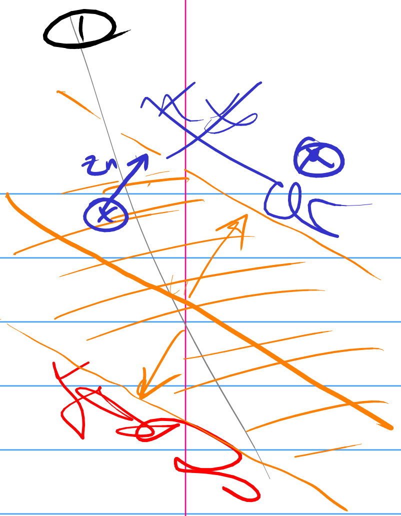


$$\frac{1}{2}\|\bar w\|^2 + C\sum_{n=1}^{N} z_n \qquad \begin{cases} z_n \ge 0 \\ t_n(\underbrace{\bar w^T \bar x_n + w_0}_{= y_n}) + z_n \ge 1 \end{cases}$$

$$z_n = \begin{cases} 0, & t_n y_n \ge 1 \\ 1 - t_n y_n, & t_n y_n < 1 \end{cases}$$

$$\sum_{n=1}^{N} z_n + \frac{\lambda}{2}\|\bar w\|^2$$

L_2-регуляризатор

$$\sum_{n=1}^{N} \underbrace{\max(0, 1 - t_n y_n)}_{= L_{SVM}(\bar x_n, t_n)} + \frac{\lambda}{2}\|\bar w\|^2$$

$t_n = \pm 1$

$$\frac{1}{1+e^{+x}} = 1 - \frac{1}{1+e^{-x}}$$

$$p(t_n = 1 \mid \bar x_n, \bar w) = \sigma(\bar w^T \bar x_n) = \sigma(y_n)$$

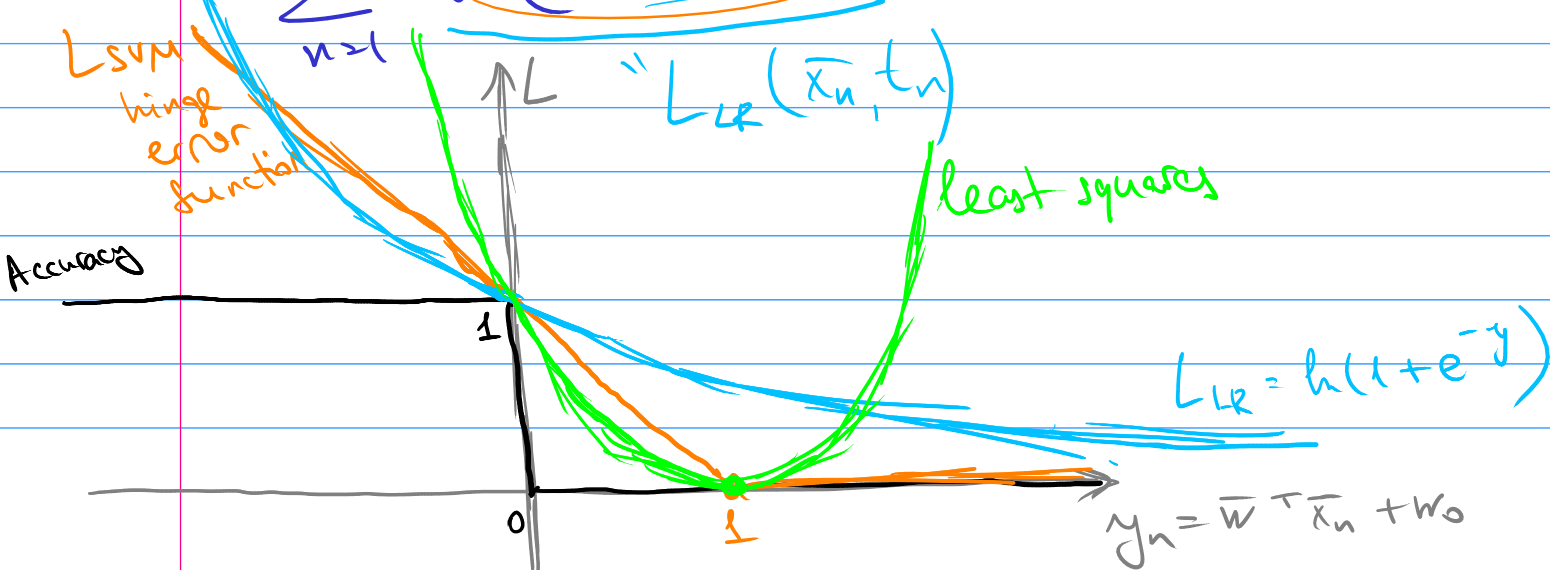
$$p(t_n = -1 \mid \text{—}) = 1 - \sigma(y_n) = \sigma(-y_n)$$

