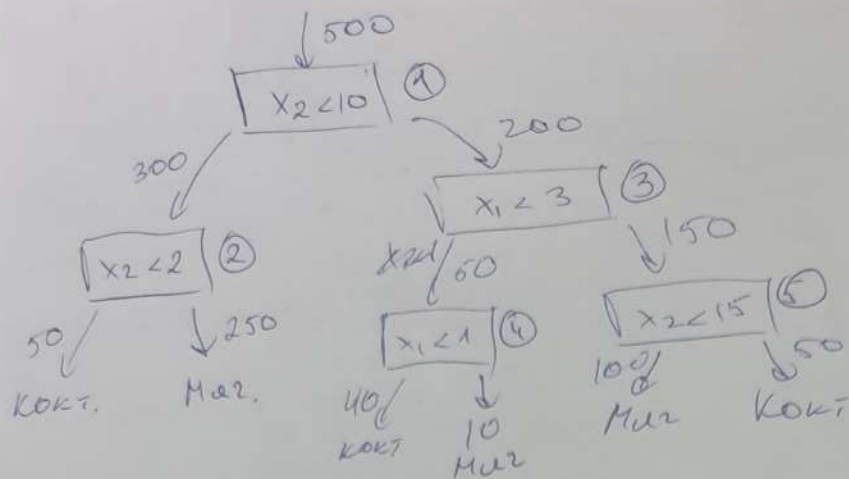


10)

Артемов
Александр
4 вёр

Зная конечные кол-во коктейлей и маз (Пусть
Коктейль - А
Маз - В)
Можем узнать кол-во эл-тов А и В на
каждом шаге. Если мы решим строить в качестве
критерия информативности - энтропийный критерий,
то $H = -\sum p_k \log p_k$. Всего А - 360 и В - 140
Посчитаем $H_1 = -\frac{360}{500} \log \frac{360}{500} - \frac{140}{500} \log \frac{140}{500} = 0,2575$

$$H_2 = -\frac{50}{300} \log \frac{5}{30} - \frac{250}{300} \log \frac{25}{30} = 0,1957$$

$$H_3 = -\frac{1}{4} \log \frac{1}{4} - \frac{3}{4} \log \frac{3}{4} = 0,2442$$

$$H_4 = -\frac{4}{5} \log \frac{4}{5} - \frac{1}{5} \log \frac{1}{5} = 0,2173$$

$$H_5 = -\frac{2}{3} \log \frac{2}{3} - \frac{1}{3} \log \frac{1}{3} = 0,2764$$

Далее найдем ΔH_i : $\Delta H_2 = H_2$ & $\Delta H_5 = H_5$
 $\Delta H_4 = H_4$ т.к. дальше от
не разделяются

$$\Delta H_1 = H_1 - \frac{300}{500} H_2 - \frac{200}{500} H_3 = 0,0424$$

$$\Delta H_3 = H_3 - \frac{1}{4} H_4 - \frac{3}{4} H_5 = -0,017425$$

Полночим ΔH_i на $\frac{n}{N} \rightarrow \Pi \Delta H_i = \Delta H_1$

$$\sum \Pi_{призн} = \Delta H_1 + \Delta H_2 + \Delta H_5 = 0,24274$$

$$\sum \Pi_{призн} = \Delta H_3 + \Delta H_4 = 0,01476$$

$$\Pi \Delta H_2 = \Delta H_2 \cdot \frac{3}{5} = 0,11742$$

$$\Pi \Delta H_3 = \Delta H_3 \cdot \frac{2}{5} = -0,00697$$

$$\Pi \Delta H_4 = \Delta H_4 \cdot \frac{1}{10} = 0,02173$$

$$\Pi \Delta H_5 = \Delta H_5 \cdot \frac{3}{10} = 0,08292$$

1

Максим образцы влияния 2 признака: $\frac{\sum H_2^2}{\sum H_1^2} = 0,826$
 Влияние 1 признака: $\frac{\sum H_1^2}{\sum H_2^2} = 0,074$

9)

