$$p(D|\bar\theta), \quad p(\bar\theta)$$

$$\max \Leftarrow p(\bar\theta|D) \propto p(\bar\theta)\, p(D|\bar\theta)$$

$\bar\theta_{MAP}$

$$p(\bar x|D) = \underset{p(\bar\theta|D)}{\mathbb{E}}\left[p(\bar x|\bar\theta)\right]$$

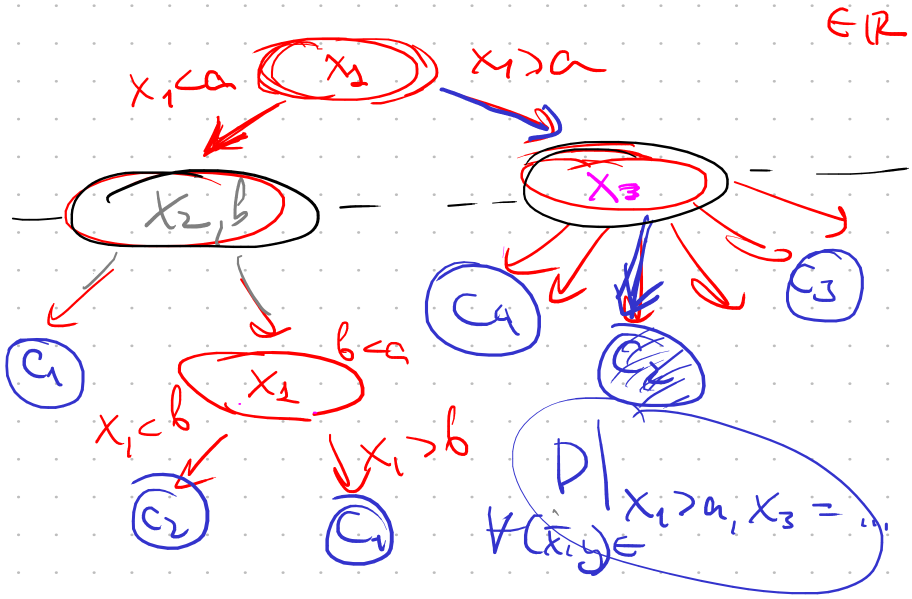
$\bar x = (\boxed{x_1}, x_2, \dots, x_d)$

$$D = \{(\bar x_n, y_n)\}_{n=1}^{N}, \quad y_n \in \{1, \dots, K\}$$

возраст $\in \{<14,\ [14,18],\ [19,25],\ [26,35],\ [36,50],\ >50\}$

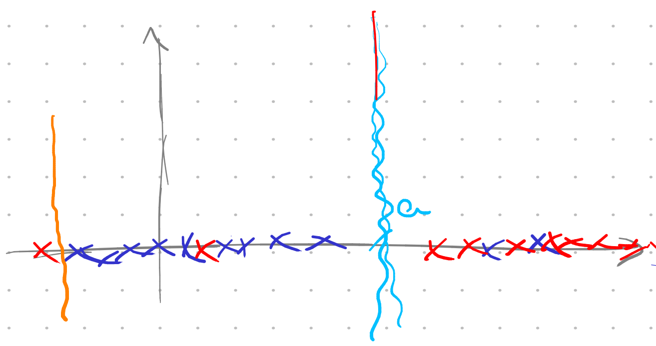
$\bar x = (x_1, x_3, \dots)$, $x_1 \in \mathbb{R}$

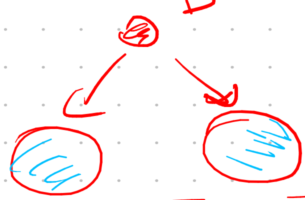
① Decision trees



Oblivious decision trees

- рекурсивно:
  - хотим ли мы расщеплять? если нет, всё
  - выбрать атрибут x_i
  - выбрать расщепление:
    - если x_i катег., то по знач.
    - если численный?





