

تمرین دو

الف) گسسته سازی داده ها

با توجه به جدول اشاره شده در کتاب:

Table 12.10. Iris dataset discretization and labels employed

Attribute	Range or value	Label
Sepal length	4.30–5.55	sl_1
	5.55–6.15	sl_2
	6.15–7.90	sl_3
Sepal width	2.00–2.95	sw_1
	2.95–3.35	sw_2
	3.35–4.40	sw_3
Petal length	1.00–2.45	pl_1
	2.45–4.75	pl_2
	4.75–6.90	pl_3
Petal width	0.10–0.80	pw_1
	0.80–1.75	pw_2
	1.75–2.50	pw_3
Class	Iris-setosa	c_1
	Iris-versicolor	c_2
	Iris-virginica	c_3

با محدوده های داده شده داده ها را گسسته سازی میکنیم

تصویر زیر خروجی مقادیر گسسته سازی شده را نشان می‌دهد:

Sepal_lengh	Sepal_width	Petal_lengh	Petal_width
5.9	3.0	4.2	1.5
6.9	3.1	4.9	1.5
6.6	2.9	4.6	1.3
4.6	3.2	1.4	0.2
6.0	2.2	4.0	1.0
...
5.4	3.4	1.7	0.2
5.5	2.6	4.4	1.2
6.7	3.0	5.2	2.3
7.7	3.8	6.7	2.2
5.1	3.4	1.5	0.2

که پس از گسسته سازی به شکل زیر خواهد بود: (در زیر تصویر مقادیر یکه را می‌بینیم).

Sepal_lengh	Sepal_width	Petal_lengh	Petal_width
sl2	sw2	pl2	pw2
sl3	sw2	pl3	pw2
sl3	sw1	pl2	pw2
sl1	sw2	pl1	pw1
sl2	sw1	pl2	pw2
...
sl1	sw3	pl1	pw1
sl1	sw1	pl2	pw2
sl3	sw2	pl3	pw3
sl3	sw3	pl3	pw3
sl1	sw3	pl1	pw1

```

Number Of Unique Values Of Sepal_lengh : 3
Number Of Unique Values Of Sepal_width : 3
Number Of Unique Values Of Petal_lengh : 3
Number Of Unique Values Of Petal_width : 3
Number Of Unique Values Of Class : 3

```

ب) استخراج قواعد انجمنی

برای اینکار ابتدا باید داده ها را به شکل one hot vector در بیاوریم
نمای کلی داده ها پس از این عمل در زیر آمده است:

	sl1	sl2	sl3	sw1	sw2	sw3	pl1	pl2	pl3	pw1	pw2	pw3	Iris-setosa	Iris-versicolor	Iris-virginica
0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
1	0	0	1	0	1	0	0	0	1	0	1	0	0	1	0
2	0	0	1	1	0	0	0	1	0	0	1	0	0	1	0
3	1	0	0	0	1	0	1	0	0	1	0	0	1	0	0
4	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0
...
145	1	0	0	0	0	1	1	0	0	1	0	0	1	0	0
146	1	0	0	1	0	0	0	1	0	0	1	0	0	1	0
147	0	0	1	0	1	0	0	0	1	0	0	1	0	0	1
148	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1
149	1	0	0	0	0	1	1	0	0	1	0	0	1	0	0

150 rows × 15 columns

حال میتوانیم با استفاده از مقدار تعیین شده 30 درصد برای پشتیبانی نسبی (با استفاده از الگوریتم Apriori) آیتم ست ها را بدست آوریم

support	itemsets	8	0.306667	(pw3)			
0	0.386667	(sl1)	9	0.333333	(Iris-setosa)		
1	0.366667	(sl3)	10	0.333333	(Iris-versicolor)		
2	0.373333	(sw1)	11	0.333333	(Iris-virginica)	17	0.326667 (pl3, Iris-virginica)
3	0.380000	(sw2)	12	0.300000	(pl1, sl1)	18	0.326667 (Iris-versicolor, pw2)
4	0.326667	(pl1)	13	0.306667	(Iris-setosa, sl1)	19	0.300000 (pw3, Iris-virginica)
5	0.300000	(pl2)	14	0.326667	(Iris-setosa, pl1)	20	0.300000 (Iris-setosa, pl1, sl1)
6	0.366667	(pl3)	15	0.300000	(pl2, pw2)	21	0.300000 (pl3, pw3, Iris-virginica)
7	0.360000	(pw2)	16	0.306667	(pl3, pw3)		

تعداد 21 آیتم ست وجود دارد که با استفاده از مقدار تعیین شده 80 درصد برای مینیمم درجه اعتماد 24 قانون انجمنی خواهیم داشت که به شرح زیر است:

	antecedents	consequents	antecedent support	consequent support	support	confidence
0	(pl1)	(sl1)	0.326667	0.386667	0.300000	0.918367
1	(Iris-setosa)	(sl1)	0.333333	0.386667	0.306667	0.920000
2	(Iris-setosa)	(pl1)	0.333333	0.326667	0.326667	0.980000
3	(pl1)	(Iris-setosa)	0.326667	0.333333	0.326667	1.000000
4	(pl2)	(pw2)	0.300000	0.360000	0.300000	1.000000
5	(pw2)	(pl2)	0.360000	0.300000	0.300000	0.833333
6	(pl3)	(pw3)	0.366667	0.306667	0.306667	0.836364
7	(pw3)	(pl3)	0.306667	0.366667	0.306667	1.000000
8	(pl3)	(Iris-virginica)	0.366667	0.333333	0.326667	0.890909
9	(Iris-virginica)	(pl3)	0.333333	0.366667	0.326667	0.980000
10	(Iris-versicolor)	(pw2)	0.333333	0.360000	0.326667	0.980000
11	(pw2)	(Iris-versicolor)	0.360000	0.333333	0.326667	0.907407
12	(pw3)	(Iris-virginica)	0.306667	0.333333	0.300000	0.978261
13	(Iris-virginica)	(pw3)	0.333333	0.306667	0.300000	0.900000
14	(Iris-setosa, pl1)	(sl1)	0.326667	0.386667	0.300000	0.918367
15	(Iris-setosa, sl1)	(pl1)	0.306667	0.326667	0.300000	0.978261
16	(pl1, sl1)	(Iris-setosa)	0.300000	0.333333	0.300000	1.000000
17	(Iris-setosa)	(pl1, sl1)	0.333333	0.300000	0.300000	0.900000
18	(pl1)	(Iris-setosa, sl1)	0.326667	0.306667	0.300000	0.918367
19	(pl3, pw3)	(Iris-virginica)	0.306667	0.333333	0.300000	0.978261
20	(pl3, Iris-virginica)	(pw3)	0.326667	0.306667	0.300000	0.918367
21	(pw3, Iris-virginica)	(pl3)	0.300000	0.366667	0.300000	1.000000
22	(pl3)	(pw3, Iris-virginica)	0.366667	0.300000	0.300000	0.818182
23	(pw3)	(pl3, Iris-virginica)	0.306667	0.326667	0.300000	0.978261
24	(Iris-virginica)	(pl3, pw3)	0.333333	0.306667	0.300000	0.900000