$$p(t_n \mid y_n) = \sigma(t_n y_n)$$

$$p(D \mid \bar w) = \prod_n p(t_n \mid \bar x_n, \bar w) = \prod_n p(t_n y_n)$$

$$\ln p(D \mid \bar w) = \sum_n \ln \sigma(t_n y_n) = \sum_n -\ln(1 + e^{-t_n y_n})$$

$$\sum_{n=1}^{N} \underbrace{\ln(1 + e^{-t_n y_n})}_{= L_{LR}(\bar x_n, t_n)} + \frac{\lambda}{2}\|\bar w\|^2$$



(2) Equivalent kernel

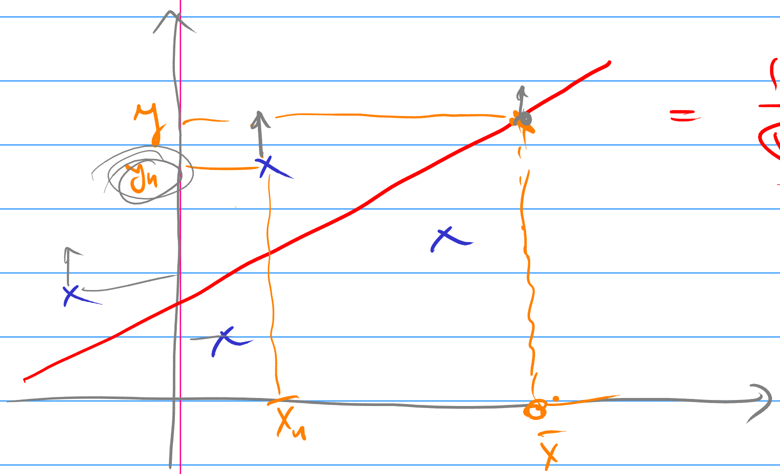
$$p(t \mid D) = \mathcal{N}(t \mid \bar{w}_{MAP}^T \bar{x}, \sigma_N^2), \qquad \sigma_N^2 = \sigma_0^2 + \bar{x}^T \Sigma_N \bar{x}$$

$$\bar{w}_{MAP} = \frac{1}{\sigma_0^2} \Sigma_N X^T \bar{y} = \frac{1}{\sigma_0^2} \sum_{n=1}^{N} \Sigma_N \bar{x}_n y_n$$

$$y(\bar{x}, \bar{w}_{MAP}) = \bar{x}^T \bar{w}_{MAP} = \frac{1}{\sigma_0^2} \sum_{n=1}^{N} (\bar{x}^T \Sigma_N \bar{x}_n) y_n =$$

$$= \frac{1}{\sigma_0^2} \sum_{n=1}^{N} \boxed{k(\bar{x}, \bar{x}_n)} y_n$$

kernel methods



precision

$$\beta = \frac{1}{\sigma_0^2}$$

hyperparameters

(3) Empirical Bayes

$$p(\bar{w} \mid D) \propto p(\bar{w})\, p(D \mid \bar{w}) = \mathcal{N}(\bar{w} \mid \bar{0}, \alpha^{-1} I) \cdot \prod_{n=1}^{N} \mathcal{N}(y_n \mid \bar{w}^T \bar{x}_n, \beta^{-1})$$

$$= \frac{1}{(2\pi)^{d/2} \cdot \sqrt{\det \frac{1}{\alpha} I}} e^{-\frac{\alpha}{2} \bar{w}^T \bar{w}} \cdot \prod_{n=1}^{N} \frac{\sqrt{\beta}}{\sqrt{2\pi}} \cdot e^{-\frac{\beta}{2}(y_n - \bar{w}^T \bar{x}_n)^2}$$

$$\ln p(\bar{w} \mid D) = \text{Const} + \frac{d}{2} \ln \alpha - \frac{\alpha}{2} \bar{w}^T \bar{w} + \frac{N}{2} \ln \beta - \frac{\beta}{2} (y_n - \bar{w}^T \bar{x}_n)^2 =$$

$$\Sigma_N = \beta X^T X + \alpha I$$

$$\bar{w}_{MAP} = \beta \Sigma_N^{-1} X^T \bar{y}$$

$$= \text{Const} - \frac{1}{2} (\bar{w} - \bar{w}_{MAP})^T \Sigma_N (\bar{w} - \bar{w}_{MAP}) + \frac{d}{2} \ln \alpha + \frac{N}{2} \ln \beta -$$

$$- \frac{\beta}{2} (\bar{y} - X \bar{w}_{MAP})^T (\bar{y} - X \bar{w}_{MAP}) - \frac{\alpha}{2} \bar{w}_{MAP}^T \bar{w}_{MAP}$$

