

CS444x PROJECT 3

PART I (solution)

Mine association rules by hand from the Contact-lenses dataset available with the Weka system by faithfully following the Apriori algorithm with minimal support = 20% and minimal confidence 90%. That is, start by generating candidate itemsets and frequent itemsets level by level and after all frequent itemsets have been generated, produce from them all the rules with confidence greater than or equal to the min. confidence. **SHOW IN DETAIL ALL THE STEPS OF THE PROCESS.**

Candidate Itemset – is a potential frequent itemset.

Frequent itemset – an itemset whose support is greater than some user-specified minimum support.

Acronyms used:

Spectacle-prescription = SP

Astigmatism = A

Tear-rate-production = TPR

Contact-Lense = CL

pre-presbyopic = PP

presbyopic = P

myope = M

hypermetrope = H

reduced = R

normal = N

One-Item Candidate Sets

Set Num	Attribute	Value	Support
1	Age	Young	8
2	Age	PP	8
3	Age	P	8
4	SP	M	12
5	SP	H	12
6	A	no	12
7	A	yes	12
8	TPR	R	12
9	TPR	N	12
10	CL	Soft	5
11	CL	Hard	4
12	CL	None	15

Since (CL=Hard) has support less than 20%, it will be pruned and not appear in subsequent item sets.

Two-Item Candidate Sets

Set Num	Attributes	Values	Support
1	Age, SP	Young, M	4
2	Age, SP	Young, H	4
3	Age, A	Young, no	4
4	Age, A	Young, yes	4
5	Age, TPR	Young, R	4
6	Age, TPR	Young, N	4
7	Age, CL	Young, Soft	2
9	Age, CL	Young, none	4
10	Age, SP	PP, M	4
11	Age, SP	PP, H	4
12	Age, A	PP, no	4
13	Age, A	PP, yes	4
14	Age, TPR	PP, R	4
15	Age, TPR	PP, N	4
16	Age, CL	PP, Soft	2
18	Age, CL	PP, none	5
19	Age, SP	P, M	4
20	Age, SP	P, H	4
21	Age, A	P, no	4
22	Age, A	P, yes	4
23	Age, TPR	P, R	4
24	Age, TPR	P, N	4
25	Age, CL	P, Soft	1
27	Age, CL	P, None	6
28	SP, A	M, no	6
29	SP, A	M, yes	6
30	SP, TPR	M, R	6
31	SP, TPR	M, N	6
32	SP, CL	M, Soft	2
34	SP, CL	M, None	7
35	SP, A	H, no	6
36	SP, A	H, yes	6
37	SP, TPR	H, R	6
38	SP, TPR	H, N	6
39	SP, CL	H, Soft	3
41	SP, CL	H, None	8
42	A, TPR	No, R	6
43	A, TPR	No, N	6
44	A, CL	No, Soft	5

46	A, CL	No, None	7
47	A, TPR	Yes, R	6
48	A, TPR	Yes, N	6
49	A, CL	Yes, Soft	0
51	A, CL	Yes, None	8
52	TPR, CL	R, Soft	0
54	TPR, CL	R, None	12
55	TPR, CL	N, Soft	5
57	TPR, CL	N, None	3

Three-Item Candidate Sets

Reference	Set Num	Attribute	Value	Support
2i#18	1	Age, CL, SP	PP, None, M	2
	2	Age, CL, SP	PP, None, H	3
	3	Age, CL, A	PP, None, No	2
	4	Age, CL, A	PP, None, Yes	3
	5	Age, CL, TPR	PP, None, R	4
	6	Age, CL, TPR	PP, None, N	1
2i#27	1	Age, CL, SP	P, None, M	3
	2	Age, CL, SP	P, None, H	3
	3	Age, CL, A	P, None, No	3
	4	Age, CL, A	P, None, Yes	3
	5	Age, CL, TPR	P, None, R	3
	6	Age, CL, TPR	P, None, N	2
2i#28	1	SP, A, Age	M, No, Young	2
	2	SP, A, Age	M, No, PP	2
	3	SP, A, Age	M, No, P	2
	4	SP, A, TPR	M, No, R	3
	5	SP, A, TPR	M, No, N	3
	6	SP, A, CL	M, No, Soft	2
	7	SP, A, CL	M, No, Hard	0
	8	SP, A, CL	M, No, None	4
2i#29	1	SP, A, Age	M, Yes, Young	2
	2	SP, A, Age	M, Yes, PP	2
	3	SP, A, Age	M, Yes, P	2
	4	SP, A, TPR	M, Yes, R	3
	5	SP, A, TPR	M, Yes, N	3
	6	SP, A, CL	M, Yes, Soft	0
	7	SP, A, CL	M, Yes, Hard	3
	8	SP, A, CL	M, Yes, None	3
2i#30	1	SP, TPR, Age	M, R, Young	2
	2	SP, TPR, Age	M, R, PP	2

