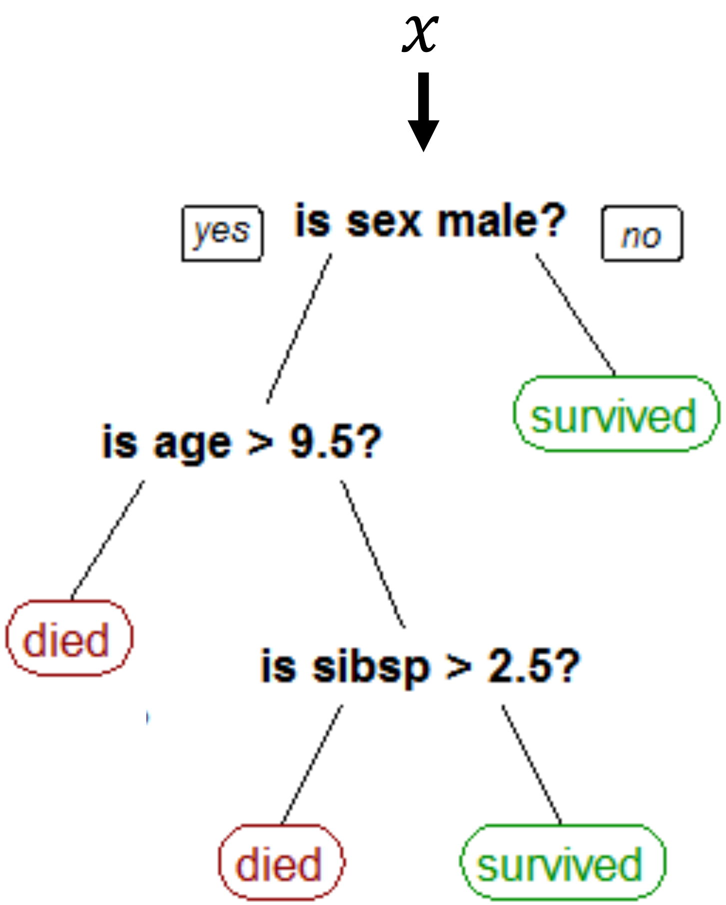
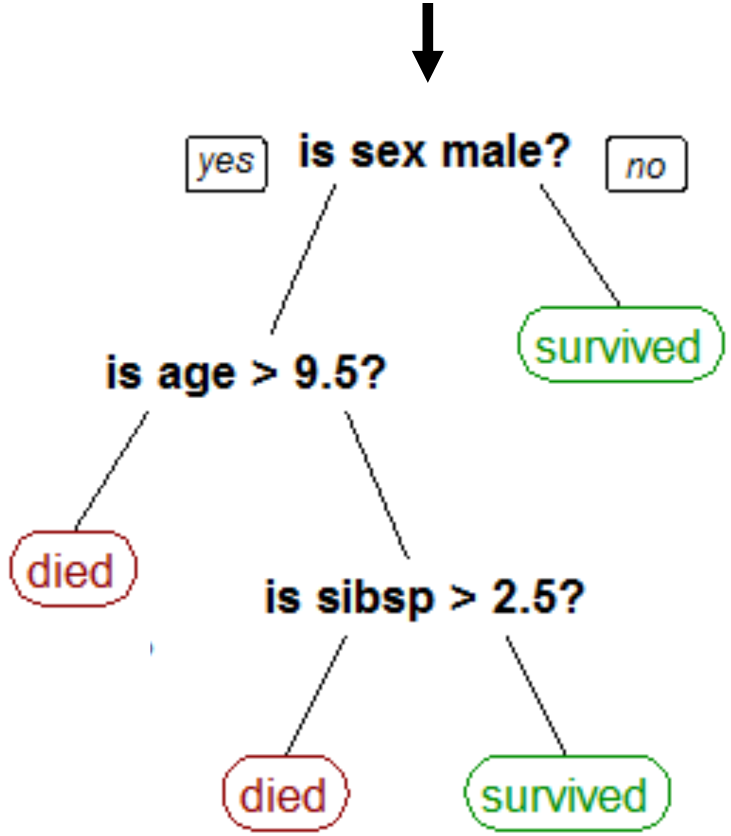
1. Что такое решающие деревья

Решающее дерево



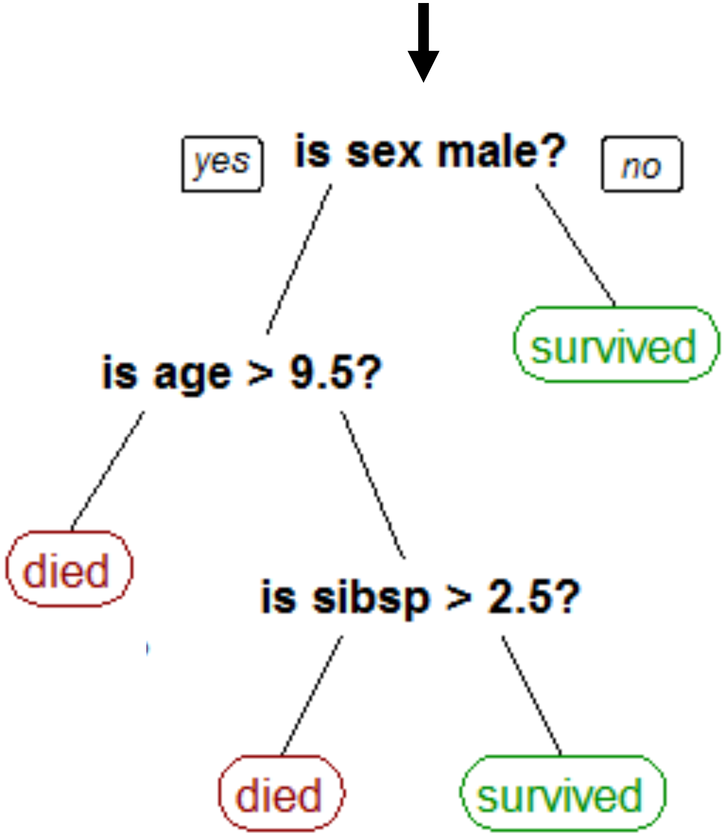
Решающее дерево

$$x = (9, \quad male, \quad 3)$$



Решающее дерево

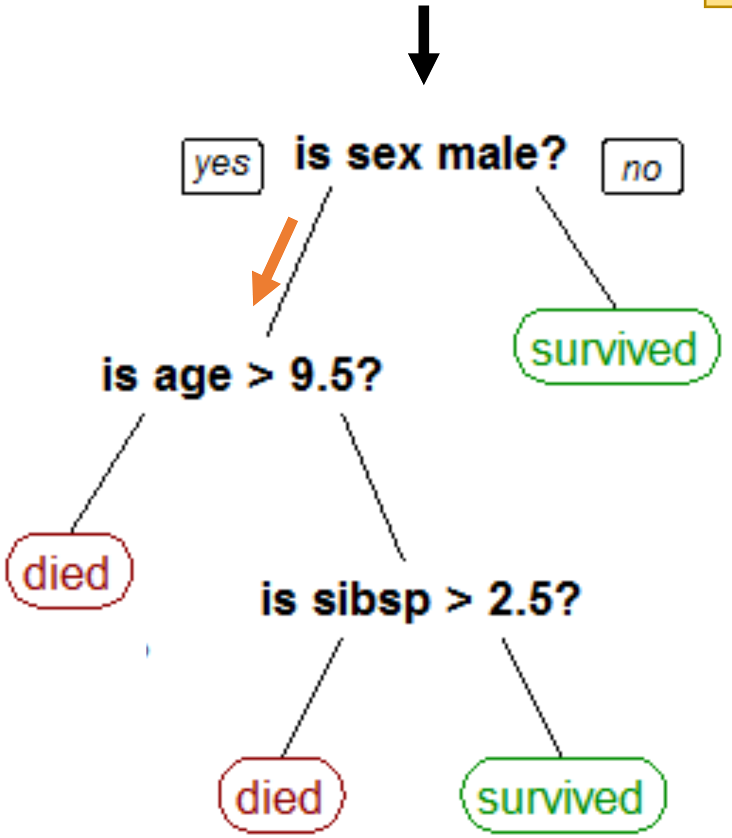
$$\begin{array}{rccc} & \text{age} & \text{sex} & \text{sibsp} \\ x = (& 9, & male, & 3) \end{array}$$



Решающее дерево

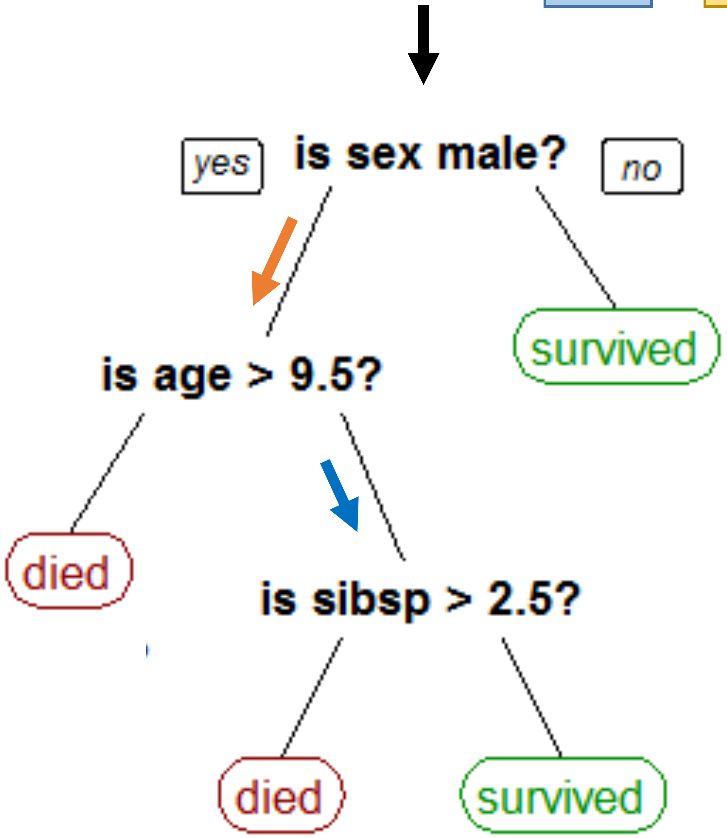
age sex sibsp

$$x = (9, \boxed{male,} \ 3)$$



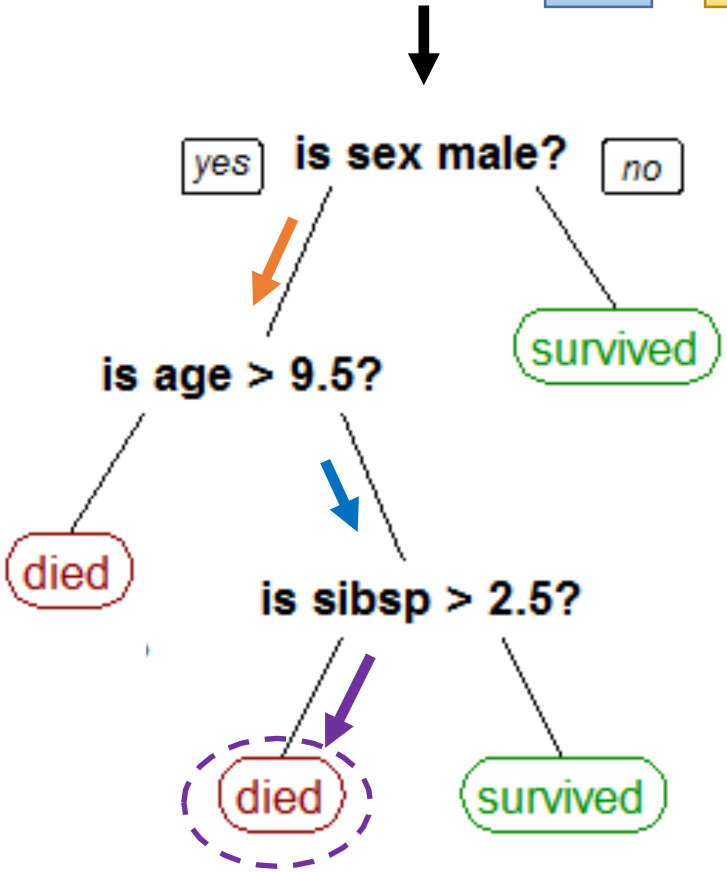
Решающее дерево

$$x = \underset{\text{age}}{\boxed{(9,}}\ \underset{\text{sex}}{\boxed{male,}}\ \underset{\text{sibsp}}{3)}$$