$D_1 = D|_{x_i < a}$ $\quad D_2 = D|_{x_i > a}$

$H(D_2)$

$$\mathrm{Gini} = \sum_k p_k(1 - p_k)$$

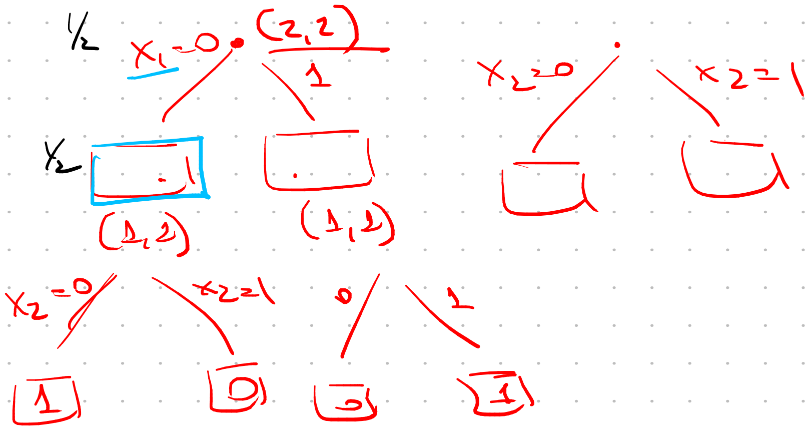
$$H(D_1) = -\sum_k p_k \log p_k$$

$$\cancel{\frac{1}{2}H(D_1) + \frac{1}{2}H(D_2)} \to \min$$

$$\text{где } p_k = \frac{\#\, n: x_n \in C_k}{|D_1|}$$

$$\mathrm{Gain}(D, x_i) = \sum_{s=1}^{S} \frac{|D_s|}{|D|} H(D_s) - H(D)$$

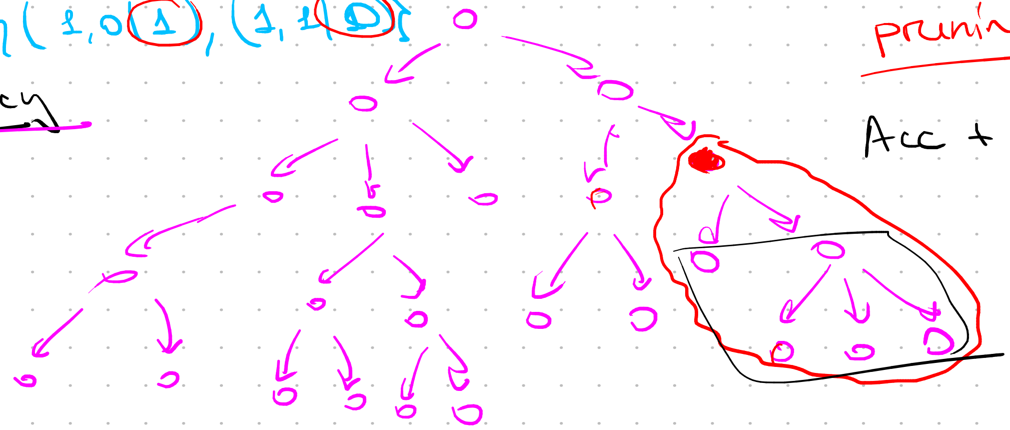
$x_1 \backslash x_2$	0	1
0	$y=0$	$y=1$
1	$y=1$	$y=0$



$D|_{x_1=0} = \{(0,0|0), (0,1|1)\}$

$D|_{x_1=1} = \{(1,0|1), (1,1|0)\}$

Accuracy



pruning

$Acc + \lambda \cdot |Tree|$

Random forest

2 Model combination

→ model selection: $M_1, M_2, \dots, M_K \rightsquigarrow k^*$

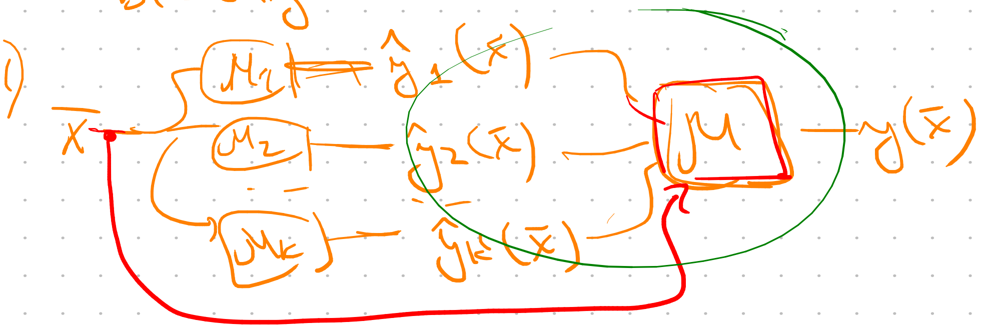
$$k^* = \arg\max_k p(M_k|D) = \arg\max_k p(D|M_k)\, p(M_k)$$