Empirical Bayes: $p(D|\alpha,\beta) \xrightarrow[\alpha,\beta]{} \max$

$$= \int p(\bar{w}|\alpha)\, p(D|\bar{w},\beta)\, d\bar{w}$$

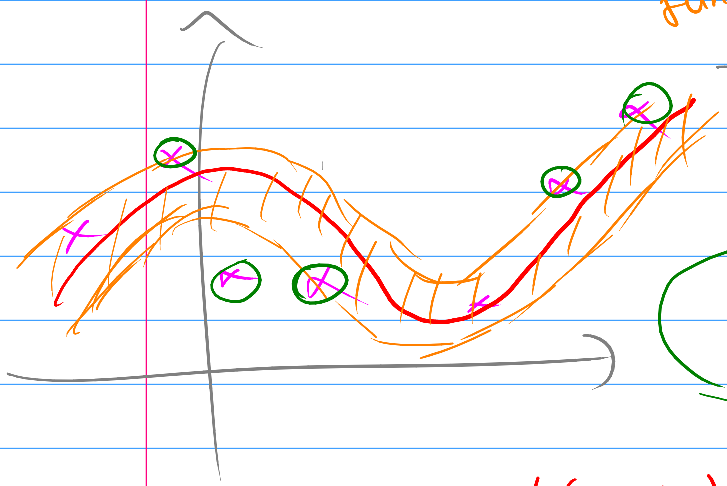
$$= e^{\dots} \underbrace{\int e^{-\frac{1}{2}(\bar{w}-\bar{w}_{MAP})^T \Sigma_N (\bar{w}-\bar{w}_{MAP})} d\bar{w}}_{=\sqrt{\frac{(2\pi)^d}{\det \Sigma_N}}}$$

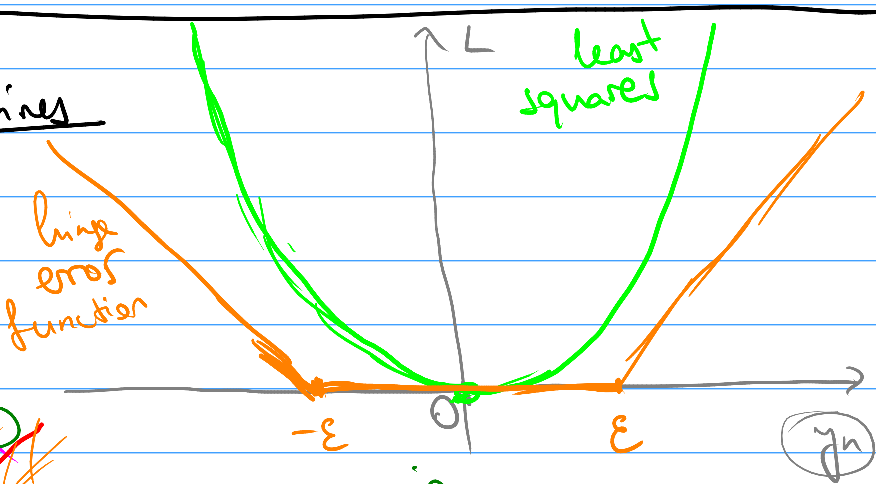
$$\ln p(D|\alpha,\beta) = \text{Const} - \frac{1}{2}\ln\det\Sigma_N + \frac{d}{2}\ln\alpha + \frac{N}{2}\ln\beta - \frac{\beta}{2}(\_\_)^T(\_\_) - \frac{\alpha}{2}\bar{w}_{MAP}^T\bar{w}_{MAP} \xrightarrow[\alpha,\beta]{} \max$$

