_1, \tilde{X}_2, \dots \tilde{X}_n$$

$$\uparrow \qquad\qquad \uparrow$$

$$n_{11}, n_{12}, \dots n_{1m} \qquad n_{21}, n_{22}, \dots n_{2m}$$

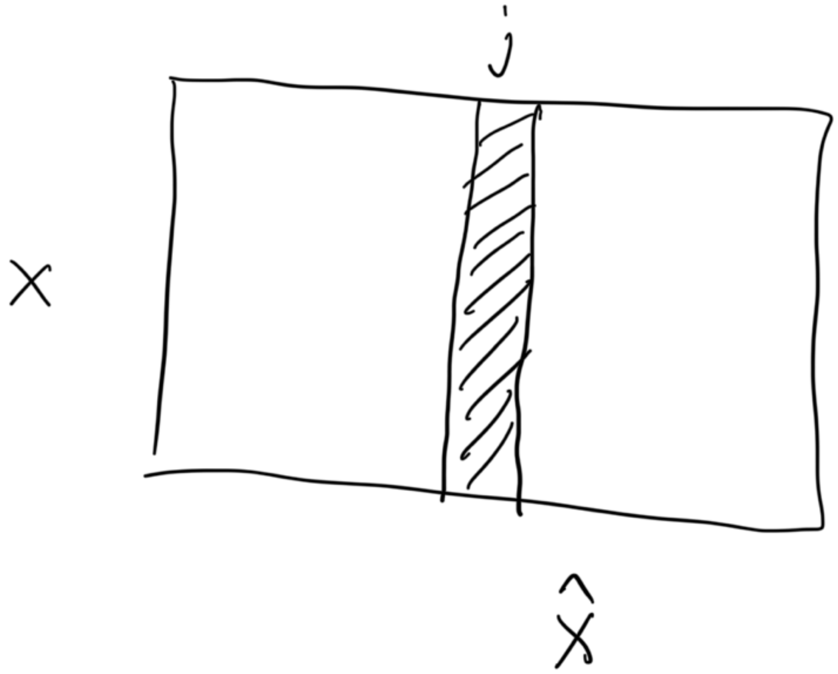




OOB

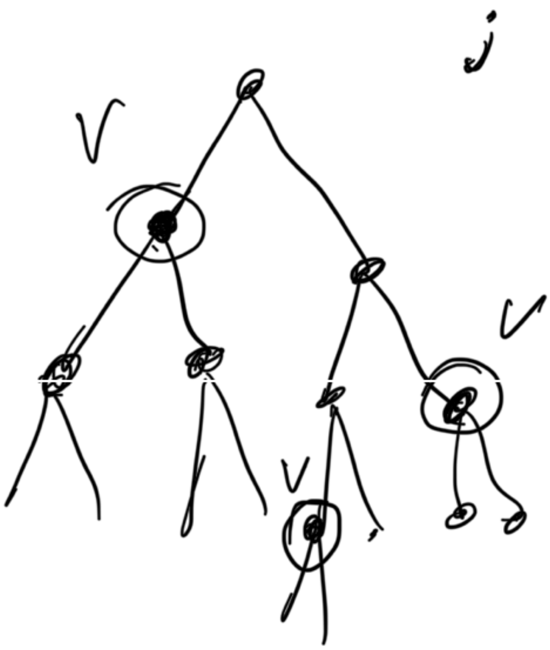
$$OOB = \sum_{i=1}^{\ell} L\left(y_i, \frac{1}{\sum_{n=1}^{N} [x_i \notin X_n]} \sum_{n=1}^{N} [x_i \notin X_n] b_n(x_i)\right)$$

$\mu(X)$
$\mu(\hat{X})$



1)

2)



$$\underline{a_N(x)} = \sum_{n=1} \underline{b_n(x)}$$

Регрессия с MSE

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\left(a(x_i)-y_i\right)^2 \to \min_a$$

$$b_1(x) = \arg\min_{b\in A} \frac{1}{\ell}\sum_{i=1}^{\ell}\left(b(x_i)-y_i\right)^2$$

$$s_i^{(1)} = \underline{y_i} - \underline{b_1(x_i)}$$

$$b_2(x) = \arg\min_{b\in A} \frac{1}{\ell}\sum_{i=1}^{\ell}\left(b(x_i)-s_i^{(1)}\right)^2$$

$$\underline{s_i^{(2)}} = y_i - b_1(x_i) - b_2(x_i) = s_i^{(1)} - b_2(x_i)$$

$$b_3(x) = \arg\min_{b\in A} \frac{1}{\ell}\sum_{i=1}^{\ell}\left(b(x_i)-s_i^{(2)}\right)^2$$

