

**Data Mining, CSCI 347, Fall 2019**  
**Covering Rules, Sept. 25**

Apply the PRISM algorithm to the contact lenses dataset to determine when contact\_lenses='hard'.

5 attributes- all nominal

Class is nominal (none, soft and hard)

24 instances – all possible situations are given, very idealized

Relation: contact-lenses					
No.	1: age Nominal	2: spectacle-prescrip Nominal	3: astigmatism Nominal	4: tear-prod-rate Nominal	5: contact-lenses Nominal
1	young	myope	no	reduced	none
2	young	myope	no	normal	soft
3	young	myope	yes	reduced	none
4	young	myope	yes	normal	hard
5	young	hypermetrope	no	reduced	none
6	young	hypermetrope	no	normal	soft
7	young	hypermetrope	yes	reduced	none
8	young	hypermetrope	yes	normal	hard
9	pre-presbyopic	myope	no	reduced	none
10	pre-presbyopic	myope	no	normal	soft
11	pre-presbyopic	myope	yes	reduced	none
12	pre-presbyopic	myope	yes	normal	hard
13	pre-presbyopic	hypermetrope	no	reduced	none
14	pre-presbyopic	hypermetrope	no	normal	soft
15	pre-presbyopic	hypermetrope	yes	reduced	none
16	pre-presbyopic	hypermetrope	yes	normal	none
17	presbyopic	myope	no	reduced	none
18	presbyopic	myope	no	normal	none
19	presbyopic	myope	yes	reduced	none
20	presbyopic	myope	yes	normal	hard
21	presbyopic	hypermetrope	no	reduced	none
22	presbyopic	hypermetrope	no	normal	soft
23	presbyopic	hypermetrope	yes	reduced	none
24	presbyopic	hypermetrope	yes	normal	none

Consider each attribute/value pair, seeing it as a rule. For example, for age=young, imagine the rule: If age=young then contact-lenses = hard

Maximize the accuracy of the rule. That is

$$p/t = \frac{\text{\# correct class predictions made by this rule}}{\text{total number instances to which the rule applies}}$$

age = young            2 correct out of 8 total, 2/8

age = pre-presbyopic    1/8

age = presbyopic        1/8

spectacle-prescription = myope            3/12

spectacle-prescription = hypermetrope   1/1

astigmatism = no        0/12

astigmatism = yes      4/12

tear-prod-rate = reduced 0/12

tear-prod-rate = normal 4/12

Best ratio is astigmatism = yes at  $4/12 = 33\%$ .

Gives the rule with 33% accuracy:

If astigmatism = yes then contact-lenses = hard

This rule can, and needs to be refined by considering only those instances where astigmatism = yes and repeating the process. Repeat until accuracy of rule is 100%.

age = young 2/4

age = pre-presbyopic 1/4

age = presbyopic 1/4

spectacle-prescription = myope 3/6

spectacle-prescription = hypermetrope 1/6

tear-prod-rate = reduced 0/6

tear-prod-rate = normal 4/6

Best ratio is tear-prod-rate = normal at 67%.

Gives the rule with 67% accuracy:

If astigmatism = yes and tear-prod-rate = normal then contact-lenses = hard

Repeat until accuracy of rule is 100% or we run out of attributes to split on.

age = young 2/2

age = pre-presbyopic 1/2

age = presbyopic 1/2

spectacle-prescription = myope 3/3

spectacle-prescription = hypermetrope 1/3

Adding spectacle-prescription = myope gives 100% accuracy.

Rule:

**If astigmatism = yes and tear-prod-rate = normal and spectacle-prescription = myope then contact-lenses = hard**

We've found a 100% accurate rule for contact-lenses=hard. However, there is still one instance with contact-lenses=hard that we haven't gotten. Start the process again but first remove the instances we've already covered.

If ? then contact-lenses = hard

age = young                    1/7  
age = pre-presbyopic        0/7  
age = presbyopic            0/7  
spectacle-prescription = myope            0/9  
spectacle-prescription = hypermetrope 1/12  
astigmatism = no            0/12  
astigmatism = yes           1/9  
tear-prod-rate = reduced 0/11  
tear-prod-rate = normal    1/11

Best ratio is age = young, giving the rule with 14% accuracy:

If age=young then contact-lenses = hard

Refine the rule.

spectacle-prescription = myope            0/3  
spectacle-prescription = hypermetrope 1/4  
astigmatism = no            0/4  
astigmatism = yes           1/3  
tear-prod-rate = reduced 0/4  
tear-prod-rate = normal    1/3

Gives the rule with 33% accuracy:

If age=young and astigmatism = yes then contact-lenses = hard

Refine the rule.

spectacle-prescription = myope            0/1  
spectacle-prescription = hypermetrope 1/2  
tear-prod-rate = reduced 0/2  
tear-prod-rate = normal    1/1

Gives the rule with 100% accuracy:

**If age=young and astigmatism = yes and tear-prod-rate = normal  
then contact-lenses = hard**