Доход	Доход _{кв}	Потр	Потр _{кв}	$P(X=x, Y=y)$	$\sum H_j$
50	0	48	0	$\frac{1}{4} = \frac{1}{16}$	
200	3	150	1	$\frac{1}{6} = \frac{1}{36}$	
70	1	310	1	$\frac{1}{6} = \frac{1}{36}$	
30	0	30	0	$\frac{1}{4} = \frac{1}{16}$	
25	0	35	0	$\frac{1}{6} = \frac{1}{36}$	
120	2	150	1	$\frac{1}{12}$	
150	3	80	0	$\frac{1}{6} = \frac{1}{36}$	
110	2	110	1	$\frac{1}{6} = \frac{1}{36}$	
85	1	95	1	$\frac{1}{6} = \frac{1}{36}$	
150	3	130	0	$\frac{1}{12}$	
130	2	50	0	$\frac{1}{12}$	
80	1	60	0	$\frac{1}{12}$	
	q_1	q_2	q_3		

Доход) 25 30 50 70 80 85 110 130 150 150 200
 $q_1 = \frac{13}{4} = 3,25$ $q_3 = 8,75$

$q_2 = \frac{13}{2} = 6,5$

Потр) 30 35 48 50 60 80 95 110 130 150 150 310
 медиана

T_{xy}	P
Доход 0	$\frac{1}{4}$
Доход 1	$\frac{1}{4}$
Доход 2	$\frac{1}{4}$
Доход 3	$\frac{1}{4}$
Потр 0	$\frac{1}{2}$
Потр 1	$\frac{1}{2}$

$$MI = \sum_{x \in X} \sum_{y \in Y} P(X=x, Y=y) \frac{\log \left(\frac{P(X=x, Y=y)}{P(X=x) \cdot P(Y=y)} \right)}{\log 8}$$

$$MI = 3 \cdot P(X=0, Y=0) \log_8 P(X=0, Y=0) + 2 P(X=3, Y=1) \log_8 P(X=3, Y=1) + 2 P(1, 1) \log_8 P(1, 1) + 2 P(2, 1) \log_8 P(2, 1) + P(3, 0) \log_8 P(3, 0) + P(2, 0) \log_8 P(2, 0) + P(1, 0) \log_8 P(1, 0) =$$

$$= \frac{3}{4} \log 6 + \frac{1}{3} \log \frac{4}{3} + \frac{1}{3} \log \frac{4}{3} + \frac{1}{3} \log \frac{4}{3} + \frac{3}{12} \log \frac{2}{3} =$$

$$= \frac{3}{4} \log 6 + \log \frac{4}{3} + \frac{1}{4} \log \frac{2}{3} = 0,665$$

хорошая средняя сила
 взаимосвязи признаков
 между доходами и потреблением

Артёмов
Александр
4 вер

8) а) Ит.е в данном случае мы хотим найти метрики для выражения $AB \rightarrow C$ где A - добавление товаров в корзину = 1
 B - создание акк = 0
 C - совершение покупки = 0

$Support(AB \rightarrow C) = \frac{N(AB \cup C)}{N} = \frac{2}{9}$ ← всего строк

$Confidence(AB \rightarrow C) = \frac{S(AB \rightarrow C)}{S(AB)} = \frac{N(AB \rightarrow C)}{N(AB)} = \frac{2}{4} = \frac{1}{2}$

$lift(AB \rightarrow C) = \frac{S(AB \rightarrow C)}{S(AB)S(C)} = \frac{2/9}{4/9 \cdot 5/9} = \frac{18}{20} = \frac{9}{10}$ ← элементы 1, 3, 4, 5

$Conviction(AB \rightarrow C) = \frac{1 - S(C)}{1 - Conf(AB \rightarrow C)} = \frac{5/9}{1/2} = \frac{10}{9}$

$Support = \frac{2}{9}$ - слабая связь для правила. $C = \frac{1}{2}$ - т.е. половина наблюдений из $A=1, B=0$ уровн. этому правилу
 $lift = \frac{9}{10} > 1 \Rightarrow AB$ + влияет на C и сила правила = $\frac{1}{8}$
 $Conviction > 1$