Решающее дерево

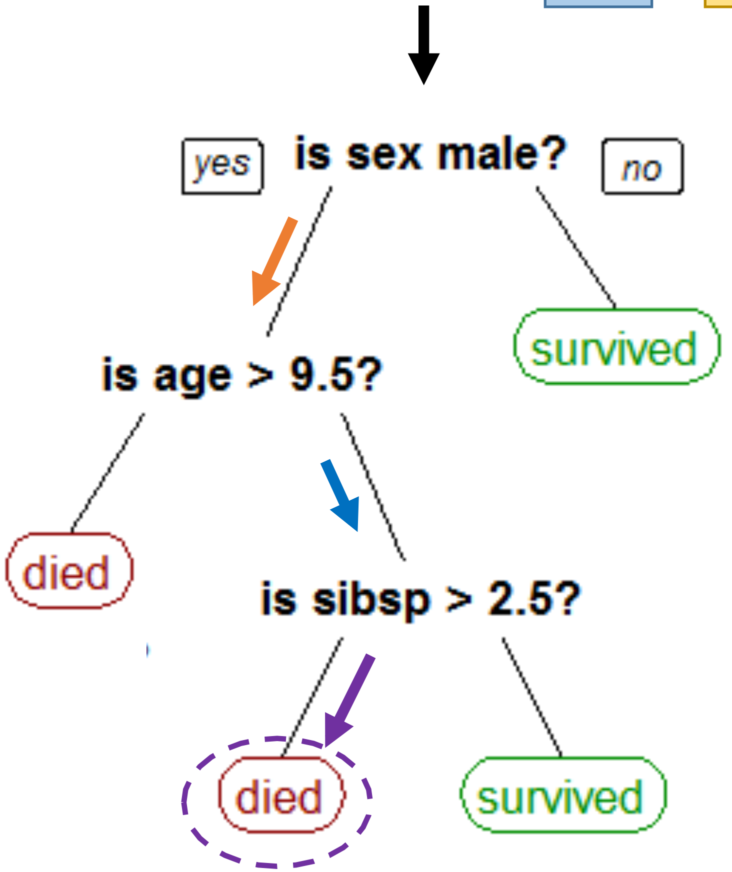
$$x = (\underset{\text{age}}{\boxed{9,}}\ \underset{\text{sex}}{\boxed{male,}}\ \underset{\text{sibsp}}{\boxed{3)}}$$



Решающее дерево

age sex sibsp

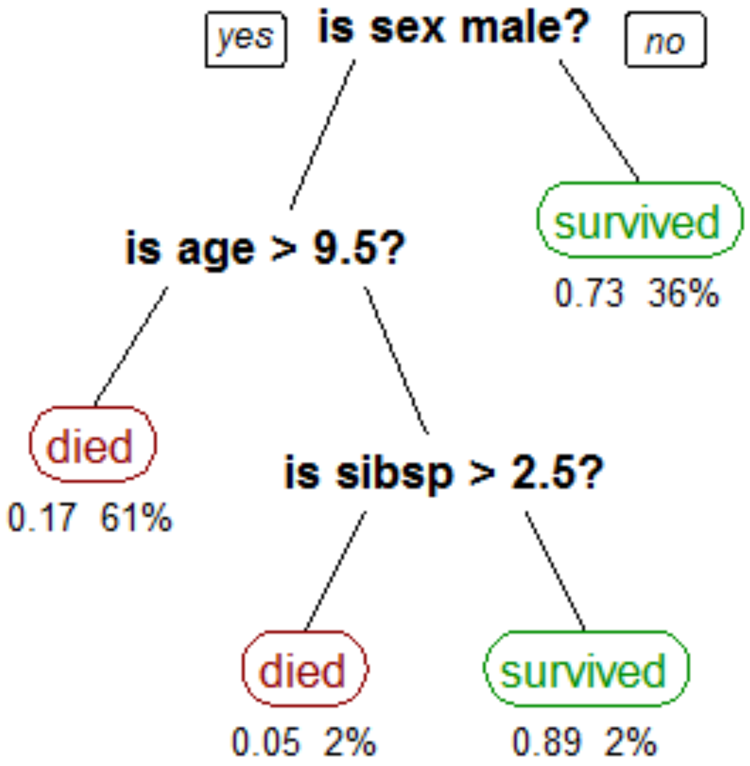
$$x = (9, \; male, \; 3)$$



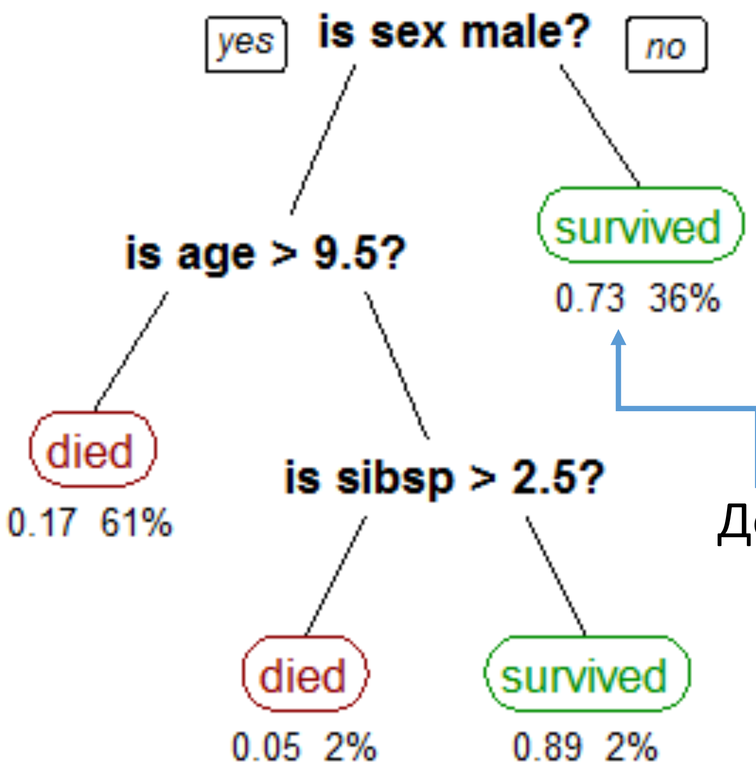
$$y = died$$

2. Решающие деревья в классификации и регрессии

Решающее дерево: классификация



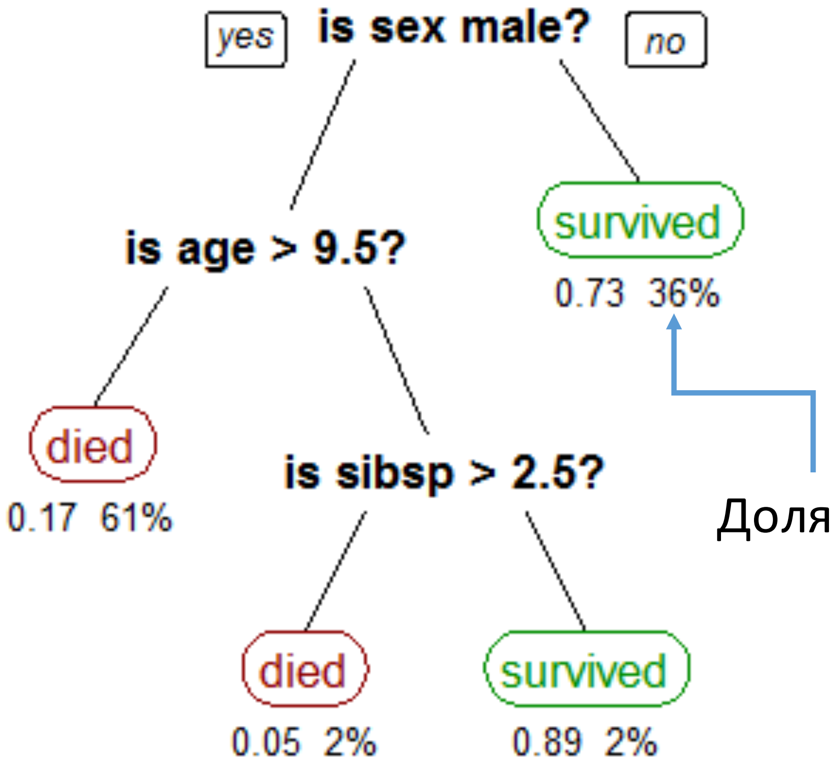
Решающее дерево: классификация



Доля выживших среди попавших в лист

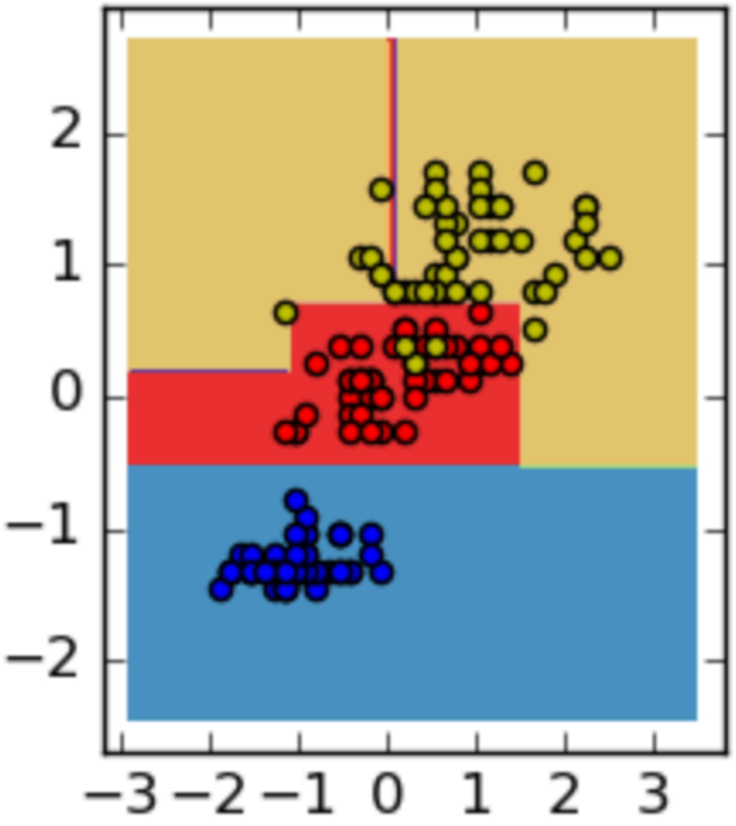
В каждом листе дерево отвечает преобладающим классом

Решающее дерево: классификация



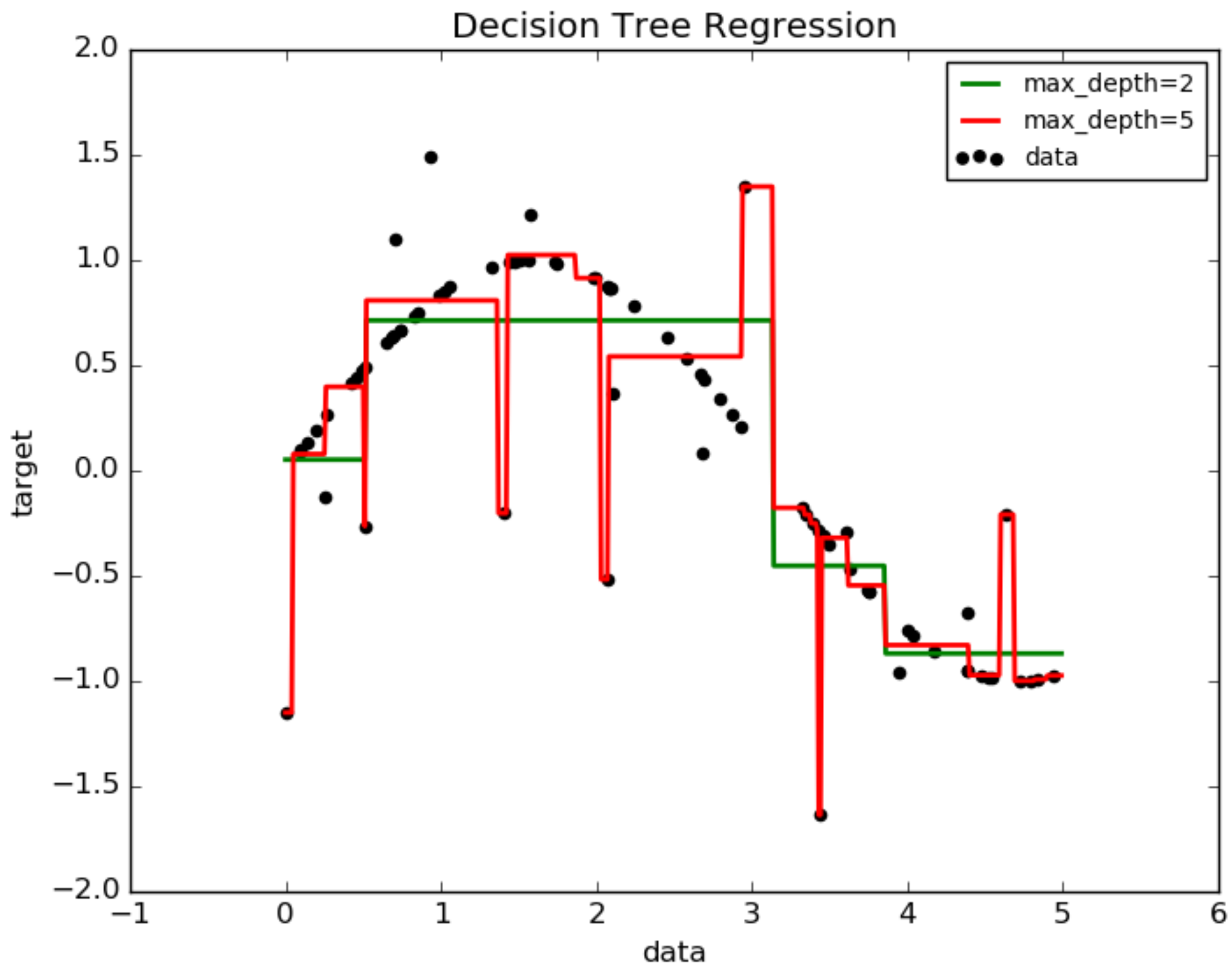
Доля объектов обучающей выборки

Решающее дерево: классификация



Пример: 3 класса и 2 признака

Решающее дерево: регрессия



Пример: восстановление зависимости $y(x) = \sin x$ с помощью решающих деревьев глубины 2 и глубины 5