$\vdots$

$$\frac{1}{\ell}\sum^{\ell}\left(b(x_i) - s_i^{(N)}\right)^2$$

$$b_N(x) = \underset{b \in \mathcal{A}}{\arg\min} \sum_{i=1}^{\ell} \left(b(x_i) - s_i\right)$$

$$a_N(x_i) = \sum_{n=1}^{N} b_n(x_i)$$

$L(y, z)$ - дифф-ма.

$$a_N(x) = \sum_{n=0}^{N} b_n(x)$$

- $b_0(x) = 0$
- $b_0(x) = \underset{y \in \mathbb{Y}}{\arg\max} \sum_{i=1}^{\ell} [y_i = y]$

$b_N(x)$ - ?

$$\sum_{i=1}^{\ell} L\left(y_i, a_{N-1}(x_i) + b_N(x_i)\right) \to \min_{b_N}$$

$$\sum_{i=1}^{\ell} L\left(y_i, a_{N-1}(x_i) + s_i\right) \to \min_{s_1, \dots, s_\ell}$$

$$s_i = y_i - a_{N-1}(x_i)$$

$$s_i = -\frac{\partial L}{\partial z}\Big|_{z = a_{N-1}(x_i)}$$

$$L\left(y_i,\ a_{N-1}(x_i) - \frac{\partial L}{\partial z}\Big|_{z=a_{N-1}(x_i)}\right)$$

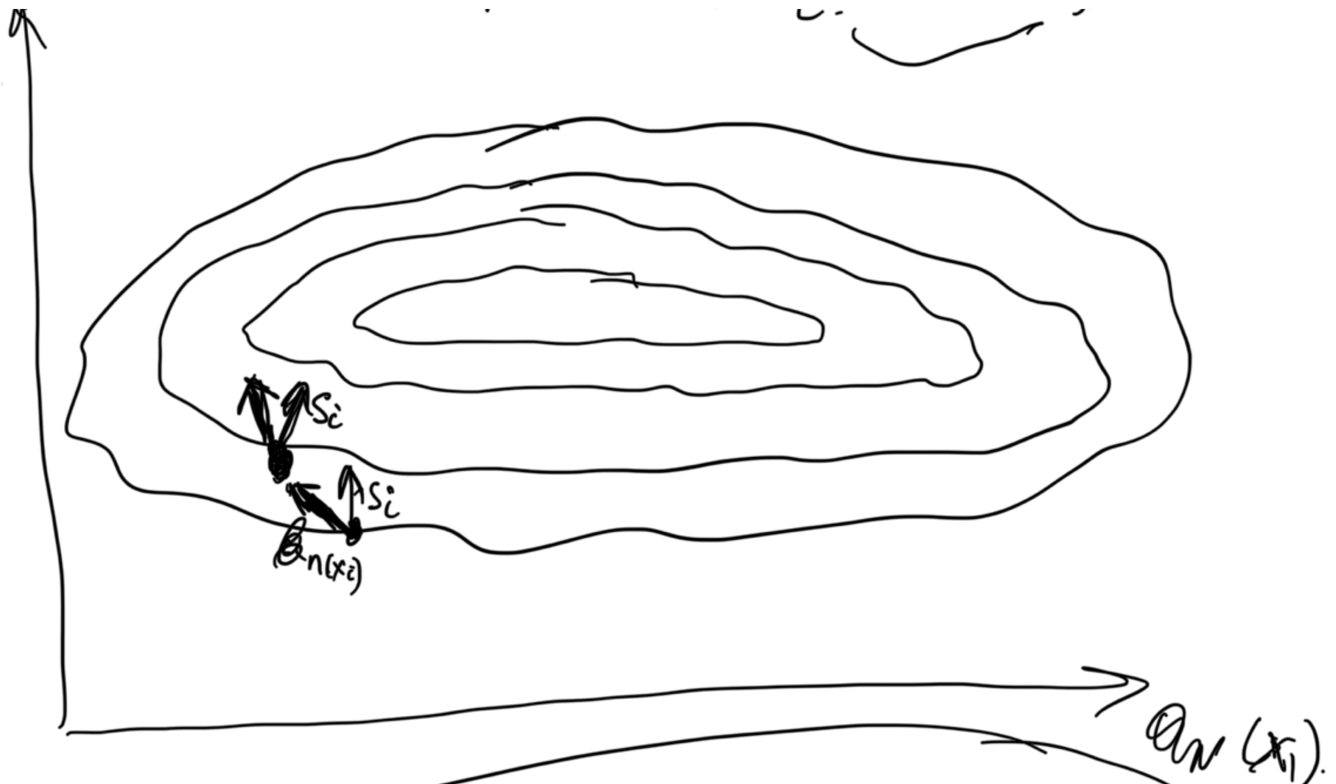
$$b_N \sim s_i \qquad \approx -\frac{\partial L}{\partial z}\Big|_{z=a_{N-1}(x_i)}$$

