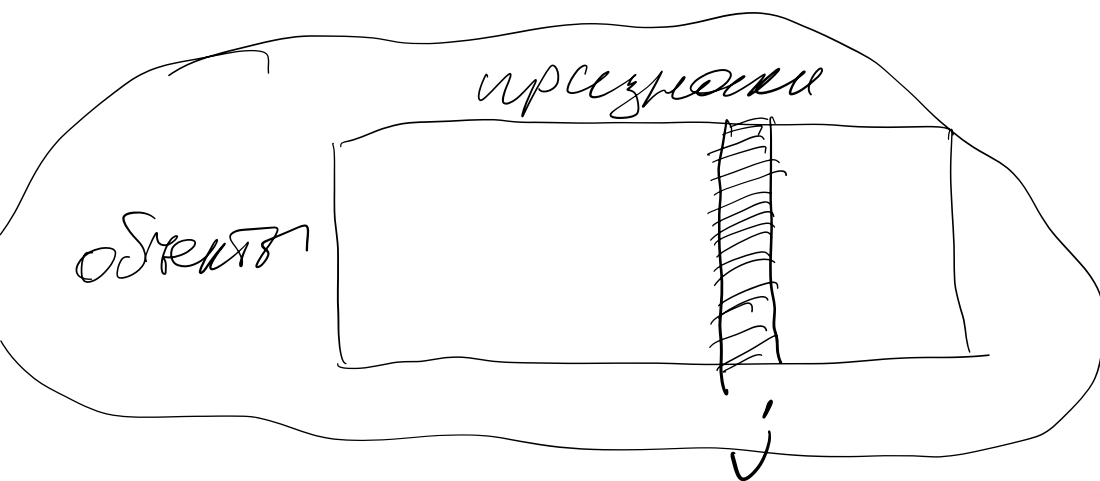
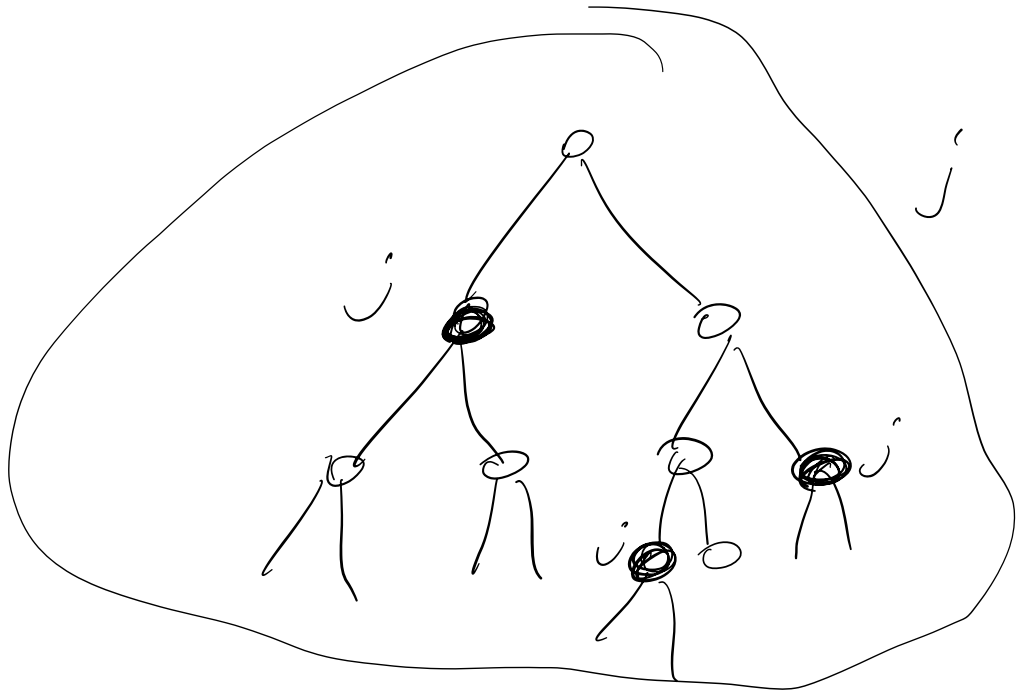
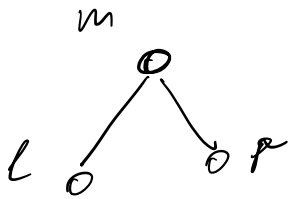


$$\underline{Q} = H(R_m) - \frac{|R_r|}{|R_m|} H(R_r) - \frac{|R_e|}{|R_m|} H(R_e)$$



model - feature - importance.

