

تمرین داده کاوی

موعد تحویل ۱ خرداد

۱. با استفاده از الگوریتم Apriori الگوهای پرتکرار برای مجموعه داده زیر را بیابید. سپس دو قانون انجمنی معتبر برای

این مجموعه داده بیابید:

Transaction ID	Items Bought
1	{a, b, d, e}
2	{b, c, d}
3	{a, b, d, e}
4	{a, c, d, e}
5	{b, c, d, e}
6	{b, d, e}
7	{c, d}
8	{a, b, c}
9	{a, d, e}
10	{b, d}

Min-support = ۳۰٪ Min-confidence = ۶۰٪

۲. برای داده های زیر FP-tree رسم کنید. سپس الگوهای مکرر آن را با استفاده از درخت بیابید:

TID	Items
1	E, A, D, B
2	D, A, C, E, B
3	C, A, B, E
4	B, A, D
5	D
6	D, B
7	A, D, E
8	B, C

۳. با در نظر گرفتن داده های زیر، محاسبه کنید الگوریتم CART کدام صفت را برای ریشه درخت تصمیم انتخاب خواهد

کرد.

Customer ID	Gender	Car Type	Shirt Size	Class
1	M	Family	Small	C0
2	M	Sports	Medium	C0
3	M	Sports	Medium	C0
4	M	Sports	Large	C0
5	M	Sports	Extra Large	C0
6	M	Sports	Extra Large	C0
7	F	Sports	Small	C0
8	F	Sports	Small	C0
9	F	Sports	Medium	C0
10	F	Luxury	Large	C0
11	M	Family	Large	C1
12	M	Family	Extra Large	C1
13	M	Family	Medium	C1
14	M	Luxury	Extra Large	C1
15	F	Luxury	Small	C1
16	F	Luxury	Small	C1
17	F	Luxury	Medium	C1
18	F	Luxury	Medium	C1
19	F	Luxury	Medium	C1
20	F	Luxury	Large	C1

۴. در مجموعه داده زیر:

Instance	a_1	a_2	a_3	Target Class
1	T	T	1.0	+
2	T	T	6.0	+
3	T	F	5.0	-
4	F	F	4.0	+
5	F	T	7.0	-
6	F	T	3.0	-
7	F	F	8.0	-
8	T	F	7.0	+
9	F	T	5.0	-

- information gain را برای صفات a_1 و a_2 محاسبه کنید.

- برای صفت a_3 که یک صفت پیوسته است، information gain را برای تمام نقاط شکست ممکن محاسبه کنید.

- از میان سه صفت a_1 ، a_2 و a_3 کدام یک بهترین مورد برای تقسیم است؟

۵. فرض کنید یک الگوریتم دسته بندی مبتنی بر قانون، برای مجموعه داده ای با صفات زیر قوانین داده شده را تولید کرده است:

- Air Conditioner = { Working, Broken }
- Engine = { Good, Bad }
- Mileage = { High, Medium, Low }
- Rust = { Yes, No }

Mileage = High \rightarrow Value = Low Mileage = Low \rightarrow Value = High

Air Conditioner = Working, Engine = Good \rightarrow Value = High

Air Conditioner = Working, Engine = Bad \rightarrow Value = Low

Air Conditioner = Broken \rightarrow Value = Low

- آیا این قوانین دوبدو ناسازگارند؟ آیا نیاز به مرتب سازی و تعیین اولویت بین آن ها وجود دارد؟

- آیا این قوانین جامع هستند یا نیاز به قانون پیش فرض دارند؟

۶. با فرض مستقل بودن صفات در مجموعه داده زیر و با استفاده از الگوریتم naïve bayse کلاس نمونه داده شده را

تعیین کنید:

chills	runny nose	headache	fever	flu?
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

chills	runny nose	headache	fever	flu?
Y	N	Mild	Y	?