В каждом листе дерево отвечает некоторой константой

3. Как строить решающие деревья

Рекурсивное построение

Строим разбиение выборки по значению одного из признаков

$$x^{(j)} < t$$

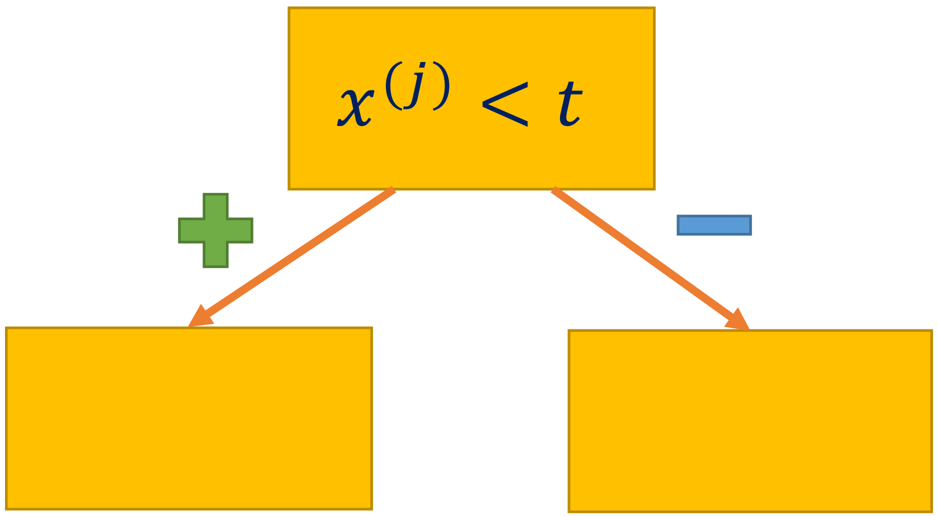
Рекурсивное построение

Строим разбиение выборки по значению одного из признаков

$$x^{(j)} < t$$

Фактически нужно только выбрать j и t наилучшим образом

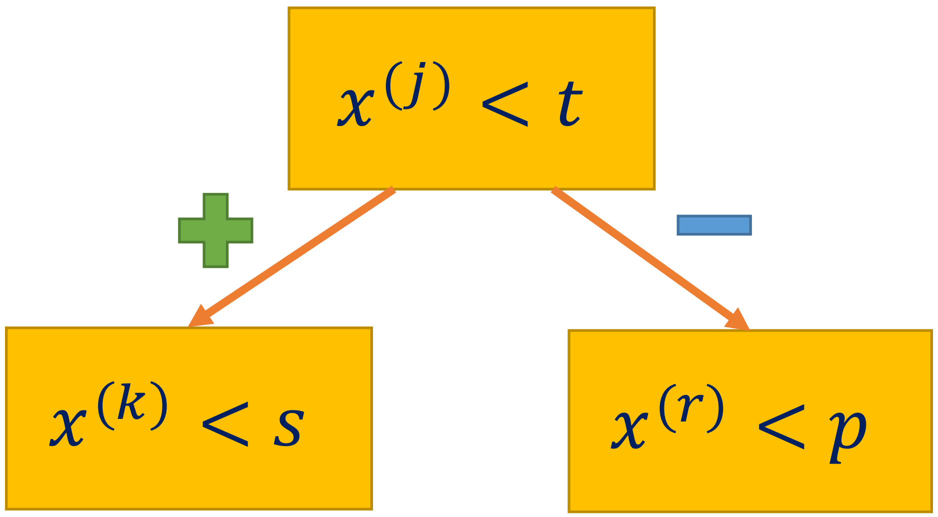
Рекурсивное построение



Выборка делится
по этому условию
на две части

Рекурсивное построение

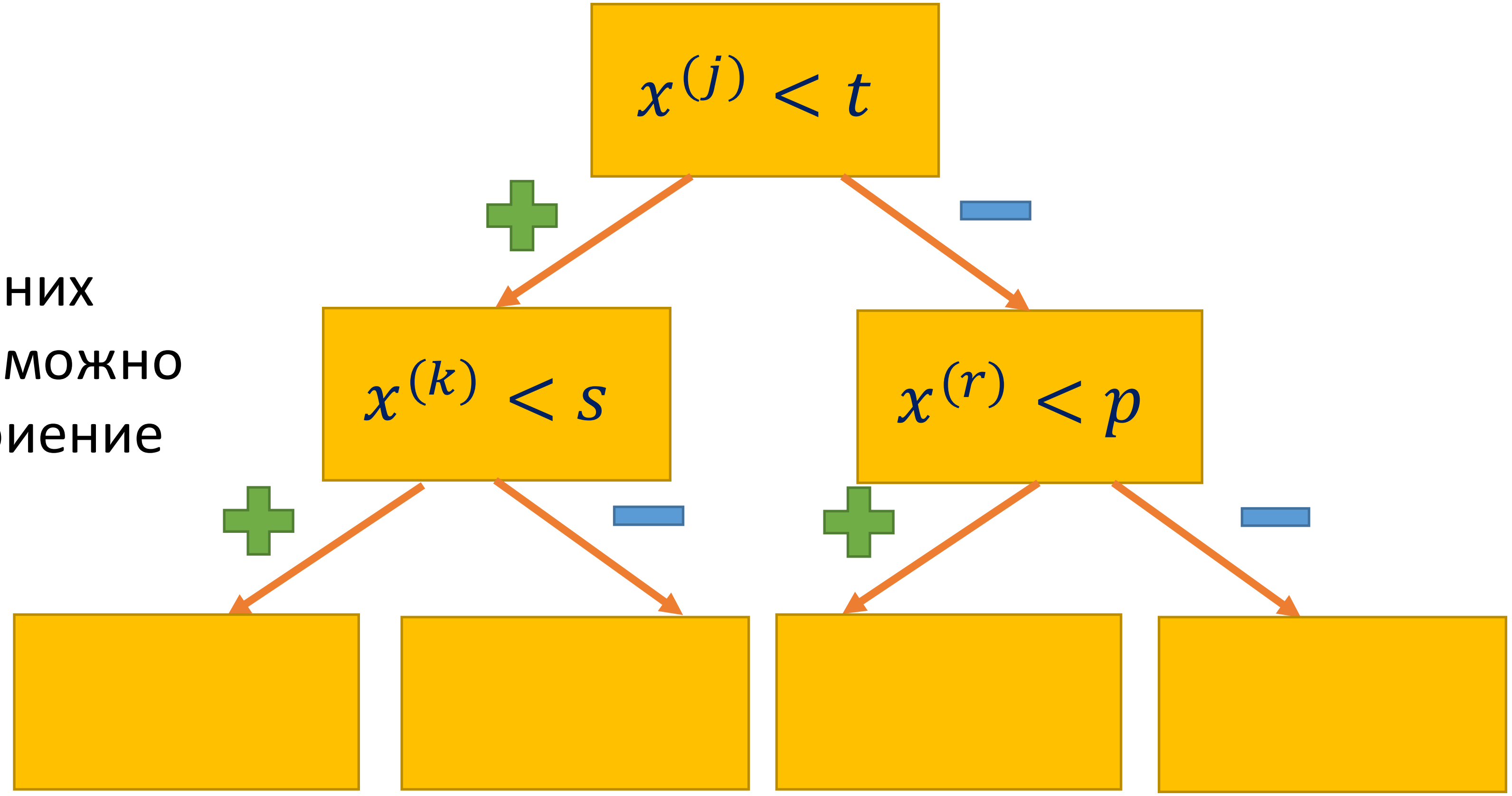
В каждой из них теперь тоже можно сделать разбиение



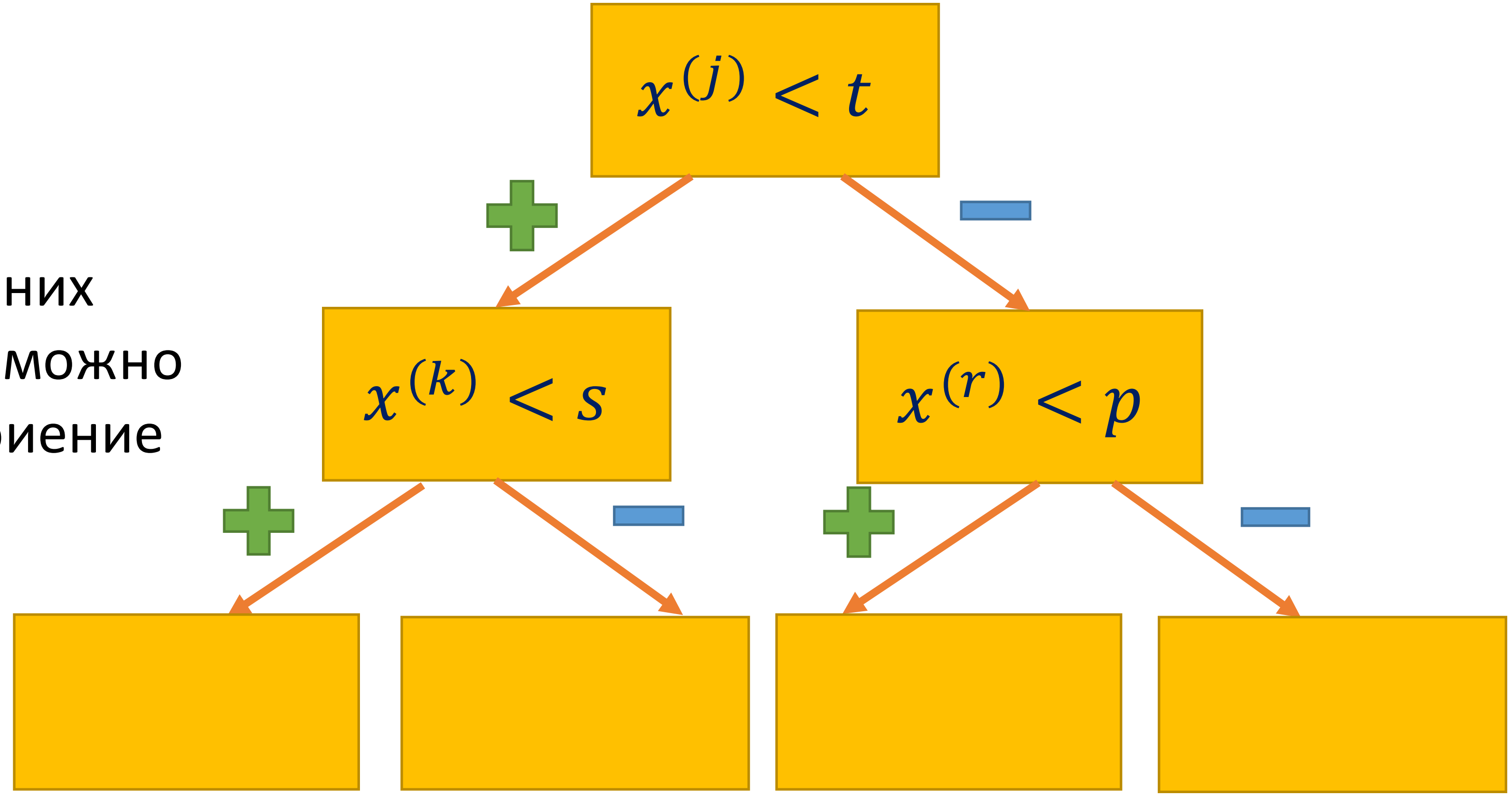
Выборка делится по этому условию на две части

Рекурсивное построение

В каждой из них теперь тоже можно сделать разбиение



Рекурсивное построение



В каждой из них теперь тоже можно сделать разбиение

Процесс можно продолжать в тех узлах, в которые попадает достаточно много объектов

