

Wprowadzenie

Zbiór Nursery¹ i jego atrybuty:

parents usual, pretentious, great_pret

has_nurs proper, less_proper, improper, critical, very_crit

form complete, completed, incomplete, foster

children 1, 2, 3, more

housing convenient, less_conv, critical

finance convenient, inconv

social nonprob, slightly_prob, problematic

health recommended, priority, not_recom

Klasa decyzyjna: not_recom, recommend, very_recom,
priority, spec_prior

¹<https://archive.ics.uci.edu/ml/datasets/Nursery>

Konstrukcja tabel licznosci

Aby włączyć poprawkę LaPlace'a, należy do każdej komórki tabel licznosci dodać jedynkę.

Tabela: Atrybut `parents`.

	not_recom	priority	recommend	spec_prior	very_recom
usual	1440	1924	2	758	196
pretentious	1440	1484	0	1264	132
great_pret	1440	858	0	2022	0

Tabela: Atrybut `has_nurs`.

	not_recom	priority	recommend	spec_prior	very_recom
proper	864	1344	2	252	130
less_proper	864	1344	0	252	132
improper	864	904	0	758	66
critical	864	464	0	1264	0
very_crit	864	210	0	1518	0

Konstrukcja tabel licznosci c.d.

Tabela: Atrybut form.

	not_recom	priority	recommend	spec_prior	very_recom
complete	1080	1152	2	888	118
completed	1080	1092	0	968	100
incomplete	1080	1038	0	1052	70
foster	1080	984	0	1136	40

Tabela: Atrybut children.

	not_recom	priority	recommend	spec_prior	very_recom
1	1080	1206	2	804	148
2	1080	1092	0	968	100
3	1080	984	0	1136	40
more	1080	984	0	1136	40

Konstrukcja tabel licznosci c.d.

Tabela: Atrybut housing.

	not_recom	priority	recommend	spec_prior	very_recom
convenient	1440	1618	2	1052	208
less_conv	1440	1396	0	1384	100
critical	1440	1252	0	1608	20

Tabela: Atrybut finance.

	not_recom	priority	recommend	spec_prior	very_recom
convenient	2160	2244	2	1856	218
inconv	2160	2022	0	2188	110

Tabela: Atrybut social.

	not_recom	priority	recommend	spec_prior	very_recom
nonprob	1440	1515	1	1200	164
slightly_prob	1440	1515	1	1200	164
problematic	1440	1236	0	1644	0

Konstrukcja tabel licznosci c.d.

Tabela: Atrybut health.

	not_recom	priority	recommend	spec_prior	very_recom
recommended	0	2412	2	1578	328
priority	0	1854	0	2466	0
not_recom	4320	0	0	0	0

Tabela: Klasa decyzyjna.

not_recom	4320
recommend	2
very_recom	328
priority	4266
spec_prior	4044

Dla przypomnienia:

$$P(Y = y|X = x) = \frac{P(Y = y) \cdot P(X = x|Y = y)}{P(X = x)}. \quad (1)$$

Opierając się na założeniu naiwnym (atrybuty są niezależne pod warunkiem klasy), prawdopodobieństwo warunkowe możemy rozpisać:

$$\begin{aligned} P(X = x|Y = y) &= P(X_1 = x_1 \wedge \dots \wedge X_n = x_n|Y = y) \\ &= P(X_1 = x_1|Y = y) \cdot \dots \cdot P(X_n = x_n|Y = y). \end{aligned}$$

Przykład

Policzmy sobie następujące prawdopodobieństwo zdarzenia, że klasa decyzyjna wynosi `priority` pod warunkiem:

- `parents = usual`
- `has_nurs = less_proper`
- `form = incomplete`
- `children = 2`
- `housing = less_conv`
- `finance = convenient`
- `social = problematic`
- `health = recommended`

Przykład

Prawdopodobieństwo warunkowe $P(\text{parents} = \text{usual} \mid y = \text{priority})$

$$P(\text{parents} = \text{usual} \mid y = \text{priority}) = \frac{1924}{1924+1484+858} = \frac{1924}{4266}$$

	not_recom	priority	recommend	spec_prior	very_recom
usual	1440	1924	2	758	196
pretentious	1440	1484	0	1264	132
great_pret	1440	858	0	2022	0