---

④ Relevance vector machines

SVM for regression





$\bar{w} \to \min$

$$\sum_n L(y_n, t_n) + \frac{\lambda}{2}\|\bar{w}\|^2$$

$$L(y_n,t_n) = \begin{cases} 0, & |y_n - t_n| < \varepsilon \\ |y_n - t_n| - \varepsilon, & |y_n - t_n| \geq \varepsilon \end{cases}$$

$$y_n - \varepsilon \leq t_n \leq y_n + \varepsilon$$

$$\begin{cases} t_n \leq y_n + \varepsilon + z_n^+ \\ t_n \geq y_n - \varepsilon - z_n^- \end{cases}$$

$z_n^+ \geq 0,\ z_n^- \geq 0$

$$p(t|\bar{x},\bar{w}) = \mathcal{N}(t|\bar{x}^T\bar{w}, \beta^{-1}) \qquad y = \bar{x}^T\bar{w}$$

$$y(\bar{x}) = \sum_{n=1}^{N} w_n k(\bar{x}, \bar{x}_n) + w_0 \qquad \bar{x} \mapsto \begin{pmatrix} k(\bar{x},\bar{x}_1) \\ \cdots \\ k(\bar{x},\bar{x}_N) \\ 1 \end{pmatrix}$$

$$p(\bar{t}|X,\bar{w}) = \prod_{n=1}^{N} \mathcal{N}(t_n|y(\bar{x}_n), \beta^{-1})$$

$$p(\bar{w}|\bar{\alpha}) = \prod_{n=0}^{N} \mathcal{N}(w_n|0, \alpha_n^{-1}) \qquad K = \begin{pmatrix} & | & \\ - & k(\bar{x}_n,\bar{x}_m) & 1 \\ & | & \end{pmatrix}_{N \times (N+1)}$$

$$p(\bar{w}|\bar{t}) = \mathcal{N}(\bar{w}|\bar{\mu}, \Sigma), \qquad \bar{\mu} = \beta \Sigma K^T \bar{t}$$

$$\Sigma^{-1} = \begin{pmatrix} \alpha_0 & & 0 \\ & \ddots & \\ 0 & & \alpha_N \end{pmatrix} + \beta \underbrace{K^T K}_{(N+1)\times(N+1)}$$

$$\ln p(\bar{t}|X,\bar{\alpha},\beta) = \ln \int p(\bar{w}|\alpha)\, p(\bar{t}|X,\bar{w},\beta)\, d\bar{w} = \text{упр}$$

$$= \ln \mathcal{N}(\bar{t}|\bar{0}, C) = -\frac{1}{2}\left(N\ln 2\pi + \ln \det C + \bar{t}^T C^{-1} \bar{t}\right)$$

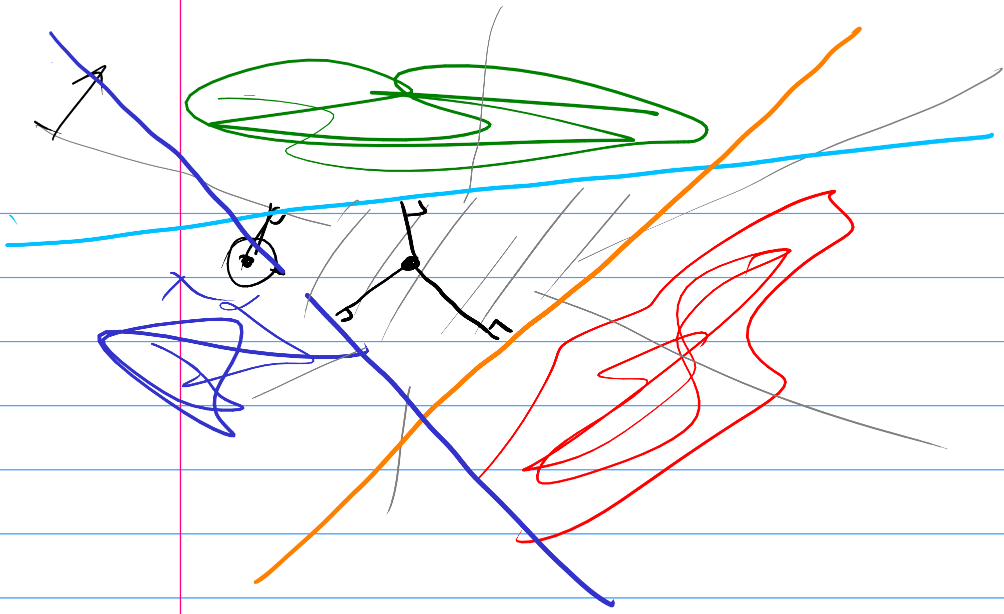
$$C = \beta^{-1} I + K \cdot \begin{pmatrix} \alpha_0 & & \\ & \ddots & \\ & & \alpha_N \end{pmatrix}^{-1} K^T \qquad \xrightarrow{\bar{\alpha},\beta} \max$$

$$\alpha_i = \frac{\gamma_i}{\mu_i^2}, \quad \beta^{-1} = \frac{\|\bar{t} - K\bar{\mu}\|^2}{N - \sum \gamma_i}$$