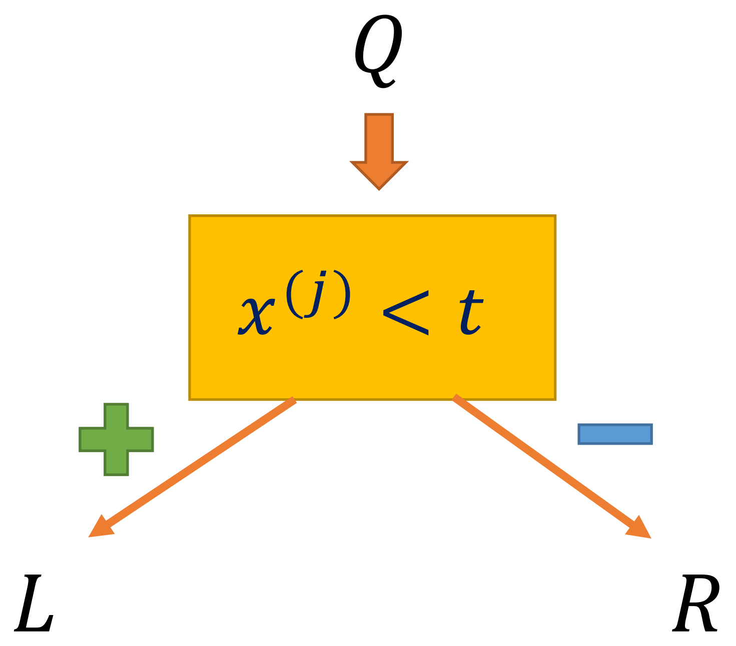
Выбор разбиения

$$Q$$

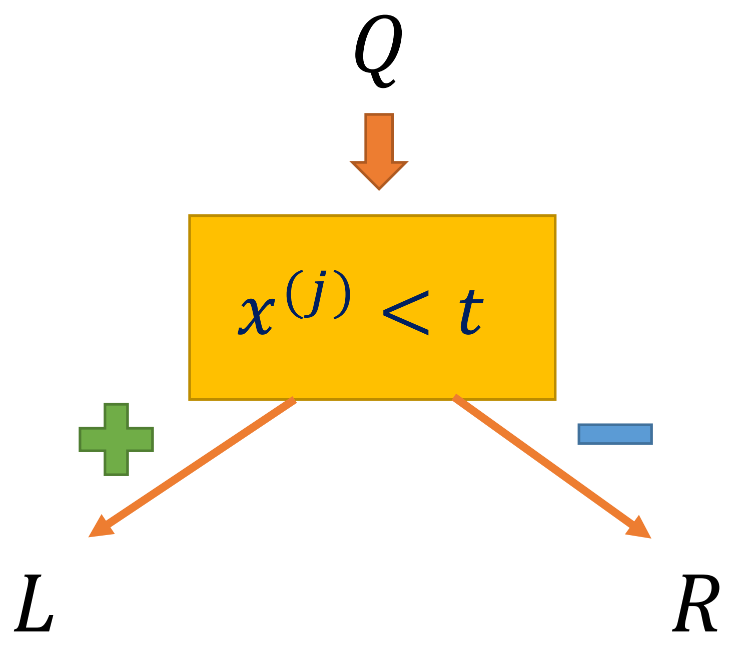
$$\downarrow$$

$$\boxed{x^{(j)} < t}$$

Выбор разбиения

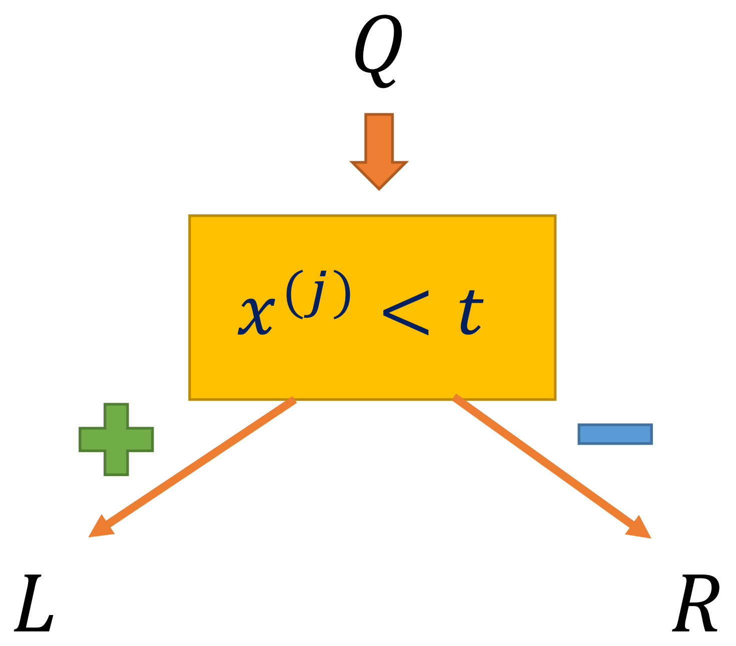


Выбор разбиения



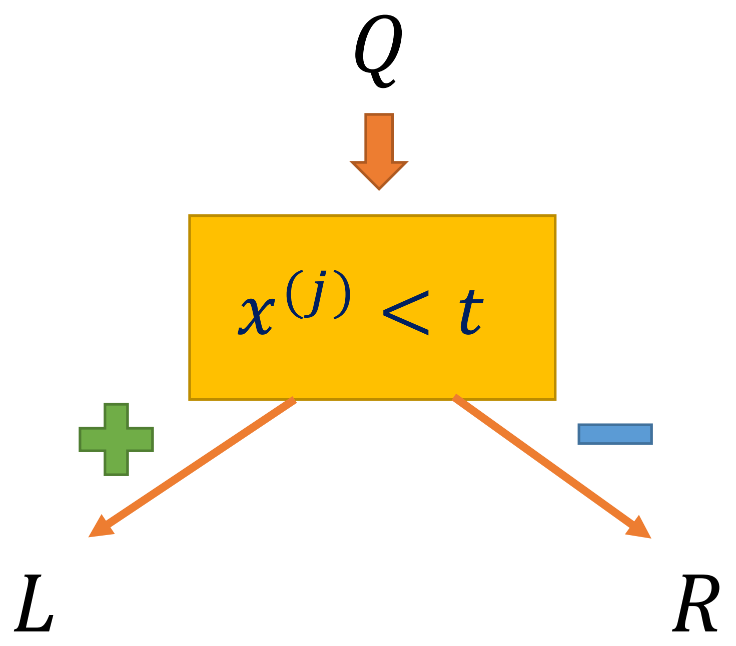
$$G(j,t) = \frac{|L|}{|Q|} H(L) + \frac{|R|}{|Q|} H(R)$$

Выбор разбиения



$$G(j,t) = \frac{|L|}{|Q|} H(L) + \frac{|R|}{|Q|} H(R) \to \min_{j,t}$$

Выбор разбиения



$$G(j,t) = \frac{|L|}{|Q|} H(L) + \frac{|R|}{|Q|} H(R) \to \min_{j,t}$$

$H(R)$ - мера «неоднородности» множества R

Критерии построения разбиений

$H(R) -$ мера «неоднородности» множества R

Критерии построения разбиений

$H(R) -$ мера «неоднородности» множества R

Пусть мы решаем задачу классификации на 2 класса,
p_0, p_1 $-$ доли объектов классов 0 и 1 в R

1) Misclassification criteria: $H(R) = 1 - p_{max}$

2) Entropy criteria: $H(R) = -p_0 \ln p_0 - p_1 \ln p_1$

3) Gini criteria: $H(R) = 1 - {p_0}^2 - {p_1}^2 = 2p_0 p_1$

Критерии построения разбиений

$H(R) -$ мера «неоднородности» множества R

Пусть мы решаем задачу классификации на K классов,
$p_1, \dots, p_K$ $-$ доли объектов классов 1, …, K в R

1) Misclassification criteria: $H(R) = 1 - p_{max}$
2) Entropy criteria:

$$H(R) = -\sum_{k=1}^{K} p_k \ln p_k$$

3) Gini criteria:

$$H(R) = \sum_{k=1}^{K} p_k(1 - p_k)$$

Критерии построения разбиений

$H(R)$ – мера «неоднородности» множества R

Чтобы решать задачу регрессии, достаточно взять среднеквадратичную ошибку в качестве $H(R)$:

$$H(R) = \frac{1}{|R|} \sum_{x_i \in R} (y_i - \bar{y})^2$$

Критерии построения разбиений

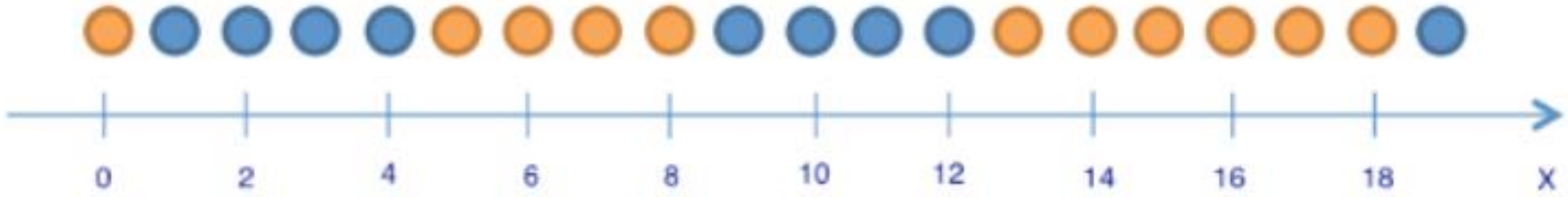
$H(R)$ – мера «неоднородности» множества R

Чтобы решать задачу регрессии, достаточно взять среднеквадратичную ошибку в качестве $H(R)$:

$$H(R) = \frac{1}{|R|} \sum_{x_i \in R} (y_i - \bar{y})^2$$

$$\bar{y} = \frac{1}{|R|} \sum_{x_i \in R} y_i$$

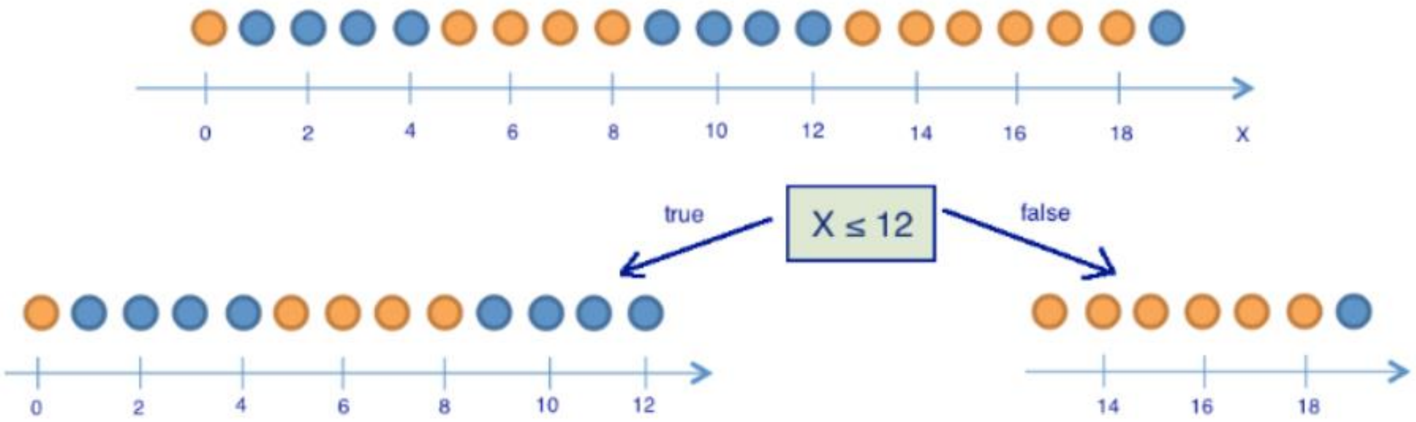
Пример из ODS курса



Entropy criteria

$$H(X) = -\sum_{i=1}^{N} p_i log_2 p_i$$

$$S_0 = -\tfrac{9}{20}\log_2 \tfrac{9}{20} - \tfrac{11}{20}\log_2 \tfrac{11}{20} \approx 1$$



$$\frac{13}{20} * H_l + \frac{7}{20} * H_r \approx 0.83$$

$$H_l = -\frac{5}{13} log_2 \frac{5}{13} - \frac{8}{13} log_2 \frac{8}{13} \approx 0.96$$

$$H_r = -\frac{1}{7} log_2 \frac{1}{7} - \frac{6}{7} log_2 \frac{6}{7} \approx 0.6$$