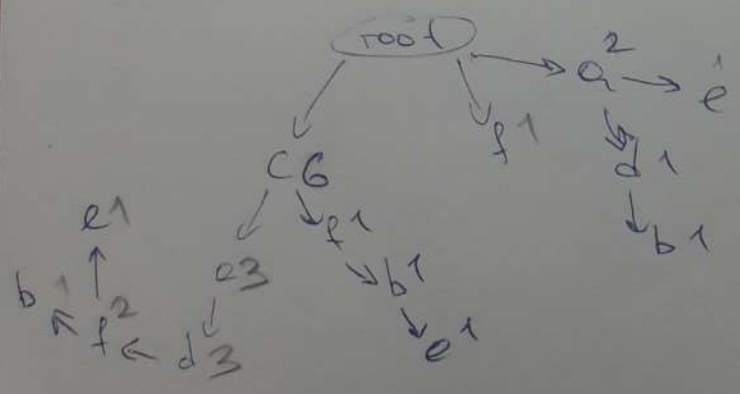
б) Пусть кризиски по порядку соответственно a, b, c, d, e, f
Получаем каждый кризисок

- Итого
- 1 c
 - 2 ae
 - 3 acd
 - 4 abcd f
 - 5 c
 - 6 f
 - 7 acdef
 - 8 bce f
 - 9 abd

- Получаем каждый кризисок
- a(5)
 - b(3)
 - c(6)
 - d(4)
 - e(3)
 - f(4)

- Упорядочим наши набора
- 1) c
 - 2) ae
 - 3) c d
 - 4) c a d f b
 - 5) c
 - 6) f
 - 7) c a d f e
 - 8) c f b e
 - 9) a d b

Строим
FP-дерево



Она какого предмета ищем
 куда где префиксы $(c, b) \checkmark \geq 3$

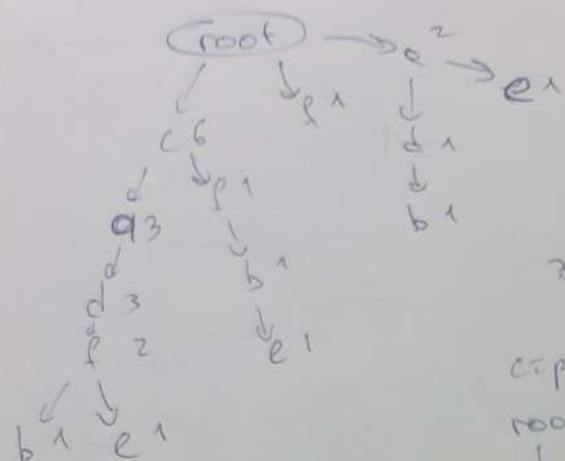
- 1) Она с только набор из c $\geq 3 \checkmark (c, d) \times \leq 3$
- 2) $a \rightarrow \{(c, a, 3)\}$ соответственно
 это и есть набор

- 3) $b \rightarrow \{(c, d, b, 1), (c, f, b, 1), (c, a, d, f, b, 1)\}$

сгруппи FP сверху

$$(c, f, b, 2) \leq 3 \times$$

$$(c, d, b, 2) \not\leq 3 \times$$



- 4) $d \rightarrow \{(c, a, d, 3), (c, d, 1)\}$

root

$$(c, a, d, 3) \checkmark \geq 3$$

3c
4a
4d

- 5) $e \rightarrow \{(c, f, b, e, 1), (c, a, d, f, e, 1), (a, e, 1)\}$

$$(c, f, e, 2) \not\leq 3 \times$$

$$(a, e, 2) \leq 3 \times$$

- 6) $f \rightarrow \{(c, a, d, f, 2), (c, f, 1), (f, 1)\}$

$$(c, f, 2) \leq 3 \times$$

III. 2 набора
 где которых support

$$\geq \frac{3}{10}$$

$$(c, b)$$

$$(c, a, 3)$$

$$(c, a, d, 3)$$

$$6) y = e^{\frac{-11xz - 5x - 19z - 22}{-2xz + x - 6z + 3}} = e^{\frac{-9xz - 6x - 13z - 25}{-2xz + x - 6z + 3}}$$

иначе уже нпн
если 2 не const

$\ln y = -9xz - 6x - 13z - 25$ и считаю что её
невозможно привести к нпн. Ввиду из-за сложного
хз, которое останется произведением где придем
к еЗ и еВ. Если мы хотим функции.

7) Мы получили 2 ГК то они соответствуют 4 канб.
т.е. $\mu_2 = 426,5$ Тогда они объясняют $\frac{\mu_2 + \mu_4}{\sum \mu_i} \cdot 100\%$
 $\mu_4 = 158,8$ дисперсии 92,7%

т.е. 2 и 4 сЗ их ГК объясняют
92,7% дисперсии.