LIME
Local Interpretable Model-agnostic
Explanations

$$\underline{f: X \rightarrow Y}$$

$$\underline{x_0 \in X} - \text{объект}$$

$$f(x_0 + \Delta x_0)$$

M - измерений Δ_i

$\{\Delta_i\}_{i=1}^M$ - уменьшение x_0

$z \in \{0, 1\}_{i=1}^M$ - вектор уменьшений x_0 .

$h: \{0, 1\}^M \rightarrow X$ - функция

из z в $x_0 + \Delta x$

$g: \{0,1\}^n \rightarrow \mathcal{Y}$ объект с метрикой
метрика

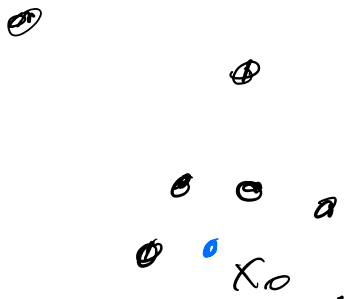
$$\sum_{i=1}^N w^{(i)} \left(g(z^{(i)}) - f(h(z^{(i)})) \right)^2 \rightarrow \min_{\mathcal{F}}$$

$$\underbrace{x_0} \quad \underbrace{x_0 + \Delta x}$$

$$\sum_{i=1}^N w^{(i)} \left(g(z^{(i)}) - f(h(z^{(i)})) \right)^2 + \mathcal{R}(g) \rightarrow \min_{\mathcal{F}}$$

штраф за сложность

кол-во ненулевых весов
в линейной модели



$x_0^{(i)}$

e

Sharp

S - коалиция (мн-во игроков)

i - игрок

$\Delta(i, S)$ - прирост эффективности
при добавлении игрока i в S

$$\Delta(i, S) = v(S \cup i) - v(S)$$

N - игроки

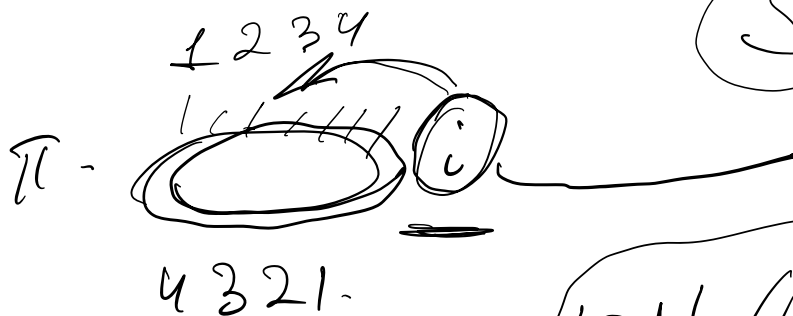
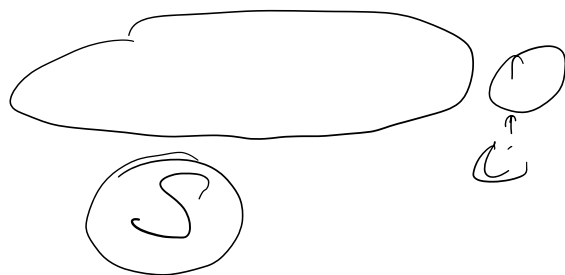
Π - мн-во упорядочиваний

$$\underline{g(i)} = \frac{1}{N!} \sum_{\pi \in \Pi} \Delta(i, \text{players before } i \text{ in } \pi)$$

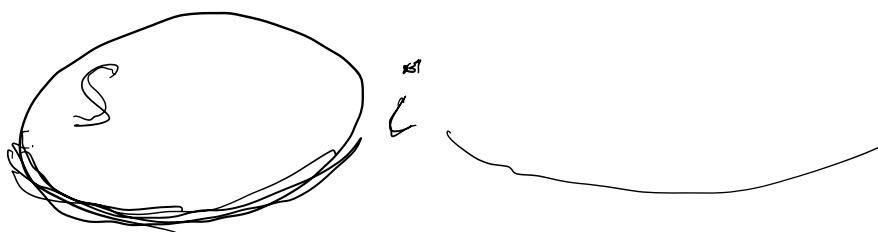
↓
мн-во игроков
идущих перед

состояния π и π'
 и $i \in \pi$

$S + i$



$$g(i) = \sum_{S \subseteq \{1, 2, \dots, N\} \setminus i} \frac{|S|! (N - |S| - 1)!}{N!} \Delta(i, S)$$



N