

- 1: **ПРОЦЕДУРА** LearnID3 ( $U \subseteq X^\ell$ );
- 2: **если** все объекты из  $U$  лежат в одном классе  $c \in Y$  **то**
- 3:     **вернуть** новый лист  $v$ ,  $c_v := c$ ;
- 4: **найти предикат с максимальной информативностью:**  
    $\beta := \arg \max_{\beta \in \mathcal{B}} I(\beta, U)$ ;
- 5: разбить выборку на две части  $U = U_0 \sqcup U_1$  по предикату  $\beta$ :  
    $U_0 := \{x \in U : \beta(x) = 0\}$ ;  
    $U_1 := \{x \in U : \beta(x) = 1\}$ ;
- 6: **если**  $U_0 = \emptyset$  или  $U_1 = \emptyset$  **то**
- 7:     **вернуть** новый лист  $v$ ,  $c_v := \text{Мажоритарный класс}(U)$ ;
- 8: создать новую внутреннюю вершину  $v$ :  $\beta_v := \beta$ ;  
   построить левое поддерево:  $L_v := \text{LearnID3}(U_0)$ ;  
   построить правое поддерево:  $R_v := \text{LearnID3}(U_1)$ ;
- 9: **вернуть**  $v$ ;

## 1. Критерий Джини:

$$I(\beta, X^\ell) = \#\{(x_i, x_j): y_i = y_j \text{ и } \beta(x_i) \neq \beta(x_j)\}.$$

## 2. $D$ -критерий В.И.Донского:

$$I(\beta, X^\ell) = \#\{(x_i, x_j): y_i \neq y_j \text{ и } \beta(x_i) = \beta(x_j)\}.$$

## 3. Энтропийный критерий:

$$I(\beta, X^\ell) = \sum_{c \in Y} h\left(\frac{P_c}{\ell}\right) - \frac{p}{\ell} h\left(\frac{p_c}{p}\right) - \frac{\ell - p}{\ell} h\left(\frac{P_c - p_c}{\ell - p}\right),$$

где  $h(z) \equiv -z \log_2 z$ ,

$$P_c(X^\ell) = \#\{x_i: y_i = c\},$$

$$p_c(X^\ell) = \#\{x_i: y_i = c \text{ и } \beta(x_i) = 1\},$$

$$p(X^\ell) = \#\{x_i: \beta(x_i) = 1\}.$$