Gini criteria

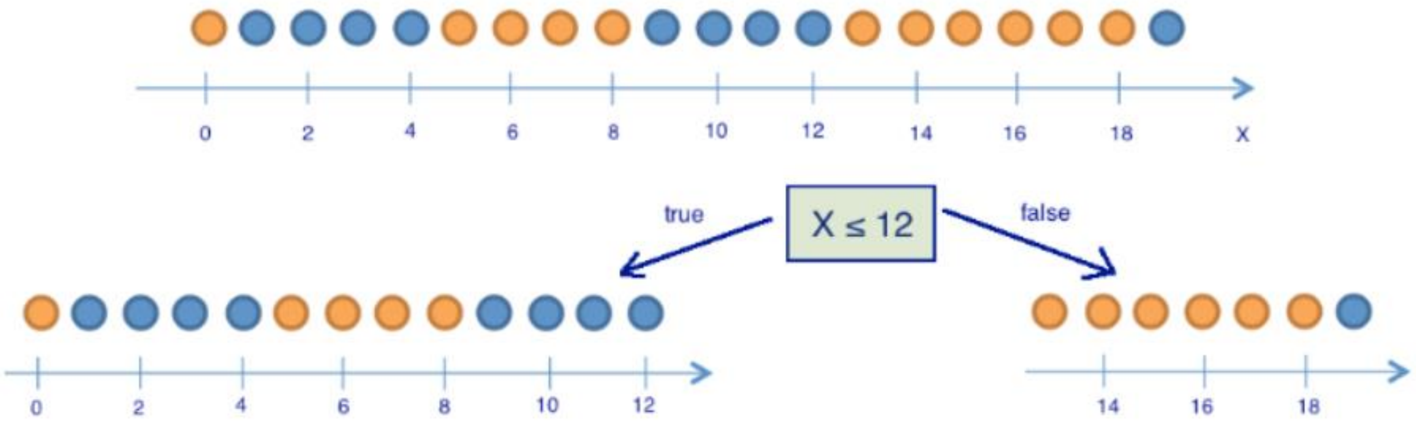
$$p_k = \frac{1}{X} \sum_{i \in X} [y_i = k]$$

$$H(X) = \sum_{k=1}^{K} p_k (1 - p_k)$$

$$H(X) = 1 - \sum_{k} (p_k)^2$$

$$Q = \frac{13}{20} * H_r + \frac{7}{20} * H_l \approx 0.33$$

$$S_0 = 1 - \left(\frac{9}{20}\right)^2 - \left(\frac{11}{20}\right)^2 = 0.495$$



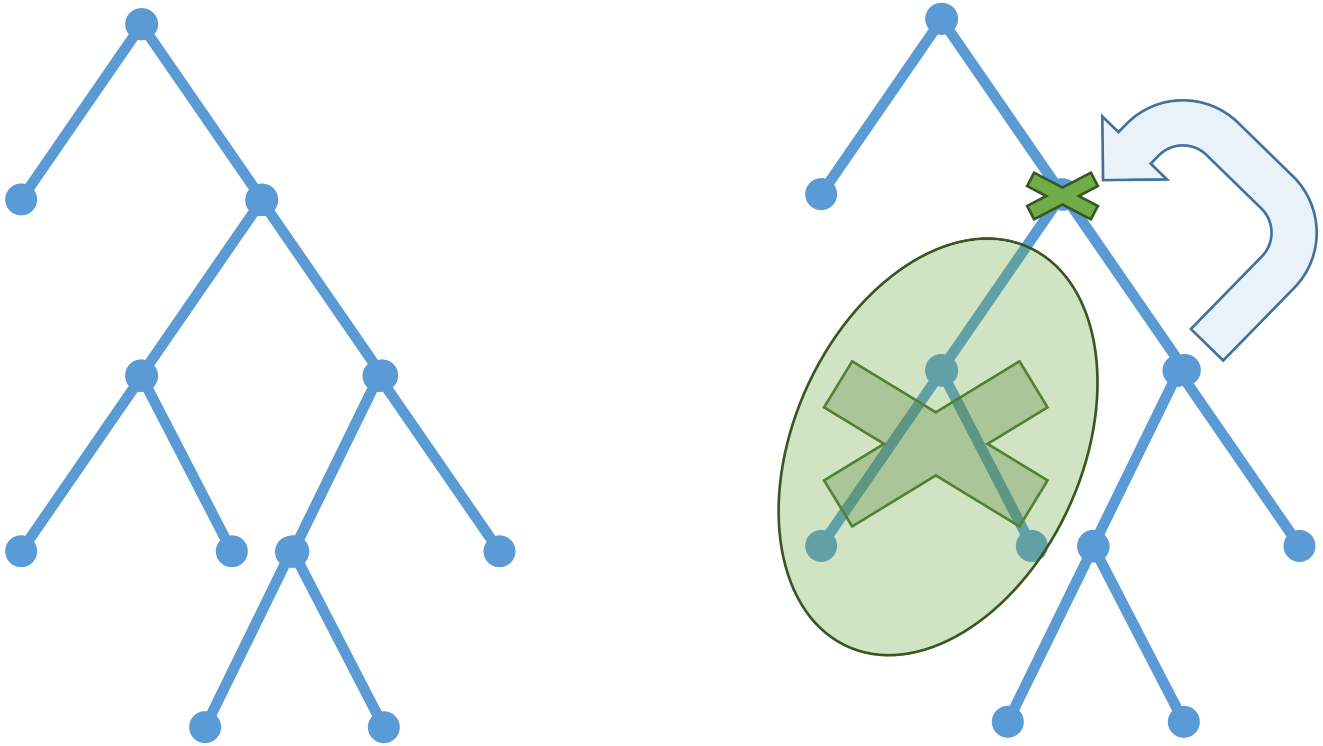
$$H_l = 1 - \left(\frac{5}{13}\right)^2 - \left(\frac{8}{13}\right)^2 \approx 0.38$$

$$H_l = 1 - \left(\frac{1}{7}\right)^2 - \left(\frac{6}{7}\right)^2 \approx 0.24$$

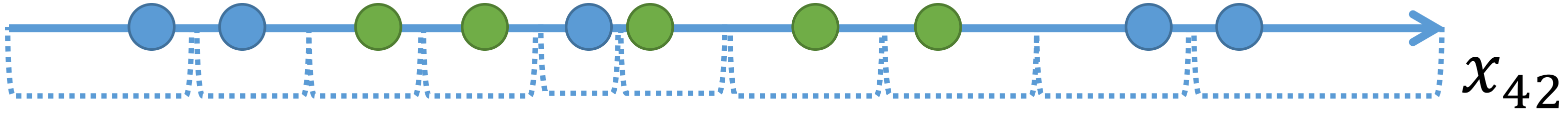
Prunning

- Pre-prunning:
  - Ограничиваем рост дерева до того как оно построено
  - Если в какой-то момент информативность признаков в разбиении меньше порога – не разбиваем вершину
- Post-prunning:
  - Упрощаем дерево после того как дерево построено

