

На стадии обучения:

- $\beta_v(x)$ не определено $\Rightarrow x_i$ исключается из U для $I(\beta, U)$
- $q_v = \frac{|U_0|}{|U|}$ — оценка вероятности левой ветви, $\forall v \in V_{\text{внутр}}$
- $P(y|x, v) = \frac{1}{|U|} \#\{x_i \in U: y_i = y\}$ для всех $v \in V_{\text{лист}}$

На стадии классификации:

- $\beta_v(x)$ не определено \Rightarrow пропорциональное распределение:

$$P(y|x, v) = q_v P(y|x, L_v) + (1 - q_v) P(y|x, R_v).$$

- $\beta_v(x)$ определено \Rightarrow либо налево, либо направо:

$$P(y|x, v) = (1 - \beta_v(x)) P(y|x, L_v) + \beta_v(x) P(y|x, R_v).$$

- Окончательное решение — наиболее вероятный класс:

$$a(x) = \arg \max_{y \in Y} P(y|x, v_0).$$

Достоинства:

- Интерпретируемость и простота классификации.
- Гибкость: можно варьировать множество \mathcal{B} .
- Допустимы разнотипные данные и данные с пропусками.
- Трудоёмкость линейна по длине выборки $O(|\mathcal{B}|hl)$.
- Не бывает отказов от классификации.

Недостатки:

- Жадный ID3 переусложняет структуру дерева, и, как следствие, сильно переобучается.
- Фрагментация выборки: чем дальше v от корня, тем меньше статистическая надёжность выбора β_v, c_v .
- Высокая чувствительность к шуму, к составу выборки, к критерию информативности.