

**МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И.УЛЬЯНОВА (ЛЕНИНА)
Кафедра МО ЭВМ**

**ОТЧЁТ
по практической работе №7
по дисциплине «Машинное обучение»
Тема: Кластеризация**

Студент гр. 6304

Преподаватель

Корытов П.В.

Жангиров Т.Р.

Санкт-Петербург

2020

1. Задание 1

Таблица 1. Q2

x_i	a_1	a_2	a_3	Class
x_1	T	T	5.0	Y
x_2	T	T	7.0	Y
x_3	T	F	8.0	N
x_4	F	F	3.0	Y
x_5	F	T	7.0	N
x_6	F	T	4.0	N
x_7	F	F	5.0	N
x_8	T	F	6.0	Y
x_9	F	T	1.0	N

Точка $x = (T, F, 10)$.

$$\hat{P}(c_Y) = \frac{n_Y}{n} = \frac{4}{9} \quad \hat{P}(c_N) = \frac{n_N}{n} = \frac{5}{9}$$

$$\mu_{3Y} = 5.25, \sigma_{3Y}^2 = 2.1875$$

$$\mu_{3Y} = 5.0, \sigma_{3Y}^2 = 6.0$$

$$P(x|c_Y) = P(T|c_Y) \cdot P(F|c_Y) \cdot P(10|c_Y) \approx \frac{3}{4} \cdot \frac{2}{4} \cdot 0.00155 = 0.00058125$$

$$P(x|c_N) = P(T|c_N) \cdot P(F|c_N) \cdot P(10|c_N) \approx \frac{1}{5} \cdot \frac{2}{5} \cdot 0.02027 = 0.00162160$$

$$P(c_Y|x) \propto P(x|c_Y) \cdot P(C_y) \approx 0.00025$$

$$P(c_N|x) \propto P(x|c_N) \cdot P(C_N) \approx 0.00090$$

$$\hat{y} \leftarrow \arg \max_{c_i} \{P(x|c_i) \cdot P(C_i)\} = N$$

2. Задание 2

$$\hat{\mu}_1 = (1, 3) \quad \hat{\mu}_2 = (5, 5)$$

$$\hat{\Sigma}_1 = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix} \quad \hat{\Sigma}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\hat{P}(C_1) = \hat{P}(C_2) = 0.5$$

$$\hat{P}(c_1|x) \propto \hat{f}(x|\mu_1, \hat{\Sigma}_1) \hat{P}(C_1) \approx 0.096 \cdot 0.5 = 0.048$$

$$\hat{P}(c_2|x) \propto \hat{f}(x|\mu_2, \hat{\Sigma}_2) \hat{P}(C_2) \approx 0.025 \cdot 0.5 = 0.0125$$

$$\hat{y} \leftarrow \arg \max_{c_i} \{f(x|\hat{\mu}_i, \hat{\Sigma}_i) \cdot P(C_i)\} = 1$$

3. Задание 3

Таблица 2. Q2

Point	Age	Car	Risk
x_1	25	Sports	L
x_2	20	Vintage	H
x_3	25	Sports	L
x_4	45	SUV	H
x_5	20	Sports	H
x_6	25	SUV	H

$$\hat{P}(C_L) = \frac{1}{3} \quad \hat{P}(C_H) = \frac{2}{3}$$

$$H(D) = -\left(\frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3}\right) \approx 0.918$$

- Проверка первого атрибута $Age = [20, 20, 25, 25, 25, 45]$

$$M = \{22.5, 35\}$$

$$N_{22.5,L} = 0 \quad N_{22.5,H} = 2$$

$$N_{35,L} = 2 \quad N_{35,H} = 3$$

$$score(Age < 22.5) \leftarrow Gain(D, D_y, D_n) = \infty.$$

Получено максимальное значение $Gain$, значит первое разбиение выполняется по $Age < 22.5$ и на этой ветви будет Н.

- На альтернативной ветви остаток набора данных имеет следующий вид:

Таблица 3. Набор данных на $\text{Age} \geq 22.5$

Point	Age	Car	Risk
x_1	25	Sports	L
x_3	25	Sports	L
x_4	45	SUV	H
x_6	25	SUV	H

Похожим образом максимальное значение Gain будет получено для $\text{Car} = \text{Sports}$, т.к. $P(\text{Car} = \text{Sports} | L) = 1$.

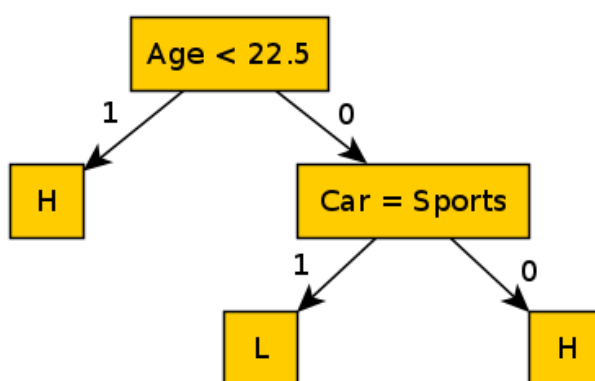


Рисунок 1 – Полученное дерево

Классификация точки ($\text{Age} = 27$, $\text{Car} = \text{Vintage}$):

- $\text{Age} < 22.5$ — 0
- $\text{Car} = \text{Sports}$ — 0
- Результат — H