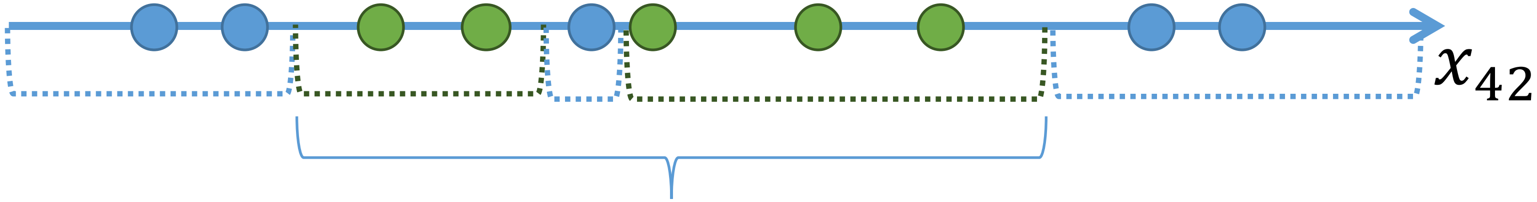
Post-prunning



Бинаризация



$$\varphi(x) = [a_k < x \le a_{k+1}]$$



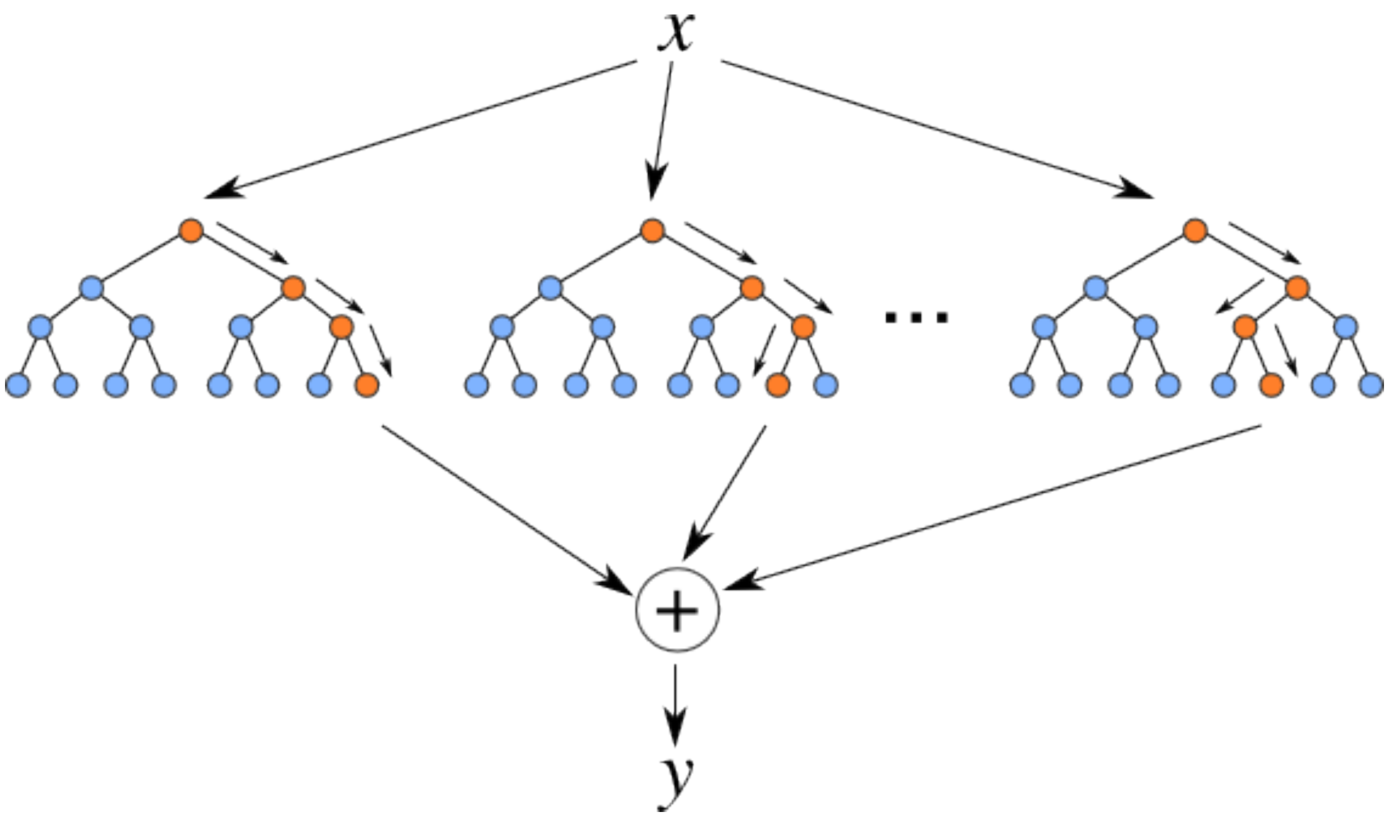
Вариации алгоритма построения

- C4.5
- C5.0
- CART

II. Ансамбли деревьев

Random Forest

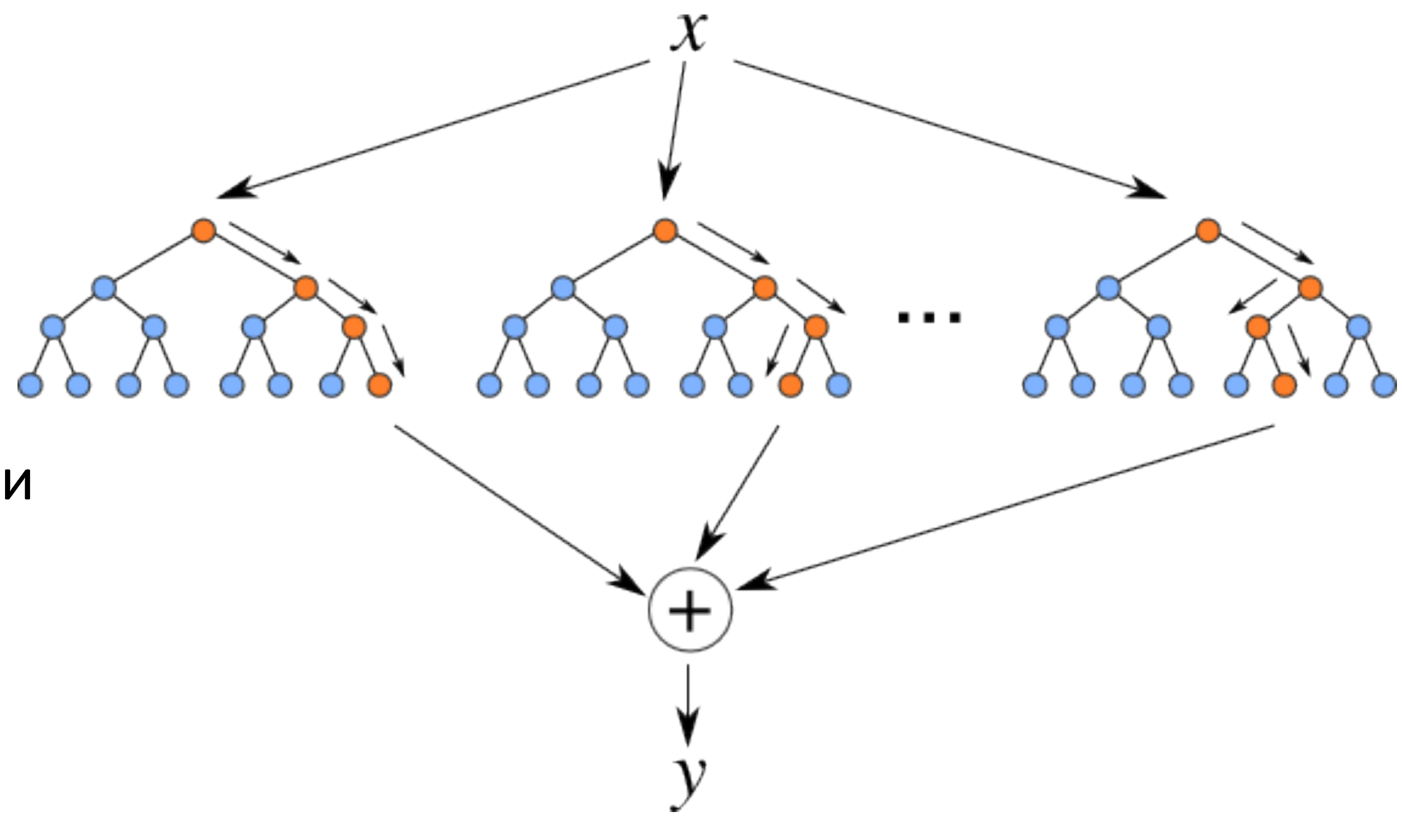
1. Генерируем M выборок на основе имеющейся (по схеме выбора с возвращением)



Random Forest

1. Генерируем M выборок на основе имеющейся (по схеме выбора с возвращением)

2. Строим на них деревья с рандомизированными разбиениями в узлах: выбираем k случайных признаков и ищем наиболее информативное разбиение по ним

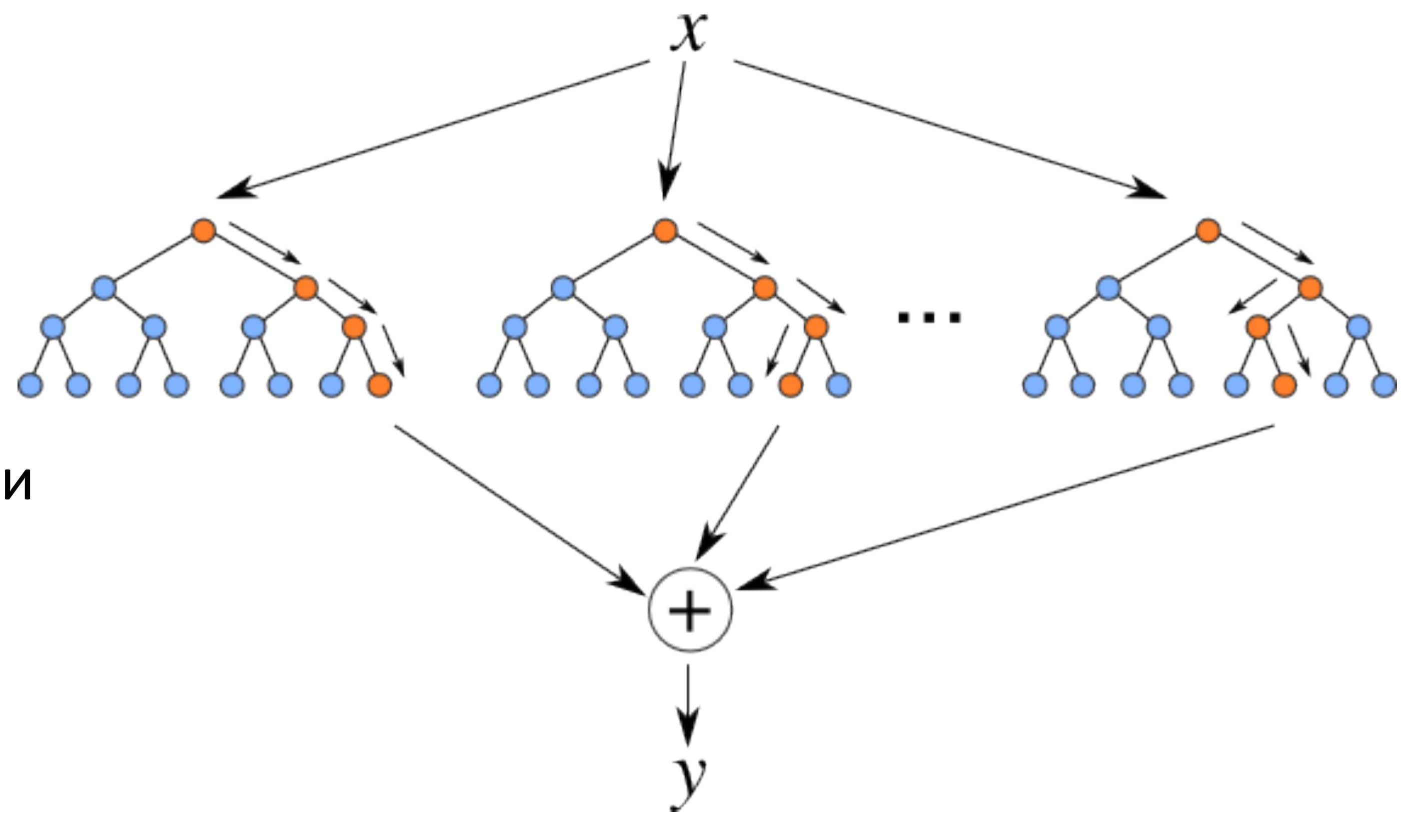


Random Forest

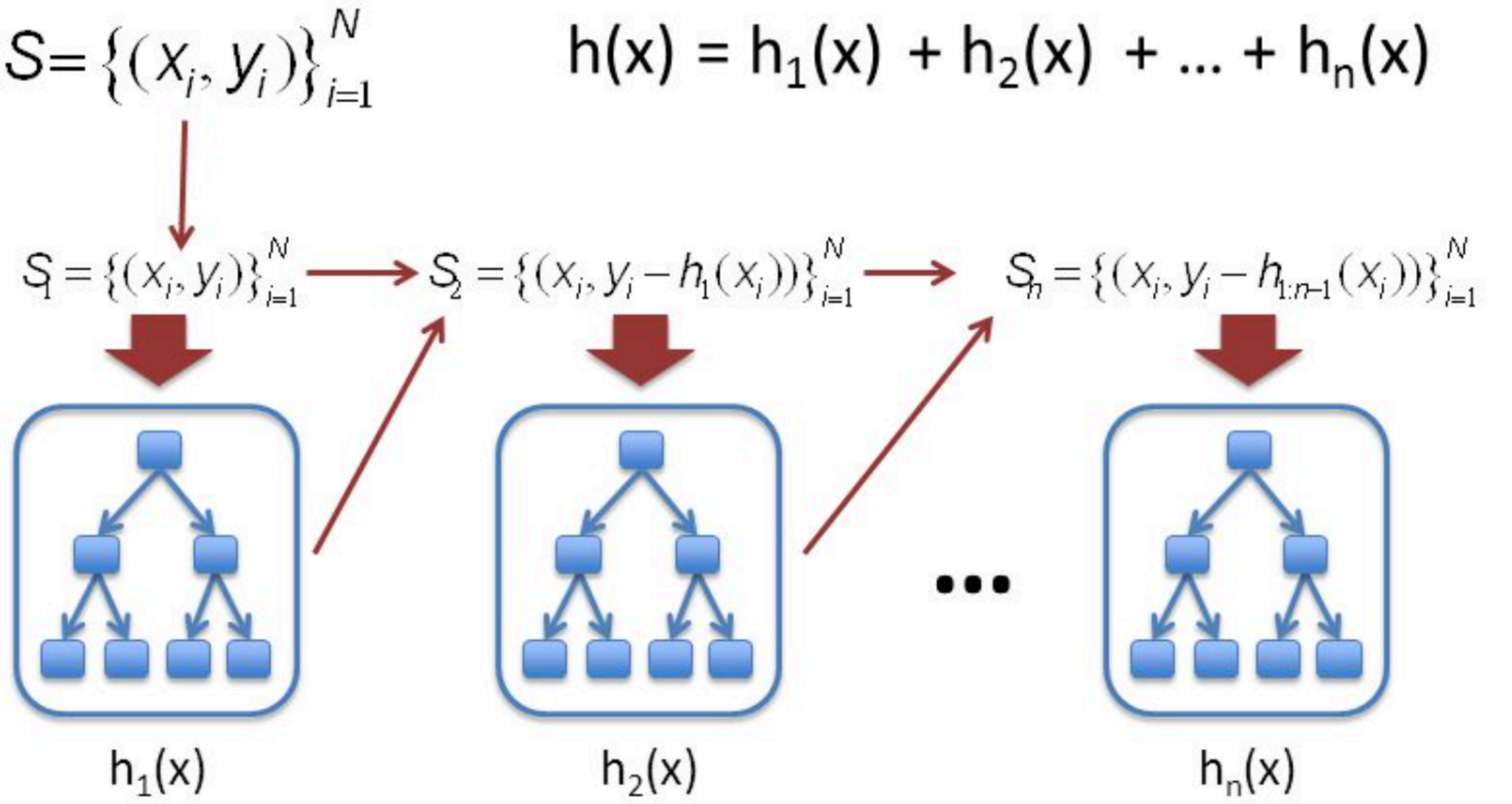
1. Генерируем M выборок на основе имеющейся (по схеме выбора с возвращением)

2. Строим на них деревья с рандомизированными разбиениями в узлах: выбираем k случайных признаков и ищем наиболее информативное разбиение по ним

3. При прогнозе усредняем ответ всех деревьев

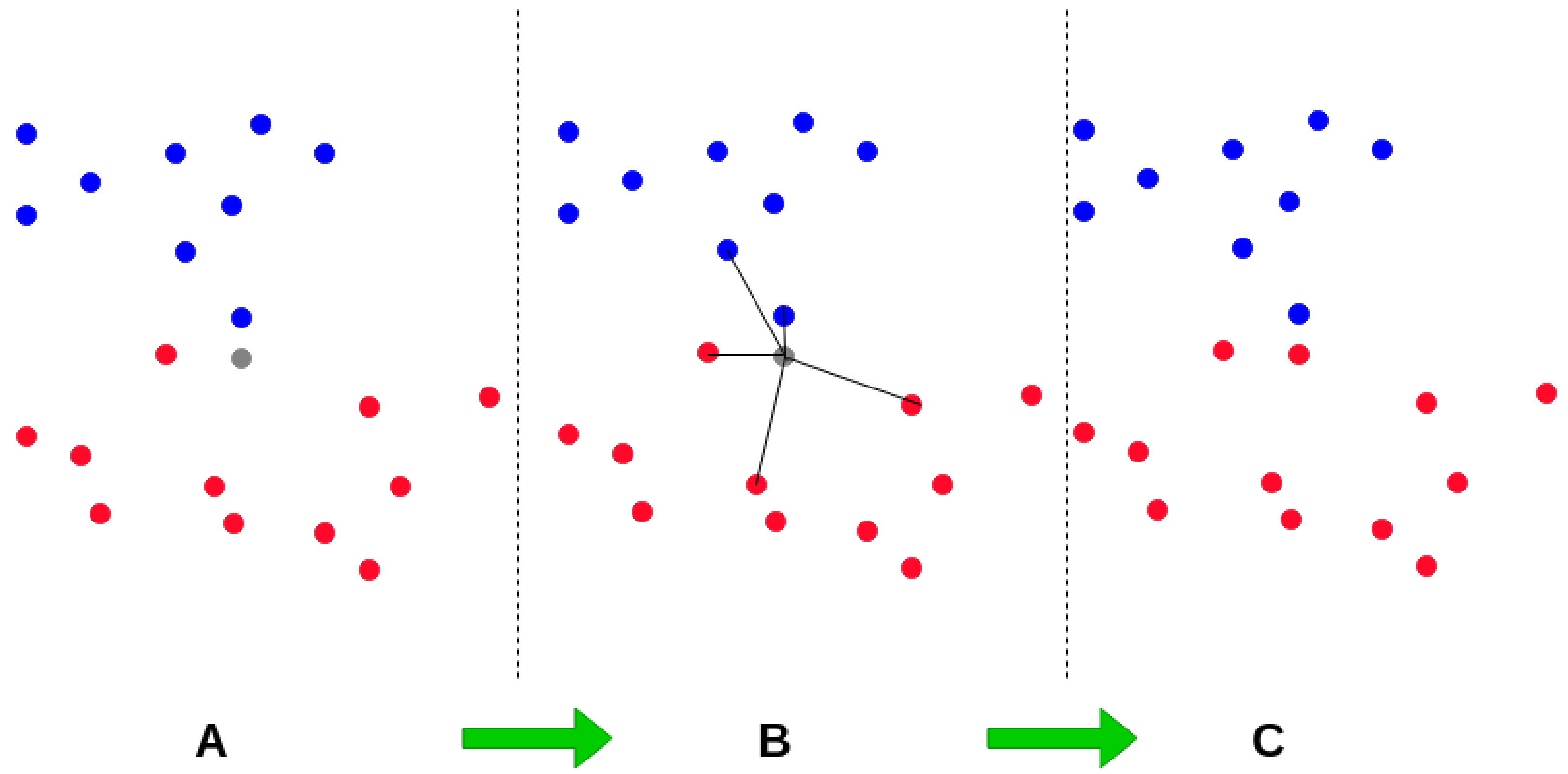


Идея Gradient Boosted Decision Trees (GBDT)



III. kNN

kNN - классификатор



IV. SVM

SVM - классификатор

S – Support
V – Vector
M –

SVM - классификатор

S – Support
V – Vector
M – Магия

SVM - классификатор

S – Support
V – Vector
M – ~~Магия~~ Machine

Точно будет решение.
Так или иначе.
Причём само.

