Classification Rules Exercises

Master degree in Computer Science

Dept. Demacs – Unical

Prof. P. Rullo

rullo@unical.it

AY 2021-2022

Exercise 1

- Consider a training set that contains 100 positive examples and 400 negative ones. Starting from the most general rule {} → +, determine which of the following candidate rules
  - r1: A=a → + (covers 4 positive and 1 negative examples)
  - r2: B=b → + (covers 30 positive and 10 negative examples)
  - r3: C=c → + (covers 100 positive and 90 negative examples)
- is the best according to:
  (a) Rule accuracy: the number of examples r is consistent with over the number of examples covered by r
  (b) FOIL's IG: $p_1 (\log_2 p_1/(p_1+n_1) - \log_2 p_0/(p_0+n_0))$

Exercise 1 - Solution

- r1: A=a $\rightarrow$ + (covers 4 positive and 1 negative example)

  (a) Rule accuracy: 4/5 = 0.8

  (b) FOIL's IG: p_0=100, n_0=400; p_1=4; n_1=1

  FOIL's IG = $p_1 (\log_2 p_1/(p_1+n_1) - \log_2 p_0/(p_0+n_0))$ = 8

- r2: B=b $\rightarrow$ + (covers 30 positive and 10 negative examples)

  (a) Rule accuracy: 30/40 = 0.75

  (b) FOIL's IG: p_0=100, n_0=400; p_1=30; n_1=10

  FOIL's IG = $p_1 (\log_2 p_1/(p_1+n_1) - \log_2 p_0/(p_0+n_0))$ = 57.21

- r3: C=c $\rightarrow$ + (covers 100 positive and 90 negative examples)

  (a) Rule accuracy: 100/190 = 0.53

  (b) FOIL's IG: p_0=100, n_0=400; p_1=100; n_1=90

  FOIL's IG = $p_1 (\log_2 p_1/(p_1+n_1) - \log_2 p_0/(p_0+n_0))$ = 140

Exercise 2

- Consider a binary classification problem with the following set of attributes and attribute values describing a second-hand car
  - Air Conditioner = {Working, Broken}
  - Engine = {Good, Bad}
  - Mileage = {High, Medium, Low}
  - Rust = {Yes, No}
- The target attribute is Value = {Low, High} – representing the value of the car
- Suppose a rule-based classifier produces the following rule set:
  1. Mileage=High → Value=Low
  2. Mileage=Low→ Value=High
  3. Air Cond = Working, Engine=G → Value=High
  4. Air Cond = Working, Engine=Bad → Value=Low
  5. Air Cond = Broken → Value = Low

Exercise 2

1. Answer the following questions:

   (a) are the rules mutually exclusive?

   (b) Is the rule set exhaustive?

   (c) is ordering needed?

   (d) is the default rule needed?

2. Define a condition for mutually exclusivity between two rules

Exercise 2 - Solution

1. Responses
   a. NO – for instance, <working, good, low, yes> triggers both rules 2 and 4.
   b. YES– the last 3 rules are able to classify any instance. Indeed, AirCond is either broken or working. In the former case, i.e., AirCond=broken, rule 5 is fired. Otherwise, either rule 3 o rule 4 is fired, depending on the value of Engine – in particular, rule 3 if Engine is good, and rule 4 otherwise.
   c. YES because of point a
   d. NO because of point b

Exercise 2 - Solution

2. condition for mutually exclusivity : two rules r1 and r2 are mutually exclusive if there is A=a in the antecedent of r1, and A= b, with b <> a, in the antecedent of r2, e.g.,

- Air Cond = W, Engine=Good → Value=High
- Air Cond = W, Engine=Bad → Value=Low