2) Тогда полученные данные

$$\text{это } X = [V_2, V_4] = \begin{pmatrix} -25,84 & 4,36 \\ -9,09 & -6,26 \\ 1,13 & -1,75 \\ -3,48 & 0,36 \\ -7,87 & -3,78 \end{pmatrix}$$

$$1) R^2 = 1 - \frac{\text{нелл. дисп}}{\text{дисп}} = 1 - \frac{3}{25} = \frac{22}{25} = 0,88$$

б) Проверка значимости на уровне $\alpha = 0,2$

$$H_0: \beta_0 = \beta_1 = \dots = \beta_7 = 0 \quad F_{\text{рас}} = \frac{\frac{1}{k} \cdot R^2}{\frac{1}{n-k-1} (1-R^2)} = \frac{\frac{1}{7} \cdot \frac{22}{25}}{\frac{1}{79} \cdot \frac{3}{25}}$$

$$H_1: \exists \beta_i: \beta_i \neq 0$$

$$= \frac{22 \cdot 79}{21} = 82,761 >> F_{\text{крит}}(7, 79, 0,2) \Rightarrow H_0 \text{ отвергается}$$

\Rightarrow модель значима

2) У нас 3 выборки $n_1(5,8; 3,4; 3,5; 4,2)$
 $n_2(6,7; 6,2; 3,3; 3,5; 1,8; 4,8)$
 $n_3(5,7; 6,6; 7; 3,4; 4,8)$

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$$H_1: \exists i, j: \mu_i \neq \mu_j$$

$$\text{Если } H_0 \Rightarrow \frac{1}{k-1} \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2$$

$$F_{\text{рас}} = \frac{D_p}{D_{\text{вн}}} = \frac{1}{n-k} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$$

ср $F_{\text{крит}}$
1- α
5

$$F_{\text{roc}} = \frac{\frac{1}{2} \cdot (4(4,225 - 4,7)^2 + 6(4,383 - 4,7)^2 + 5(5,46 - 4,7)^2)}{\frac{1}{15-3} \sum_i \sum_j (x_{ij} - \bar{x}_j)^2}$$

$$\bar{x}_1 = 4,225$$

$$\bar{x}_2 = 4,383$$

$$\bar{x}_3 = 5,46$$

$$\bar{x} = 4,7$$

$$= \frac{1}{2} \cdot 2,196717$$

$$\frac{1}{12} (3,6875 + 17,4683 + 8,712)$$

$$F_{\text{roc}} = \frac{2,196717}{2,488} = 0,883$$

$$F_{\text{крит}}(2,14,0,2) \approx 1,8 > F_{\text{roc}} \Rightarrow \text{не отвергаем } H_0$$

5) Вероятно, построена модель с которой

$$b = (X^T X)^{-1} (X^T y). \text{ Почему в условии написано}$$

построено по 18 переменным а дано только 6?

$$\text{лейное } y_{\text{pred}} = b_0 + b_1 x_1 + \dots + b_n x_n$$

$$y_{\text{pred}} = \frac{1}{1 + e^{-y_{\text{pred}}}}$$

$$f_h = 0,5$$

	y=1	y=0
$\hat{y}=1$	3	2
$\hat{y}=0$	0	1

$$FPR = \frac{2}{3}$$

$$TPR = 1$$

$$AUC = \frac{1}{2} \cdot \frac{2}{3} \cdot 1 + \frac{1}{2} \cdot 2 \cdot \frac{1}{3} = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

$$\text{precision} = \frac{TP}{TP+FP} = \frac{3}{5}$$

$$\text{recall} = 1$$

$$F_{0,5} = \frac{5}{4} \cdot \frac{1 \cdot \frac{3}{5}}{\frac{1}{4} \cdot \frac{3}{5} + 1} = \frac{5}{4} \cdot \frac{3}{5} \cdot \frac{20}{23} = \frac{15}{23}$$

или котам

	0	0	0	1	1	1
y'	0,45	0,5	-2	0,3	0,4	4
$e^{-y'}$	0,63	0,60	7,38	0,70	0,67	0,18
	0,61	0,62	0,11	0,54	0,6	0,98
\hat{y}	1	1	0	1	1	1