$$a_N = a_{N-1}(x_i) + b_N(x_i)$$

$$a_{N+1} = a_{N-1}(x_i) + \underbrace{b_N(x_i)}_{\approx -\frac{\partial L}{\partial z}\big|_{z=a_{N-1}(x_i)}} + \underbrace{b_{N+1}(x_i)}_{\approx -\frac{\partial L}{\partial z}\big|_{z=a_N(x_i)}}$$

$a_N(x_2)$

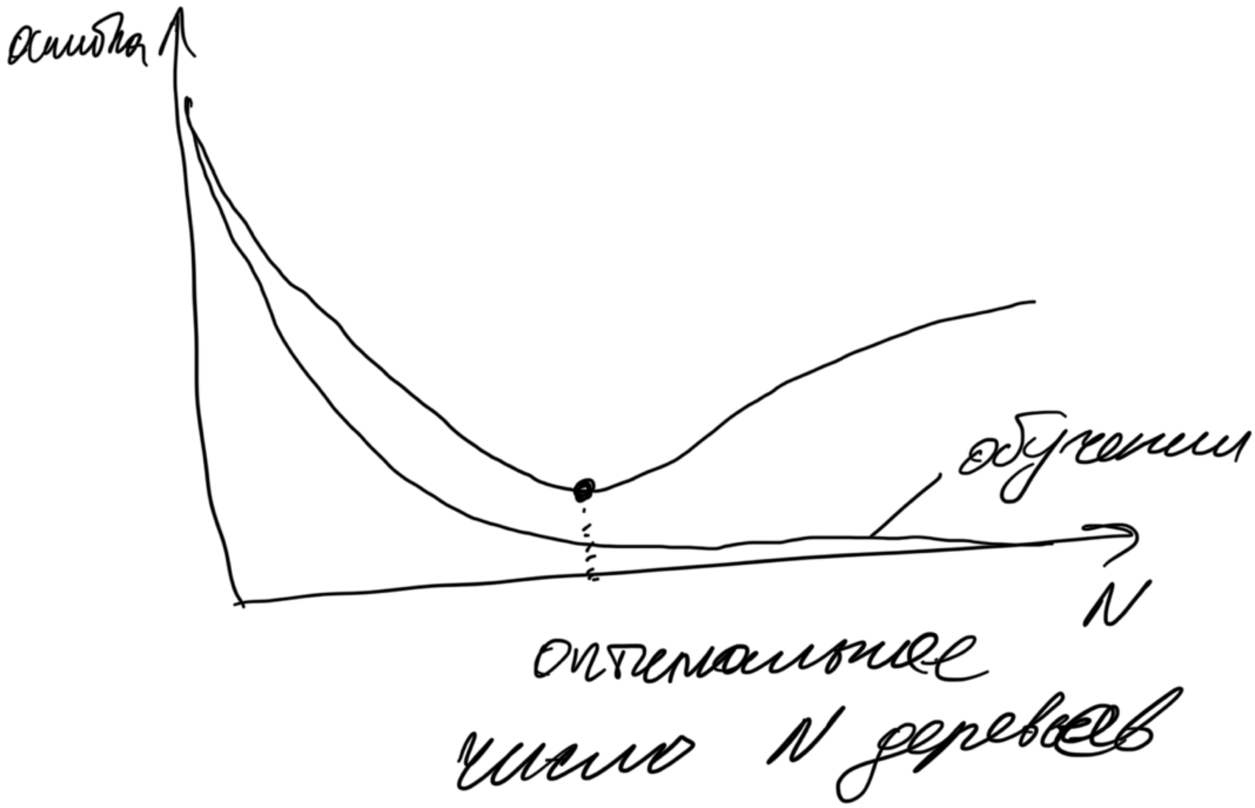
$$w_N = w_{N-1} - \eta \nabla f(w_{N-1})$$



$$b_N(x) = \underset{b}{\operatorname{argmin}} \sum_{i=1}^{\ell} \left(b(x_i) - s_i\right)^2$$