$$p(D|M_k) = \int p(D|\bar\theta_k, M_k)\, p(\bar\theta_k|M_k)\, d\bar\theta_k$$

← BIC, TIC, AIC

- blending

1)



2) $\bar x \to \boxed{M} \to \boxed{k} \to \boxed{M_k(\bar x)} \to \hat y_k(\bar x)$

— Committee

$$\mathcal{M}(\bar{x}) = \frac{1}{K}\sum_{k=1}^{K} \mu_k(\bar{x})$$

$$y_k(\bar{x}) = h(\bar{x}) + \varepsilon_k(\bar{x})$$

$$\text{Error}_k = \mathbb{E}_{\bar{x}}\left[\left(y_k(\bar{x}) - h(\bar{x})\right)^2\right] = \mathbb{E}_{\bar{x}}\left[\varepsilon_k^2(\bar{x})\right]$$

$$\text{Error}_{avg} = \frac{1}{K}\sum_k \mathbb{E}_x\left[\varepsilon_k^2(\bar{x})\right]$$

$$\text{Error}_{com} = \mathbb{E}_x\left[\left(\frac{1}{K}\sum y_k(\bar{x}) - h(\bar{x})\right)^2\right] = \mathbb{E}_x\left[\left(\frac{1}{K}\sum \varepsilon_k(\bar{x})\right)^2\right] =$$

$$= \frac{1}{K^2}\left(\sum_k \mathbb{E}_x\left[\varepsilon_k(\bar{x})^2\right] + \underbrace{\sum_{k\neq l} \mathbb{E}_{\bar{x}}\left[\varepsilon_k(\bar{x})\cdot\varepsilon_l(\bar{x})\right]}_{0}\right) = \frac{1}{K^2}\sum_k \mathbb{E}_{\bar{x}}\left[\varepsilon_k^2\right]$$

$$\frac{E_{avg}}{K} \leq E_{com} \leq E_{avg}$$

bagging – bootstrap aggregation

$$D \to \mathcal{M} \to y(\bar{x})$$

sampling with replacement

$$D \to \begin{cases} D_1 \to \mathcal{M}_1 \\ D_2 \to \mathcal{M}_2 \\ \cdots \\ D_m \to \mathcal{M}_m \end{cases} \to \frac{1}{m}\sum_{i=1}^{m} \mathcal{M}_i(\bar{x})$$

---

3 Boosting

$$F(\bar{x}) = \alpha_1 f_1(\bar{x}) + \alpha_2 f_2(\bar{x}) + \ldots + \alpha_T f_T(\bar{x})$$