Przykład

Prawdopodobieństwo warunkowe $P(\text{has_nurs} = \text{less_proper} \mid y = \text{priority})$

$$P(\text{has_nurs} = \text{less_proper} \mid y = \text{priority}) = \frac{1344}{1344+1344+904+464+210} = \frac{1344}{4266}$$

	not_recom	priority	recommend	spec_prior	very_recom
proper	864	1344	2	252	130
less_proper	864	1344	0	252	132
improper	864	904	0	758	66
critical	864	464	0	1264	0
very_crit	864	210	0	1518	0

Przykład

Prawdopodobieństwo warunkowe $P(\text{form} = \text{incomplete} \mid y = \text{priority})$

$$P(\text{form} = \text{incomplete} \mid y = \text{priority}) = \frac{1038}{1152+1092+1038+984} = \frac{1038}{4266}$$

	not_recom	priority	recommend	spec_prior	very_recom
complete	1080	1152	2	888	118
completed	1080	1092	0	968	100
incomplete	1080	1038	0	1052	70
foster	1080	984	0	1136	40

Przykład

Prawdopodobieństwo warunkowe $P(\text{children} = 2 \mid y = \text{priority})$

$$P(\text{children} = 2 \mid y = \text{priority}) = \frac{1092}{1206+1092+984+984} = \frac{1092}{4266}$$

	not_recom	priority	recommend	spec_prior	very_recom
1	1080	1206	2	804	148
2	1080	1092	0	968	100
3	1080	984	0	1136	40
more	1080	984	0	1136	40

Przykład

Prawdopodobieństwo warunkowe $P(\text{housing} = \text{less_conv} \mid y = \text{priority})$

$$P(\text{housing} = \text{less_conv} \mid y = \text{priority}) = \frac{1396}{1618+1396+1252} = \frac{1396}{4266}$$

	not_recom	priority	recommend	spec_prior	very_recom
convenient	1440	1618	2	1052	208
less_conv	1440	1396	0	1384	100
critical	1440	1252	0	1608	20

Przykład

Prawdopodobieństwo warunkowe $P(\text{finance} = \text{convenient} \mid y = \text{priority})$

$$P(\text{finance} = \text{convenient} \mid y = \text{priority}) = \frac{2244}{2244+2022} = \frac{2244}{4266}$$

	not_recom	priority	recommend	spec_prior	very_recom
convenient	2160	2244	2	1856	218
inconv	2160	2022	0	2188	110

Przykład

Prawdopodobieństwo warunkowe $P(\text{social} = \text{problematic} \mid y = \text{priority})$

$$P(\text{social} = \text{problematic} \mid y = \text{priority}) = \frac{1236}{1515+1515+1236} = \frac{1236}{4266}$$

	not_recom	priority	recommend	spec_prior	very_recom
nonprob	1440	1515	1	1200	164
slightly_prob	1440	1515	1	1200	164
problematic	1440	1236	0	1644	0

Przykład

Prawdopodobieństwo warunkowe $P(\text{health} = \text{recommended} \mid y = \text{priority})$

$$P(\text{health} = \text{recommended} \mid y = \text{priority}) = \frac{2412}{2412+1854+0} = \frac{2412}{4266}$$

	not_recom	priority	recommend	spec_prior	very_recom
recommended	0	2412	2	1578	328
priority	0	1854	0	2466	0
not_recom	4320	0	0	0	0

Przykład

$$P(y=\text{priority}) = \frac{4266}{4320+2+328+4266+4044} = \frac{4266}{12960}$$

not_recom	4320
recommend	2
very_recom	328
priority	4266
spec_prior	4044

Zaniedbawszy czynnik $P(X = x)$ odpowiedź wynosi

$$\frac{1924}{4266} \cdot \frac{1344}{4266} \cdot \frac{1038}{4266} \cdot \frac{1092}{4266} \cdot \frac{1396}{4266} \cdot \frac{2244}{4266} \cdot \frac{1236}{4266} \cdot \frac{2412}{4266} \cdot \frac{4266}{12960} \approx 0,000082144.$$

Uwzględnivszy poprawkę LaPlace'a mamy 0,000082078.

Przykład

Wyniki dla wszystkich klas (włączając poprawkę Laplace'a) wynoszą [1.78488304e-08 8.20786111e-05 6.61375661e-09 5.67208617e-06 2.38002491e-07].

Dzieląc każdy wynik przez sumę wszystkich uzyskamy prawdopodobieństwa, które wynoszą [2.02797285e-04 9.32572003e-01 7.51450856e-05 6.44458853e-02 2.70416929e-03].