

# Classifiers: toy example of Classification rules

Data mining lab 3

# Input data

	<b>skin</b>	<b>color</b>	<b>size</b>	<b>flesh</b>	<b>class</b>
1	hairy	brown	large	hard	safe
2	hairy	green	large	hard	safe
3	smooth	red	large	soft	dangerous
4	hairy	green	large	soft	safe
5	hairy	red	small	hard	safe
6	smooth	red	small	hard	safe
7	smooth	brown	small	hard	safe
8	hairy	green	small	soft	dangerous
9	smooth	green	small	hard	dangerous
10	hairy	red	large	hard	safe
11	smooth	brown	large	soft	safe
12	smooth	green	small	soft	dangerous
13	hairy	red	small	soft	safe
14	smooth	red	large	hard	dangerous
15	smooth	red	small	hard	safe
16	hairy	green	small	hard	dangerous

Query: hairy skin, red color, large with soft flesh.

The question is: is it safe to eat this unknown animal?

# Association Rules I

For rule induction follow the following steps:

1. Choose class value, and formulate the first rule as if ? then class\_value:

**if ? Then safe**

# Association Rules II

2. For each attribute-value pair calculate the accuracy of the rule, and choose the one with the greater accuracy, and between them with the greater coverage (absolute number of instances covered by the rule divided by the total number of instances)

For example,

- if color=red then safe

accuracy:  $(\text{red} + \text{safe}) / (\text{all red}) = 5/7$

- if color=brown then safe**

accuracy:  $(\text{brown} + \text{safe}) / (\text{all brown}) = 3/3$  [100%]

etc.

- We choose the best rule with the accuracy 100%

Rule I: if color=brown then safe

# Association Rules III

3. If correctness is less than 100%, adjust the rule by adding **and** clause:

**if attribute1=value1 and ? then class\_value.**

- For this perform step 2 but only for data instances with attribute1=value1. Repeat the process recursively till the correctness of the rule is 100%.

# Association Rules IV

4. To deduce the next rule, perform the same steps 2,3 but only on the data instances not covered by the first rule: remove all rows with color=brown

<b>skin</b>	<b>color</b>	<b>size</b>	<b>flesh</b>	<b>class</b>
2 hairy	green	large	hard	safe
3 smooth	red	large	soft	dangerous
4 hairy	green	large	soft	safe
5 hairy	red	small	hard	safe
6 smooth	red	small	hard	safe
8 hairy	green	small	soft	dangerous
9 smooth	green	small	hard	dangerous
10 hairy	red	large	hard	safe
12 smooth	green	small	soft	dangerous
13 hairy	red	small	soft	safe
14 smooth	red	large	hard	dangerous
15 smooth	red	small	hard	safe
16 hairy	green	small	hard	dangerous

# Association Rules V

if ? Then safe

**If hairy then safe: (accuracy: 5/7=0.7)**

If smooth then safe: (accuracy: 2/6=0.5)

If green then safe: (accuracy: 2/6=0.5)

**If red then safe: (accuracy: 5/7=0.7)**

If large then safe: (accuracy: 3/5=0.6)

If small then safe: (accuracy: 4/8=0.5)

If hard then safe: (accuracy: 5/8=0.6)

If soft then safe: (accuracy: 2/5=0.4)

There are 2 rules with the same accuracy,  
but also with the same coverage: 7/13

We choose one of them:

**Rule II: if hairy then safe**

# Association Rules VI

Since the accuracy is not 100% we adjust it by adding the **AND** clause:  
**If hairy AND ? Then safe**

<b>skin</b>	<b>color</b>	<b>size</b>	<b>flesh</b>	<b>class</b>
2 hairy	green	large	hard	safe
4 hairy	green	large	soft	safe
5 hairy	red	small	hard	safe
8 hairy	green	small	soft	dangerous
10 hairy	red	large	hard	safe
13 hairy	red	small	soft	safe
16 hairy	green	small	hard	dangerous

We find that the rule if hairy and large has 100% accuracy

**Rule II: if skin=hairy and size=large then safe**