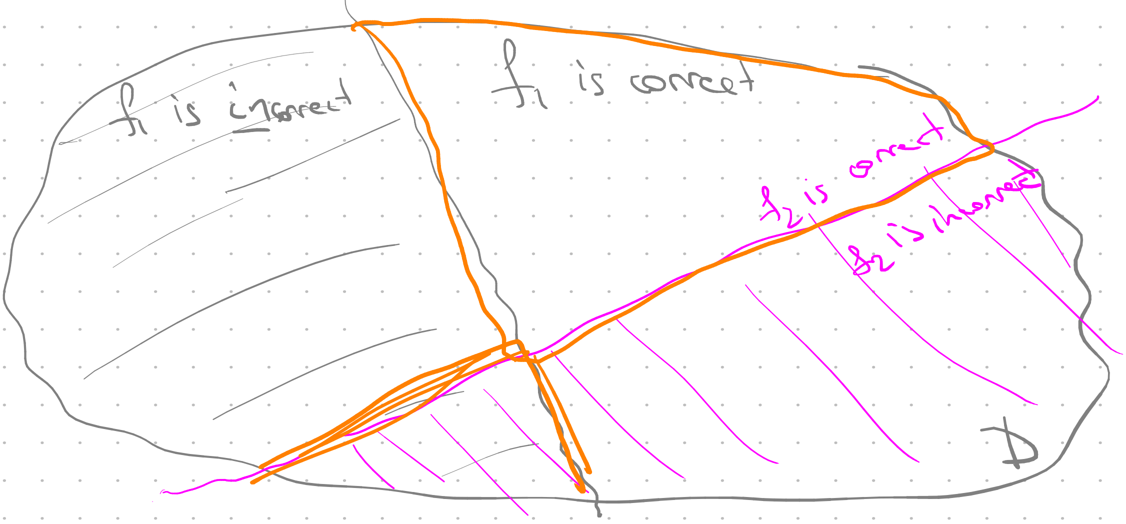
$$F_k(\bar{x}) = \underbrace{F_{k-1}(\bar{x})}_{\text{fixed}} + \alpha_k f_k(\bar{x})$$

AdaBoost

Freund, Schapire

1997

Gödel Prize



– умело обучать слаб. классиф. f_t на D, минимизируя

$$\varepsilon_t = \sum_{i=1}^{N} w_i^{(t)} [f_t(\bar{x}_i) \neq y_i]$$

AdaBoost:

$c_1 \quad c_2$

$y_i \in \{-1, 1\}$

– init $w_i^{(1)} := 1/N$

– for $t = 1..T$:

– обучаем f_t, получаем ε_t

– $\alpha_t = \frac{1}{2} \log \frac{1-\varepsilon_t}{\varepsilon_t}$

– $w_i^{(t+1)} := \frac{1}{Z_t} w_i^{(t)} \cdot e^{-\alpha_t y_i f_t(\bar{x}_i)}$,

где $Z_t = \sum_i$

– $F_{class}(\bar{x}) = \operatorname{sign}\left(\underbrace{\sum_t \alpha_t f_t(\bar{x})}_{= F(\bar{x})}\right)$

Предп.: $\varepsilon_t = \frac{1}{2} - \gamma_t$, где $\gamma_t > 0$

$$w_i^{(T+1)} = w_i^{(1)} \cdot \frac{1}{Z_1} e^{-y_i \alpha_1 f_1(\bar{x}_i)} \cdot \ldots \cdot \frac{1}{Z_T} e^{-y_i \alpha_T f_T(\bar{x}_i)}$$

$$= \frac{1}{N} \cdot \frac{e^{-y_i \cdot \sum_{t=1}^{T} \alpha_t f_t(\bar{x}_i)}}{Z_1 Z_2 \cdots Z_T} = \frac{1}{N} \cdot \frac{e^{-y_i F(\bar{x}_i)}}{Z_1 \cdots Z_T}$$

Если $\operatorname{sign}(F(x_i)) \neq y_i$, то $e^{-y_i F(\bar{x}_i)} \geq 1$

$$\Rightarrow [\operatorname{sign}(F(x_i)) \neq y_i] \leq e^{-y_i F(\bar{x}_i)}$$

$$\text{Error} = \frac{1}{N} \sum_{i=1}^{N} [\operatorname{sign}(F(x_i)) \neq y_i] \leq \frac{1}{N} \sum_i e^{-y_i F(\bar{x}_i)} = \sum_i w_i^{(T+1)} \cdot \prod_{t=1}^{T} Z_t =$$

$$= \prod_{t=1}^{T} Z_t$$

$$Z_t = \sum_i w_i^{(t)} e^{-\alpha_t y_i f_t(\bar{x}_i)} = \sum_{i: y_i = f_t(\bar{x}_i)} w_i^{(t)} e^{-\alpha_t} + \sum_{i: y_i \neq f_t(\bar{x}_i)} w_i^{(t)} \cdot e^{\alpha_t} =$$

$$= e^{-\alpha_t}(1-\varepsilon_t) + e^{\alpha_t} \cdot \varepsilon_t = e^{-\alpha_t}\left(\tfrac{1}{2}+\gamma_t\right) + e^{\alpha_t}\left(\tfrac{1}{2}-\gamma_t\right)$$

$$= \left[\alpha_t = \tfrac{1}{2}\log\frac{1-\varepsilon_t}{\varepsilon_t}\right] = \sqrt{1-4\gamma_t^2}$$

$$\text{Error} \leq \prod_{t=1}^{T}\sqrt{1-4\gamma_t^2} \leq e^{-2\sum_{t=1}^{T}\gamma_t^2}$$