∂L

Бустинг уменьшает bai's



Регуляризация

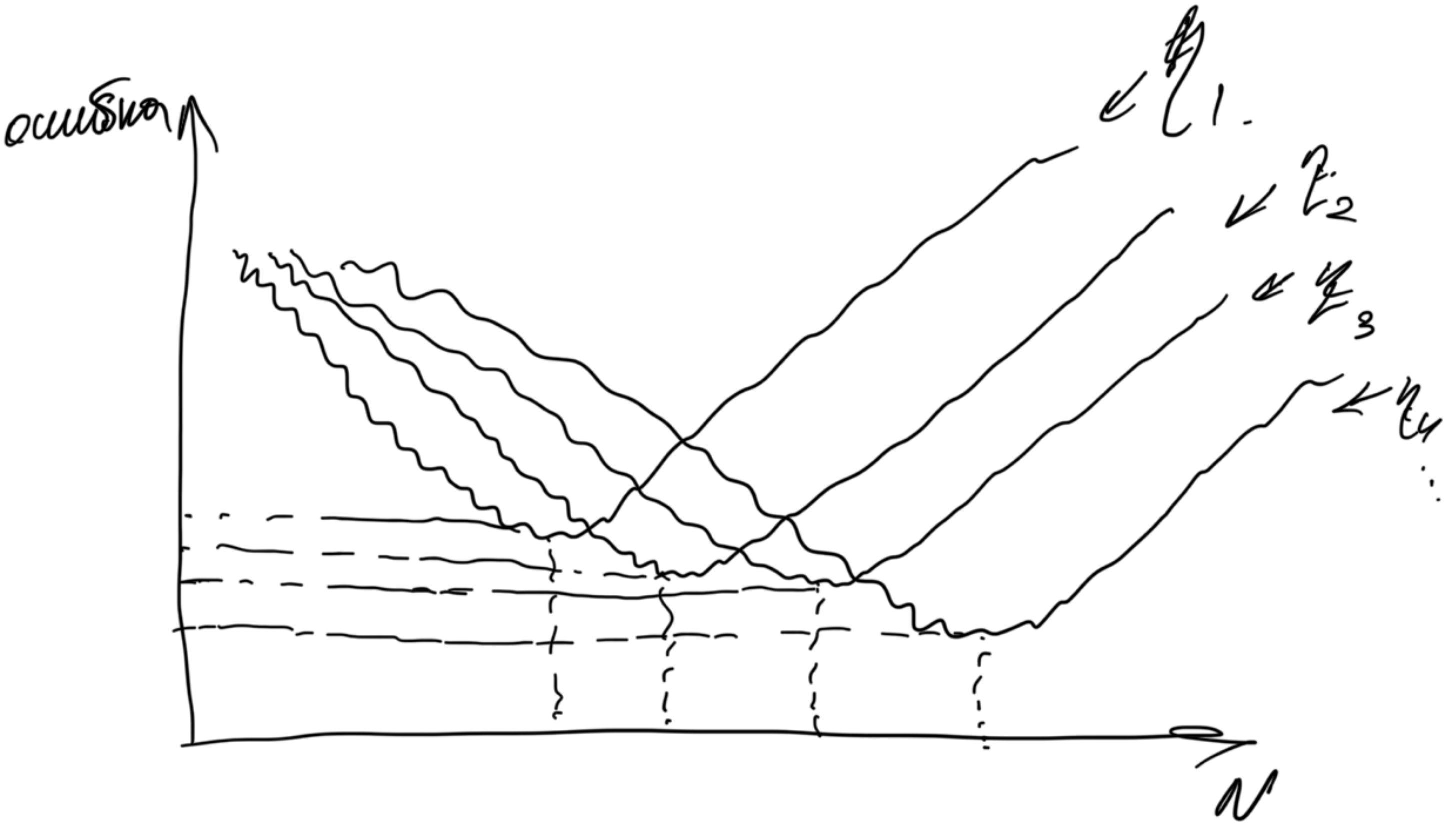
$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

$$a_N(x) = \sum_{n=1}^{N} \eta \, b_n(x)$$

$$a_N(x) = a_{N-1}(x) + \underline{\eta} \, \underline{b_N(x)}$$

$\eta \in (0, 1]$

S_i



$$\eta_1 > \eta_2 > \eta_3 > \eta_4$$