часто $\alpha_n \to \infty$, т.е. $\sigma_n^2 \to 0$

$$y(\bar{x},\bar{w}) = \sigma\left(\sum \bar{w}_n k(\bar{x},\bar{x}_n) + w_0\right) \qquad \text{softmax}$$

$$p(\bar{w}|\bar{\alpha}) = \prod \mathcal{N}(w_i|0, \alpha_i^{-1})$$



one vs. one

$K \to \frac{K(K-1)}{2}$ классиф.

one vs. all

$K \to K$ классиф.

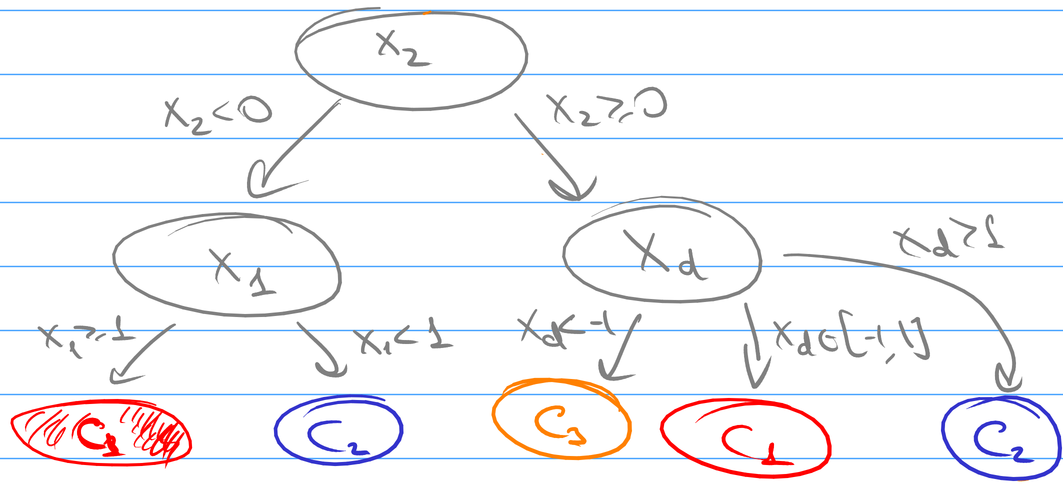
C_k vs $C_1 \cup \ldots \cancel{C_k} \cup \ldots C_K$

$p(C_k)$ $p(C_{-k})$

Decision tree

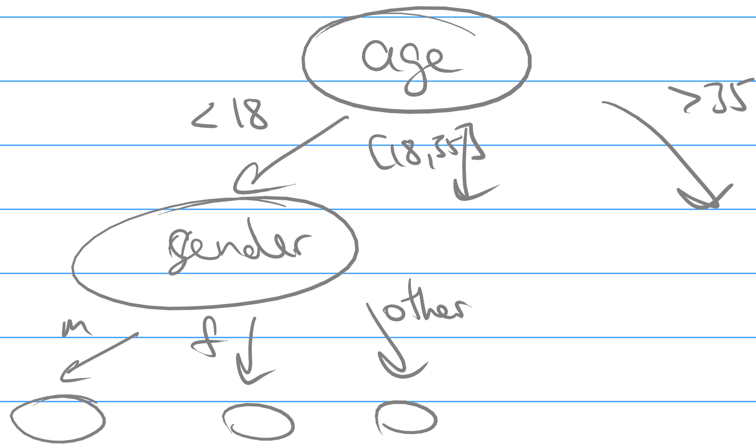
$\bar{x} = (x_1, x_2, \ldots, x_d)$

$\bar{x} \in C_k$, $k = 1, \ldots, K$



Recommender system

$\bar{x}$ = (gender, age, income, geo)