(Friedman, 2000): $F(\bar{x}) = \frac{1}{2}\sum \alpha_t f_t(\bar{x})$ $\quad \frac{1}{\text{"}w_1^{(t)}}$

$$L(\bar{x}) = \sum_{i=1}^{N} e^{-y_i F_t(\bar{x}_i)} = \text{"} N\, w_i^{(t)} \prod_{s=1}^{t-1} Z_s$$

$$= \sum_{i=1}^{N} e^{-y_i\left(F_{t-1}(\bar{x}_i) + \frac{1}{2}\alpha_t f_t(\bar{x}_i)\right)} = \sum e^{-y_i F_{t-1}(\bar{x}_i)} \cdot e^{-\frac{1}{2} y_i \alpha_t f_t(\bar{x}_i)}$$

Const

$$= \sum_{i=1}^{N} \frac{w_i^{(t)}}{w_1^{(t)}} \cdot \prod_{s=1}^{t-1} Z_s \cdot e^{-\frac{1}{2} y_i \alpha_t f_t(\bar{x}_i)} = \text{"} e^{\frac{\alpha_t}{2}} - e^{-\frac{\alpha_t}{2}} + e^{-\frac{\alpha_t}{2}}$$

$$= \text{Const}\cdot\left(\sum_{i: f_t(\bar{x}_i)=y_i} w_i^{(t)} \cdot e^{-\frac{1}{2}\alpha_t} + \sum_{i: f_t(\bar{x}_i)\neq y_i} w_i^{(t)} \cdot e^{\frac{1}{2}\alpha_t}\right) =$$

ε_t $\quad$ min $\quad f_t, \alpha_t$

$$= \text{Const}\cdot\left(e^{-\frac{1}{2}\alpha_t} \cdot \text{"}\sum_i w_i^{(t)} + \left(e^{\alpha_t/2} - e^{-\alpha_t/2}\right) \cdot \boxed{\sum_{i: f_t(\bar{x}_i)\neq y_i} w_i^{(t)}}\right)$$

$$\alpha_t = \frac{1}{2}\log\frac{1-\varepsilon_t}{\varepsilon_t}$$

$$f_t: \min \sum_{i: f_t(\bar{x}_i)\neq y_i} w_i^{(t)}$$

(4) Gradient boosting

$f_1, f_2, \ldots, f_T$

$$\hat{y}_i = \sum_{t=1}^{T} f_t(\bar{x}_i) = \left(f_1(\bar{x}_i) + \ldots + f_{T-1}(\bar{x}_i)\right) + \underbrace{f_T(\bar{x}_i)}$$

$$L = \sum_{i=1}^{N} \underbrace{l(y_i, \hat{y}_i)}_{\text{" } (y_i - \hat{y}_i)^2} + \sum_{t=1}^{T} \Omega(f_t) \longrightarrow \min$$

$$L^{(T)} = \sum_{i=1}^{N} l\left(y_i, \hat{y}_i^{(T-1)} + f_T(\bar{x}_i)\right) + \Omega(f_T) \xrightarrow{f_T} \min$$

$$l(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$$

$$L^{(T)} = \sum_i \left(\left(\underline{y_i - \hat{y}_i^{(T-1)}}\right) - f_T(\bar{x}_i)\right)^2 + \Omega(f_T) =$$

$$= \text{Const} + \sum_i \left(f_T(\bar{x}_i)^2 - 2 f_T(\bar{x}_i)\left(y_i - \hat{y}_i^{(T-1)}\right)\right) + \Omega(f_T) \xrightarrow{f_T} \min$$

$$l\left(\underline{y_i, \hat{y}_i^{(T-1)}} + \overline{f_T(\bar{x}_i)}\right) \approx l\left(y_i, \hat{y}_i^{(T-1)}\right) +$$