	3	SP, TPR, Age	M, R, P	2
	4	SP, TPR, A	M, R, No	3
	5	SP, TPR, A	M, R, Yes	3
	6	SP, TPR, CL	M, R, Soft	0
	7	SP, TPR, CL	M, R, Hard	0
	8	SP, TPR, CL	M, R, None	6
2i#31	1	SP, TPR, Age	M, N, Young	2
	2	SP, TPR, Age	M, N, PP	2
	3	SP, TPR, Age	M, N, P	2
	4	SP, TPR, A	M, N, No	3
	5	SP, TPR, A	M, N, Yes	3
	6	SP, TPR, CL	M, N, Soft	2
	7	SP, TPR, CL	M, N, Hard	3
	8	SP, TPR, CL	M, N, None	1
2i#34	1	SP, CL, Age	M, None, Young	2
	2	SP, CL, Age	M, None, PP	2
	3	SP, CL, Age	M, None, P	3
	4	SP, CL, A	M, None, No	4
	5	SP, CL, A	M, None, Yes	3
Duplicate	6	SP, CL, TPR	M, None, R	6
	7	SP, CL, TPR	M, None, N	2
2i#35	1	SP, A, Age	H, No, Young	2
	2	SP, A, Age	H, No, PP	2
	3	SP, A, Age	H, No, P	2
	4	SP, A, TPR	H, No, R	4
	5	SP, A, TPR	H, No, N	3
	6	SP, A, CL	H, No, Soft	3
	7	SP, A, CL	H, No, Hard	0
	8	SP, A, CL	H, No, None	3
2i#36	1	SP, A, Age	H, Yes, Young	2
	2	SP, A, Age	H, Yes, PP	2
	3	SP, A, Age	H, Yes, P	2
	4	SP, A, TPR	H, Yes, R	3
	5	SP, A, TPR	H, Yes, N	3
	6	SP, A, CL	H, Yes, Soft	3
	7	SP, A, CL	H, Yes, Hard	1
	8	SP, A, CL	H, Yes, None	5
2i#37	1	SP, TPR, Age	H, R, Young	2
	2	SP, TPR, Age	H, R, PP	2
	3	SP, TPR, Age	H, R, P	2
	4	SP, TPR, A	H, R, No	3
	5	SP, TPR, A	H, R, Yes	3
	6	SP, TPR, CL	H, R, Soft	0
	7	SP, TPR, CL	H, R, Hard	0
	8	SP, TPR, CL	H, R, None	6

2i#38	1	SP, TPR, Age	H, N, Young	2
	2	SP, TPR, Age	H, N, PP	2
	3	SP, TPR, Age	H, N, P	2
	4	SP, TPR, A	H, N, No	3
	5	SP, TPR, A	H, N, Yes	3
	6	SP, TPR, CL	H, N, Soft	3
	7	SP, TPR, CL	H, N, Hard	2
	8	SP, TPR, CL	H, N, None	1
2i#41	1	SP, CL, Age	H, None, Young	2
	2	SP, CL, Age	H, None, PP	3
	3	SP, CL, Age	H, None, P	3
	4	SP, CL, A	H, None, A	3
Duplicate	5	SP, CL, A	H, None, Yes	5
Duplicate	6	SP, CL, TPR	H, None, R	6
	7	SP, CL, TPR	H, None, N	2
2i#42	1	A, TPR, Age	No, R, Young	2
	2	A, TPR, Age	No, R, PP	2
	3	A, TPR, Age	No, R, P	2
	4	A, TPR, SP	No, R, M	3
	5	A, TPR, SP	No, R, H	3
	6	A, TPR, CL	No, R, Soft	0
	7	A, TPR, CL	No, R, Hard	0
	8	A, TPR, CL	No, R, None	6
2i#43	1	A, TPR, Age	No, N, Young	2
	2	A, TPR, Age	No, N, PP	2
	3	A, TPR, Age	No, N, P	2
	4	A, TPR, SP	No, N, M	3
	5	A, TPR, SP	No, N, H	3
	6	A, TPR, CL	No, N, Soft	5
	7	A, TPR, CL	No, N, Hard	0
	8	A, TPR, CL	No, N, None	1
2i#44	1	A, CL, Age	No, Soft, Young	2
	2	A, CL, Age	No, Soft, PP	2
	3	A, CL, Age	No, Soft, P	1
	4	A, CL, SP	No, Soft, M	2
	5	A, CL, SP	No, Soft, H	3
	6	A, CL, TPR	No, Soft, R	0
Duplicate	7	A, CL, TPR	No, Soft, N	5
2i#46	1	A, CL, Age	No, None, Young	2
	2	A, CL, Age	No, None, PP	2
	3	A, CL, Age	No, None, P	3
	4	A, CL, SP	No, None, M	4
	5	A, CL, SP	No, None, H	3
Duplicate	6	A, CL, TPR	No, None, R	6
	7	A, CL, TPR	No, None, N	1