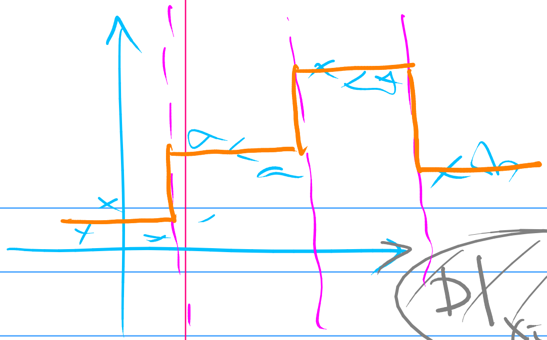


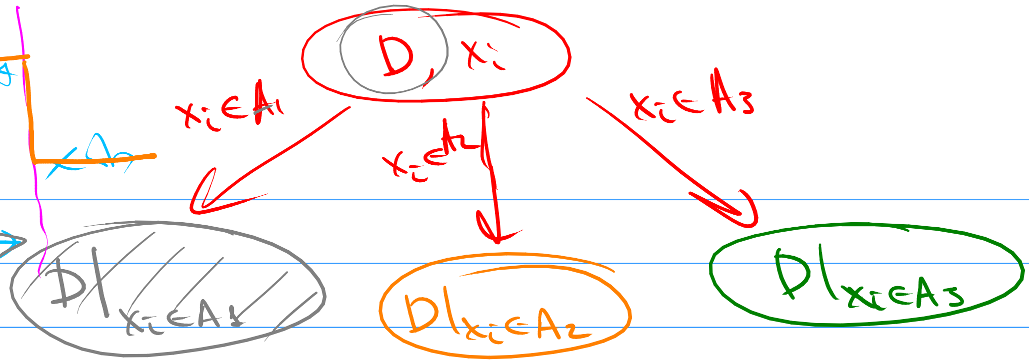
0) Проверить, не пора ли остан.

D : 1) Выбр. атрибут x_i

2) —"— расщепление

D





в листе: большинство или
среднее

$$D = \{(\bar{x}, y)\}, \quad y \in \{C_1, \dots, C_k\}$$

$$x_i \in \{a_1, a_2, \dots, a_l\}$$

$D_1 = D|_{x_i = a_1}$ $D_2 = D|_{x_i = a_2}$ — $D_l = D|_{x_i = a_l}$

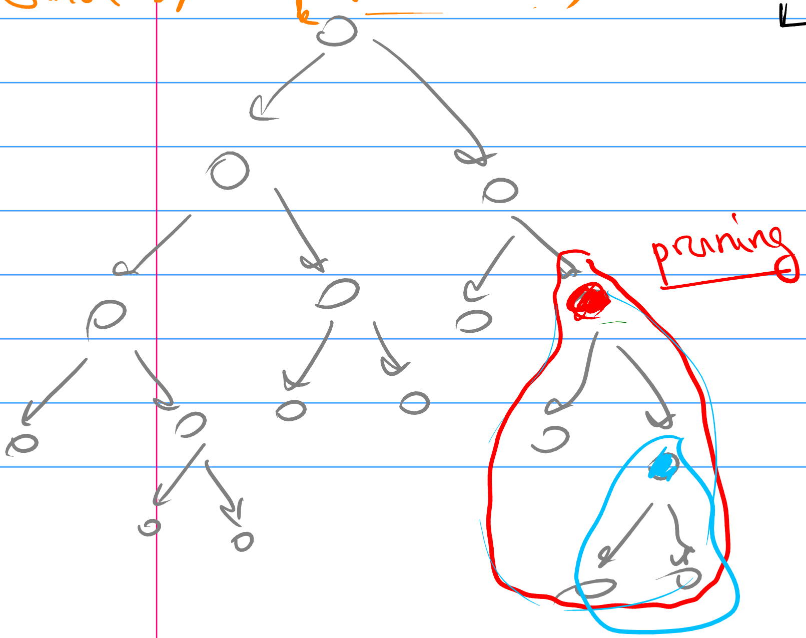
blue   red   green

$$Q(D_1) = -\sum_k p_{1,k} \ln p_{1,k} \qquad Q(D_2) \quad \dots \quad Q(D_l) = -\sum_k p_{l,k} \ln p_{l,k}$$

$$p_{1,k} = \frac{\#\{x \in C_k \text{ в } D_1\}}{|D_1|}$$

$$Gini(D_1) = \sum_k p_{1,k}(1 - p_{1,k})$$

$$Gain(x_i) = \sum_{s=1}^{l} \frac{|D_s|}{|D|} Q(D_s)$$



$$Accuracy + \lambda \cdot |T|$$