$$+ \underbrace{\boxed{\left.\frac{\partial l}{\partial \hat{y}}\right|_{\hat{y}^{(T-1)}}}}_{= g_i} \cdot f_T(\bar{x}_i) + \frac{1}{2} \underbrace{\boxed{\left.\frac{\partial^2 l}{\partial \hat{y}^2}\right|_{\hat{y}^{(T-1)}}}}_{= h_i} \cdot f_T^2(\bar{x}_i)$$

$$L^{(T)} \approx \sum_{i=1}^{N} \left(g_i \cdot f_T(\bar{x}_i) + \frac{h_i}{2} f_T^2(\bar{x}_i)\right) + \Omega(f_T) \xrightarrow{f_T} \min$$

(5) XGBoost $\qquad f_t(\bar{x})$ — decision tree

$\bar{w} \in \mathbb{R}^{M}$

$q : \mathbb{R}^d \to \{1, \dots, M\}$

$$f_T(\bar{x}) = w_{q(\bar{x})}$$

$$\Omega(f_t) = \gamma \cdot M + \frac{\lambda}{2} \cdot \sum_{j=1}^{M} w_j^2$$

$$L^{(T)} \approx \sum_{i=1}^{N} \left(g_i \cdot \underline{w_{q(\bar{x}_i)}} + \frac{h_i}{2} w^2_{q(\bar{x}_i)} \right) + \gamma M + \frac{\lambda}{2} \underbrace{\sum^{M} w_j^2}_{1} =$$

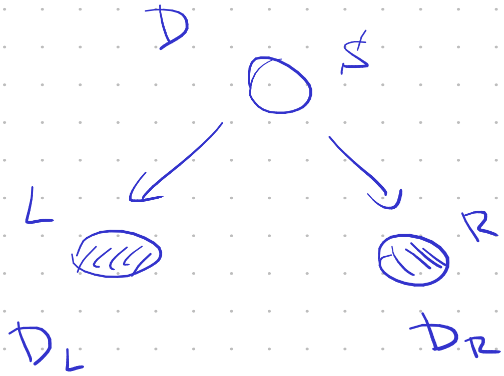
$$= \sum_{j=1}^{M} \left(w_j \cdot \underbrace{\sum_{i: q(\bar{x}_i) = j} g_i}_{\text{“} G_j(q)} + \frac{1}{2} w_j^2 \cdot \underbrace{\sum_{i: q(\bar{x}_i) = j} h_i}_{\text{“} H_j(q)} + \frac{\lambda}{2} w_j^2 \right) + \gamma M =$$

$$= \sum_{j=1}^{M} \Big(\underbrace{w_j \cdot G_j + \frac{1}{2} w_j^2 (H_j + \lambda)}_{w_j \to \min} \Big) + \gamma M = \xrightarrow{q, w_j} \min$$

$$\underbrace{w_j^* = \frac{G_j}{H_j + \lambda}}$$

$q \to \min$

$$= \sum_{j=1}^{M} \left(- \frac{G_j^2}{H_j + \lambda} + \frac{G_j^2}{2(H_j + \lambda)} \right) + \gamma M = \boxed{- \frac{1}{2} \sum_{j=1}^{M} \frac{G_j^2}{H_j + \lambda} + \gamma M}$$



$$L = -\frac{1}{2}\underbrace{\sum_j \frac{G_j^2}{H_j+\lambda}} + \gamma M$$

$$L' = -\frac{1}{2}\sum_{j\neq s} \frac{G_j^2}{H_j+\lambda} - \frac{1}{2}\left(\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda}\right) + \gamma(M+1)$$

$$\text{Gain}_s = L - L' = \frac{1}{2}\left(\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{G_s^2}{H_s+\lambda}\right) - \gamma$$

<u>XGBoost</u>:
- init
- for $t=1,\dots,T$:
  - while $\exists s$: $\underline{\text{Gain}_s > 0}$:
    - grow tree f_t
  - $F_t = f_1 + \dots + f_t$
  - compute $\hat{y}_i^{(t)}, g_i^{(t)}, h_i^{(t)}$