SVM - классификатор

S – Support
V – Vector (геометрия?)
M – ~~Магия~~ Machine

Точно будет решение.
Так или иначе.
Причём само.

SVM - классификатор

S – Support
V – Vector (геометрия?)
M – ~~Магия~~ Machine

$$w_1x_1 + w_2x_2 + \ldots + w_nx_n + w_0 = 0$$

разделяющая гиперплоскость

Точно будет решение.
Так или иначе.
Причём само.

SVM - классификатор

S – Support
V – Vector (геометрия?)
M – ~~Магия~~ Machine

$$w_1x_1 + w_2x_2 + \ldots + w_nx_n + w_0 = 0$$

разделяющая гиперплоскость

$$a(x) = \mathrm{sign}\left(\sum_{j=1}^{n} w_j x^j - w_0\right) = \mathrm{sign}(\langle w, x\rangle - w_0)$$

Точно будет решение.
Так или иначе.
Причём само.

SVM - классификатор

S – Support (???)
V – Vector (геометрия?)
M – ~~Магия~~ Machine

$$w_1x_1 + w_2x_2 + \ldots + w_nx_n + w_0 = 0$$

разделяющая гиперплоскость

$$a(x) = \operatorname{sign}\left(\sum_{j=1}^{n} w_j x^j - w_0\right) = \operatorname{sign}(\langle w, x\rangle - w_0)$$

Точно будет решение.
Так или иначе.
Причём само.

SVM - классификатор

$$Q(w, w_0) = \sum_{i=1}^{\ell} \left[y_i(\langle w, x_i \rangle - w_0) < 0 \right]$$

S – Support (???)
V – Vector (геометрия?)
M – ~~Магия~~ Machine

$$w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + w_0 = 0$$

разделяющая гиперплоскость

$$a(x) = \operatorname{sign}\left(\sum_{j=1}^{n} w_j x^j - w_0 \right) = \operatorname{sign}(\langle w, x \rangle - w_0)$$

Точно будет решение.
Так или иначе.
Причём само.

SVM - классификатор

$$Q(w, w_0) = \sum_{i=1}^{\ell} [y_i(\langle w, x_i \rangle - w_0) < 0]$$

S – Support (???)
V – Vector (геометрия?)
M – ~~Магия~~ Machine

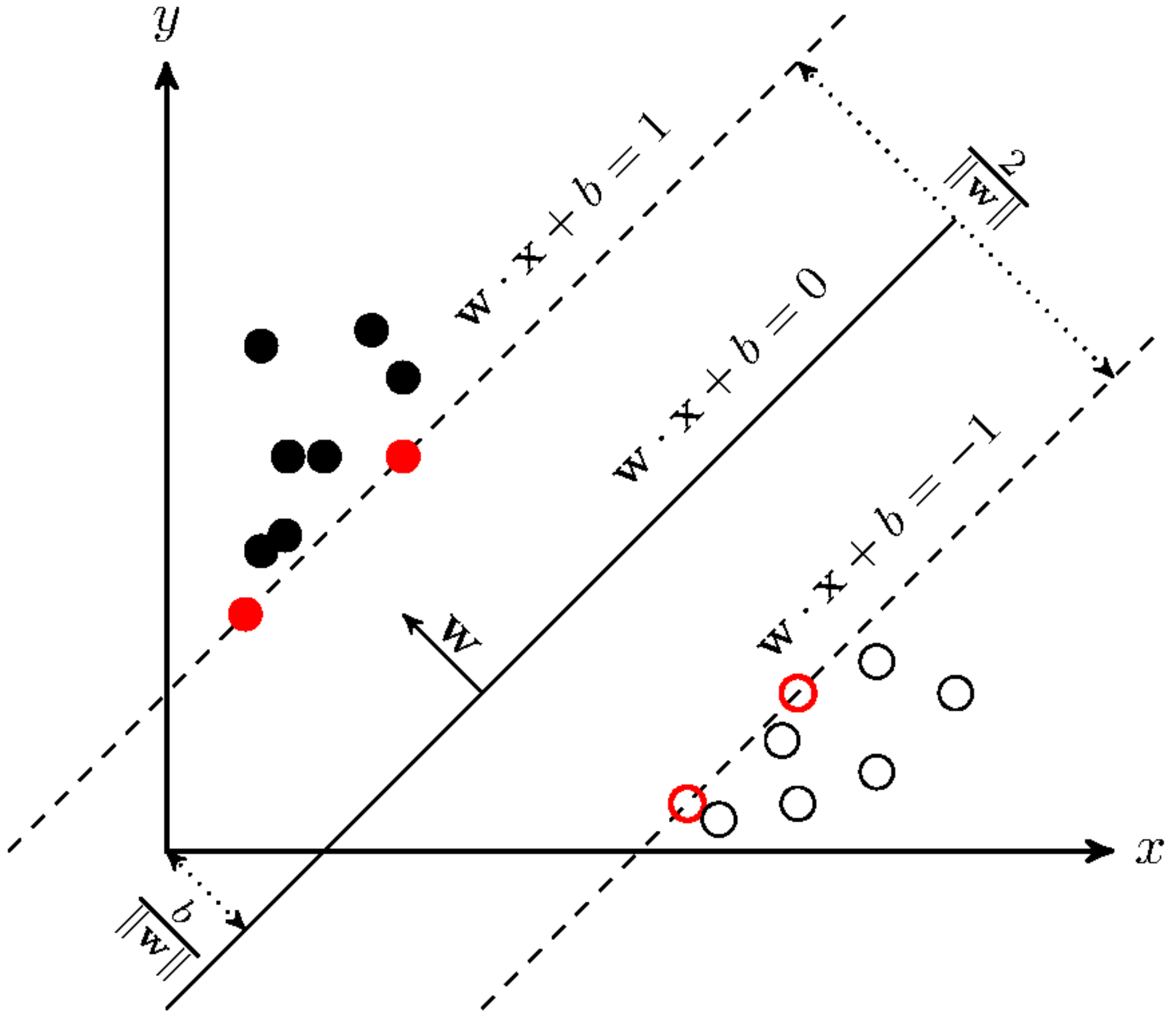
$$w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + w_0 = 0$$

разделяющая гиперплоскость

$$a(x) = \operatorname{sign}\left(\sum_{j=1}^{n} w_j x^j - w_0\right) = \operatorname{sign}(\langle w, x \rangle - w_0)$$

Точно будет решение.
Так или иначе.
Причём само.

SVM - классификатор



SVM - классификатор

$$\left\langle (x_+ - x_-), \frac{w}{\|w\|} \right\rangle = \frac{\langle w, x_+\rangle - \langle w, x_-\rangle}{\|w\|} = \frac{(w_0 + 1) - (w_0 - 1)}{\|w\|} = \frac{2}{\|w\|}$$

SVM - классификатор

$$\left\langle (x_+ - x_-), \frac{w}{\|w\|} \right\rangle = \frac{\langle w, x_+\rangle - \langle w, x_-\rangle}{\|w\|} = \frac{(w_0 + 1) - (w_0 - 1)}{\|w\|} = \frac{2}{\|w\|}$$

$$\langle w, x_i\rangle - w_0 \begin{cases} \leqslant -1, & \text{если } y_i = -1; \\ \geqslant 1, & \text{если } y_i = +1; \end{cases}$$

SVM - классификатор

$$\left\langle (x_+ - x_-), \frac{w}{\|w\|} \right\rangle = \frac{\langle w, x_+\rangle - \langle w, x_-\rangle}{\|w\|} = \frac{(w_0 + 1) - (w_0 - 1)}{\|w\|} = \frac{2}{\|w\|}$$

$$\begin{cases} \langle w, w\rangle \to \min; \\ y_i\big(\langle w, x_i\rangle - w_0\big) \geqslant 1, \quad i = 1, \dots, \ell. \end{cases}$$

$$\langle w, x_i\rangle - w_0 \begin{cases} \leqslant -1, & \text{если } y_i = -1; \\ \geqslant 1, & \text{если } y_i = +1; \end{cases}$$

SVM - классификатор

$$\begin{cases} \langle w, w \rangle \to \min; \\ y_i\big(\langle w, x_i \rangle - w_0\big) \geqslant 1, \quad i = 1, \ldots, \ell. \end{cases}$$

SVM - классификатор

$$\begin{cases} \langle w, w \rangle \to \min; \\ y_i\big(\langle w, x_i \rangle - w_0\big) \geqslant 1, \quad i = 1, \ldots, \ell. \end{cases}$$

Теорема Куна-Таккера

SVM - классификатор

$$\begin{cases} \langle w, w \rangle \to \min; \\ y_i\big(\langle w, x_i \rangle - w_0\big) \geqslant 1, \quad i = 1, \dots, \ell. \end{cases}$$

$+$ Теорема Куна-Таккера

SVM - классификатор

$$\begin{cases} \langle w, w \rangle \to \min; \\ y_i\big(\langle w, x_i \rangle - w_0\big) \geqslant 1, \quad i = 1, \dots, \ell. \end{cases}$$

+ Теорема Куна-Таккера

=

$$\begin{cases} \mathscr{L}(w, w_0; \lambda) = \frac{1}{2}\langle w, w \rangle - \sum\limits_{i=1}^{\ell} \lambda_i \Big(y_i\big(\langle w, x_i \rangle - w_0\big) - 1 \Big) \to \min\limits_{w, w_0} \max\limits_{\lambda}; \\ \lambda_i \geqslant 0, \quad i = 1, \dots, \ell; \\ \lambda_i = 0, \quad \text{либо} \quad \langle w, x_i \rangle - w_0 = y_i, \quad i = 1, \dots, \ell; \end{cases}$$

двойственная задача поиска седловой точки функции Лагранжа

SVM - классификатор

$$\begin{cases} -\mathscr{L}(\lambda) = -\sum\limits_{i=1}^{\ell} \lambda_i + \frac{1}{2} \sum\limits_{i=1}^{\ell} \sum\limits_{j=1}^{\ell} \lambda_i \lambda_j y_i y_j (\langle x_i, x_j \rangle) \;\to\; \min\limits_{\lambda}; \\ \lambda_i \geqslant 0, \quad i = 1, \dots, \ell; \\ \sum\limits_{i=1}^{\ell} \lambda_i y_i = 0. \end{cases}$$

SVM - классификатор

$$\begin{cases} -\mathscr{L}(\lambda) = -\sum\limits_{i=1}^{\ell} \lambda_i + \frac{1}{2} \sum\limits_{i=1}^{\ell} \sum\limits_{j=1}^{\ell} \lambda_i \lambda_j y_i y_j (\langle x_i, x_j \rangle) \ \to \ \min\limits_{\lambda}; \\ \lambda_i \geqslant 0, \quad i = 1, \ldots, \ell; \\ \sum\limits_{i=1}^{\ell} \lambda_i y_i = 0. \end{cases}$$

$$w = \sum_{i=1}^{\ell} \lambda_i y_i x_i$$

SVM - классификатор

$$\begin{cases} -\mathscr{L}(\lambda) = -\sum\limits_{i=1}^{\ell} \lambda_i + \frac{1}{2} \sum\limits_{i=1}^{\ell} \sum\limits_{j=1}^{\ell} \lambda_i \lambda_j y_i y_j (\langle x_i, x_j \rangle) \;\to\; \min\limits_{\lambda}; \\ \lambda_i \geqslant 0, \quad i = 1, \ldots, \ell; \\ \sum\limits_{i=1}^{\ell} \lambda_i y_i = 0. \end{cases}$$

$$w = \sum_{i=1}^{\ell} \lambda_i y_i x_i$$

$$a(x) = \operatorname{sign}\left(\sum_{i=1}^{\ell} \lambda_i y_i \langle x_i, x \rangle - w_0 \right)$$