2i#47	1	A, TPR, Age	Yes, R, Young	2
	2	A, TPR, Age	Yes, R, PP	2
	3	A, TPR, Age	Yes, R, P	2
	4	A, TPR, SP	Yes, R, M	3
	5	A, TPR, SP	Yes, R, H	3
	6	A, TPR, CL	Yes, R, Soft	0
	7	A, TPR, CL	Yes, R, Hard	0
	8	A, TPR, CL	Yes, R, None	6
2i#48	1	A, TPR, Age	Yes, Normal, Young	2
	2	A, TPR, Age	Yes, Normal, PP	2
	3	A, TPR, Age	Yes, Normal, P	2
	4	A, TPR, SP	Yes, Normal, M	2
	5	A, TPR, SP	Yes, Normal, H	2
	6	A, TPR, SP	Yes, Normal, Soft	0
	7	A, TPR, CL	Yes, Normal, Hard	4
	8	A, TPR, CL	Yes, Normal, None	2
2i#51	1	A, CL, Age	Yes, None, Young	2
	2	A, CL, Age	Yes, None, PP	2
	3	A, CL, Age	Yes, None, P	2
	4	A, CL, SP	Yes, None, M	3
	5	A, CL, SP	Yes, None, H	3
	6	A, CL, TPR	Yes, None, R	6
	7	A, CL, TPR	Yes, None, N	1
2i#54	1	TPR, CL, Age	R, None, Young	4
	2	TPR, CL, Age	R, None, PP	4
	3	TPR, CL, Age	R, None, P	4
Duplicate	4	TPR, CL, SP	R, None, M	5
Duplicate	5	TPR, CL, SP	R, None, H	6
Duplicate	6	TPR, CL, A	R, None, No	6
Duplicate	7	TPR, CL, A	R, None, Yes	5
2i#55	1	TPR, CL, Age	N, None, Young	0
	2	TPR, CL, Age	N, None, PP	1
	3	TPR, CL, Age	N, None, P	2
	4	TPR, CL, SP	N, None, M	1
	5	TPR, CL, SP	N, None, H	2
	6	TPR, CL, A	N, None, No	1
	7	TPR, CL, A	N, None, Yes	2

3 Item-Frequent Sets

Those that are obtained after pruning of the non-frequent subsets end according to the minimal support criteria.

Reference	Set Num	Attribute	Value	Support
2i#30,8	1	SP, TPR, CL	M, R, CL	6
2i#35,8	2	SP, A, CL	H, Yes, None	5
2i#36,8	3	SP, TPR, CL	H, R, None	6
2i#41,8	4	A, TPR, CL	No, R, None	6
2i#42,6	5	A, TPR, CL	No, N, Soft	5
2i#46,8	6	A, TPR, CL	Yes, R, None	6

Rules produced	Confidence
(TPR=R) → (CL=None)	12/12 = 1.00
(SP=M) ^ (TPR=R) → (CL=None)	6/6 = 1.00
(SP=H) ^ (TPR=R) → (CL=None)	6/6 = 1.00
(A=No) ^ (TPR=R) → (CL=None)	6/6 = 1.00
(A=Yes) ^ (TPR=R) → (CL=None)	6/6 = 1.00
(CL=Soft) → (TPR=Normal)	5/5 = 1.00
(CL=Soft)^(TPR=N) → (A=No)	5/5 = 1.00
(CL=Soft) → (A=No)^(TPR=Normal)	5/5 = 1.00
(CL=Soft)^(A=No) → (TPR=Normal)	5/5 = 1.00
(CL=Soft) → (A=No)	5/5 = 1.00