SVM - классификатор

$$\begin{cases} \frac{1}{2}\langle w, w\rangle + C\sum_{i=1}^{\ell}\xi_i \to \min\limits_{w, w_0, \xi}; \\ y_i\big(\langle w, x_i\rangle - w_0\big) \geqslant 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \xi_i \geqslant 0, \quad i = 1, \dots, \ell. \end{cases}$$

SVM - классификатор

$$\begin{cases} \frac{1}{2}\langle w, w\rangle + C\sum_{i=1}^{\ell}\xi_i \to \min\limits_{w, w_0, \xi}; \\ y_i\big(\langle w, x_i\rangle - w_0\big) \geqslant 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \xi_i \geqslant 0, \quad i = 1, \dots, \ell. \end{cases}$$

$$\begin{cases} -\mathscr{L}(\lambda) = -\sum_{i=1}^{\ell}\lambda_i + \frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\lambda_i\lambda_j y_i y_j \langle x_i, x_j\rangle \;\to\; \min\limits_{\lambda}; \\ 0 \leqslant \lambda_i \leqslant C, \quad i = 1, \dots, \ell; \\ \sum_{i=1}^{\ell}\lambda_i y_i = 0. \end{cases}$$

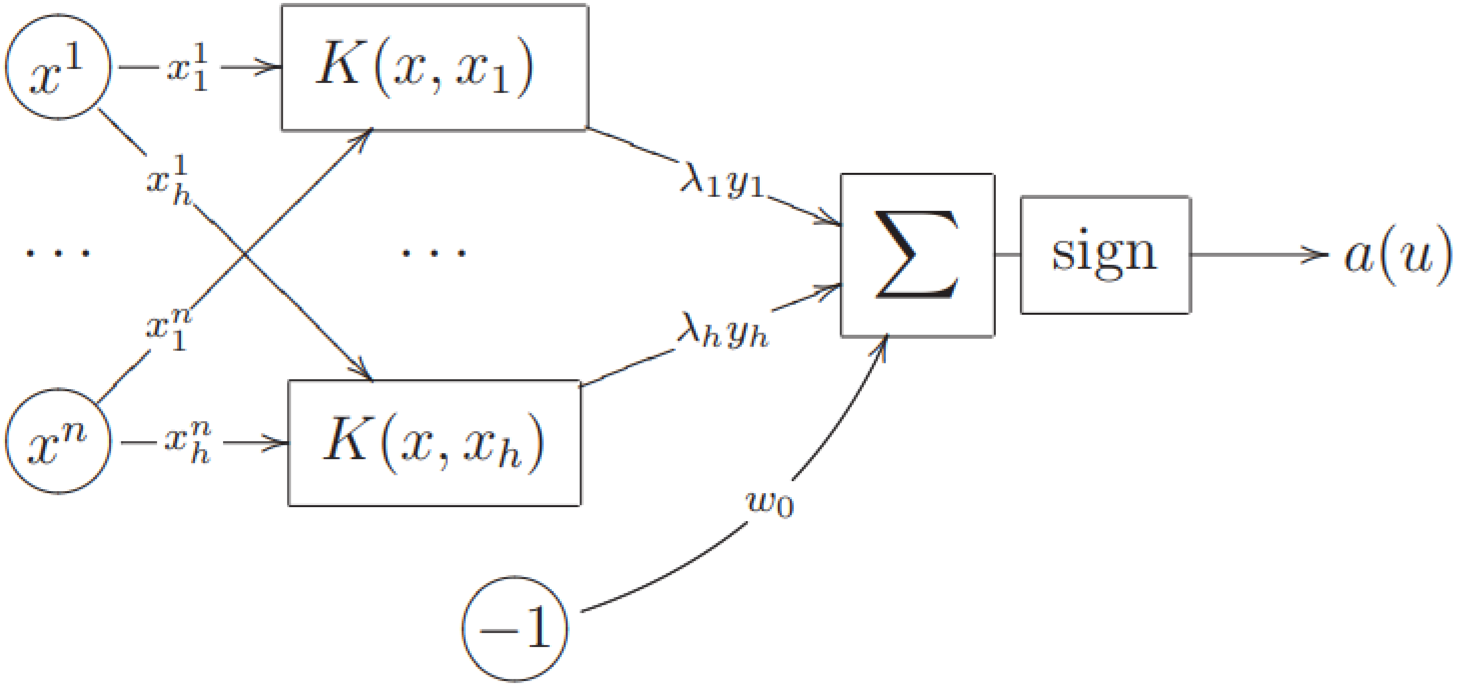
SVM - классификатор

$$\begin{cases} \frac{1}{2}\langle w, w\rangle + C\sum_{i=1}^{\ell}\xi_i \to \min\limits_{w,w_0,\xi}; \\ y_i\big(\langle w, x_i\rangle - w_0\big) \geqslant 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \xi_i \geqslant 0, \quad i = 1, \dots, \ell. \end{cases}$$

$$\begin{cases} -\mathscr{L}(\lambda) = -\sum_{i=1}^{\ell}\lambda_i + \frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\lambda_i\lambda_j y_i y_j \langle x_i, x_j\rangle \ \to \ \min\limits_{\lambda}; \\ 0 \leqslant \lambda_i \leqslant C, \quad i = 1, \dots, \ell; \\ \sum_{i=1}^{\ell}\lambda_i y_i = 0. \end{cases}$$

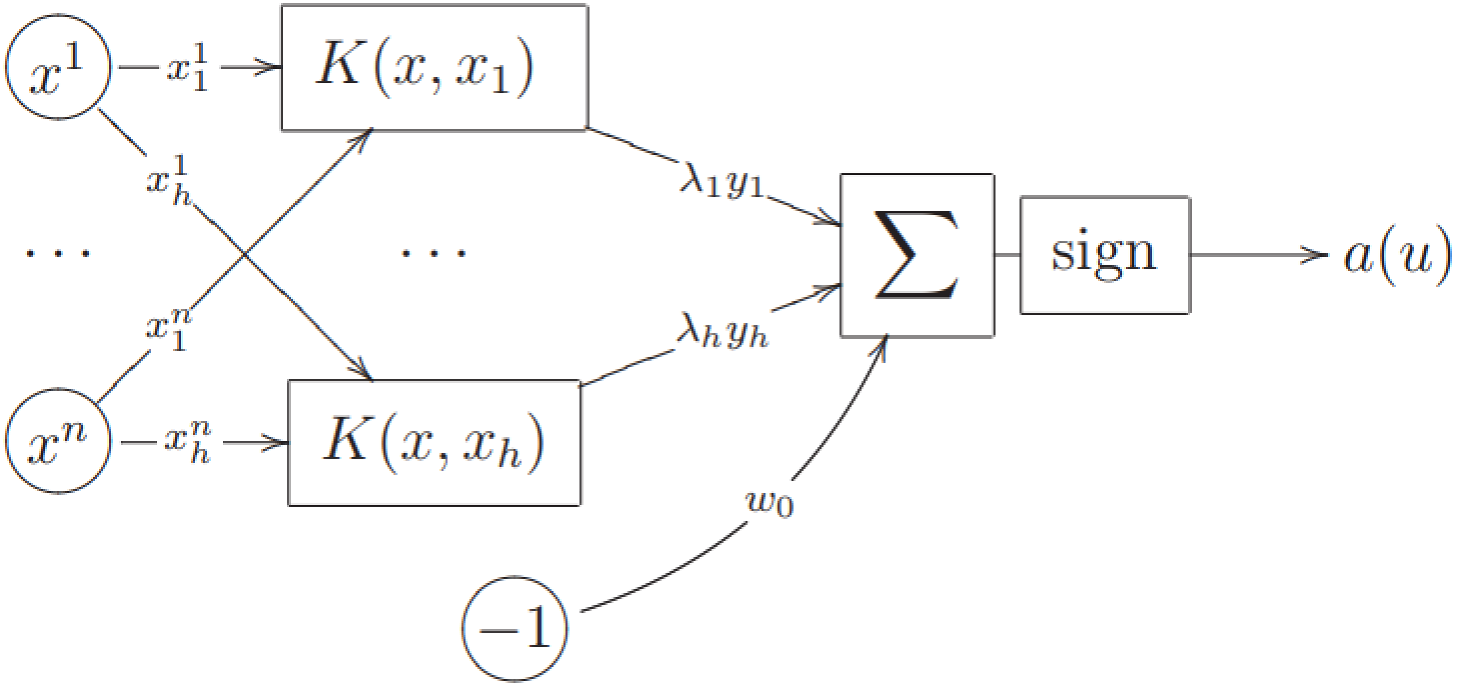
$$a(x) = \mathrm{sign}\left(\sum_{i=1}^{\ell}\lambda_i y_i \langle x_i, x\rangle - w_0\right)$$

SVM - классификатор

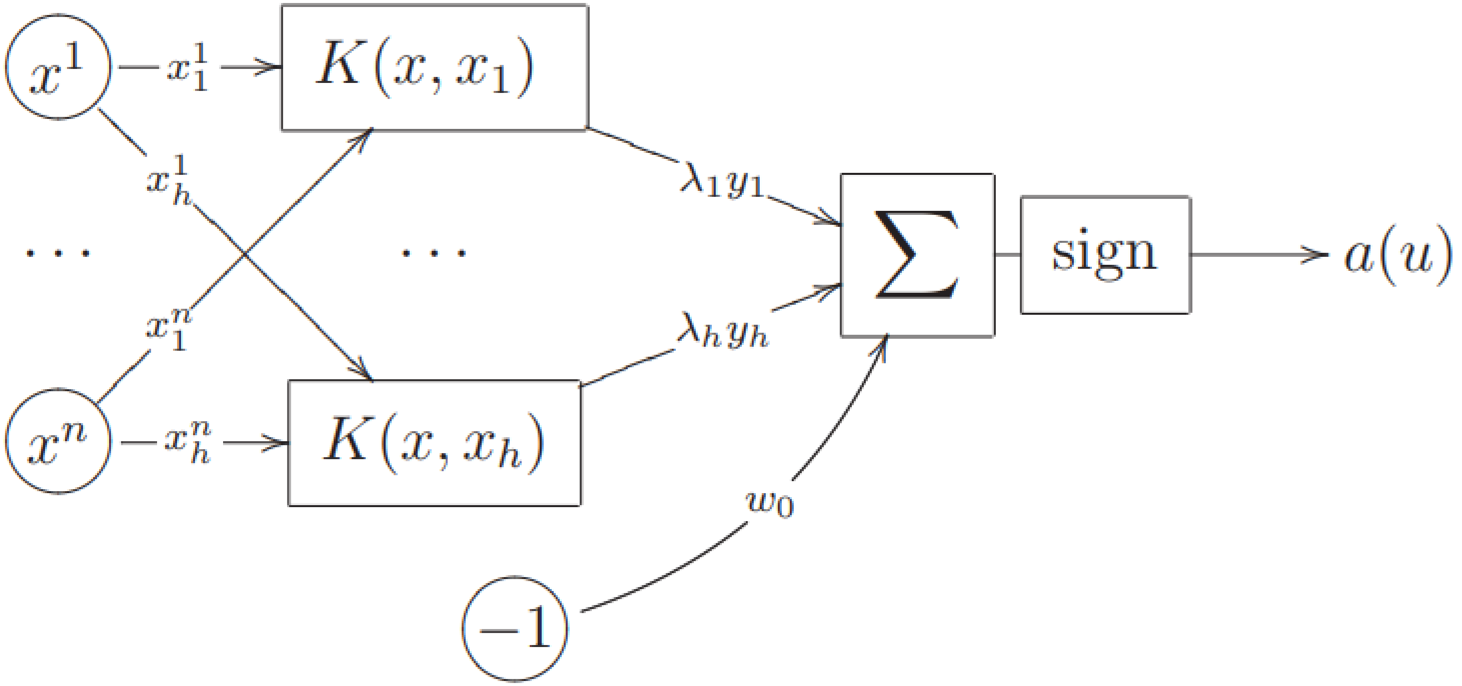


$$a(x) = \operatorname{sign}\left(\sum_{i=1}^{h} \lambda_i y_i K(x_i, x) - w_0\right)$$

SVM - классификатор



SVM - классификатор



$$a(x) = \mathrm{sign}\left(\sum_{i=1}^{h} \lambda_i y_i K(x_i, x) - w_0\right)$$