

# Knowledge based systems

## Lab 2

Rasmus Hammar

### 1. Formal definitions

$U = \{\text{Smith, Jones, Parker, Hanson, Moore, Fields, Starr}\}$

$A = \{\text{Age, Sex, LEMS}\}$

$V_{\text{Age}} = \{16-30, 31-45, 46-60\}$

$V_{\text{Sex}} = \{\text{Male, Female}\}$

$V_{\text{LEMS}} = \{0, 1-25, 26-49, 50\}$

### 2. Discernability 1

a)  $\text{IND}\{\text{Age}\} = \{\{\text{Smith, Jones, Fields}\}, \{\text{Parker, Hanson}\}, \{\text{Moore, Starr}\}\}$

b) *upper* for  $\text{Walk} = \text{Yes}$  by  $\text{Age} = \{\{\text{Smith, Fields, Jones}\}, \{\text{Hanson, Parker}\}\}$  *lower* for  $\text{Walk} = \text{Yes}$  by  $\text{Age} = \{\}$

Accuracy of approximation =  $\frac{0}{5} = 0$

c) Rough membership functions for  $\text{Walk} = \text{Yes}$  by  $\text{IND}\{\text{Age}\}$ :  $\{\text{Smith, Jones, Fields}\} = \frac{2}{3}$   $\{\text{Parker, Hanson}\} = \frac{1}{2}$

Good model? No.

### 3. Discernability 2

a) For  $B = \{\text{Age, Sex}\}$ ,  $\text{POS}_B(\{d\}) = \{\{\text{Moore, Starr}\}, \{\text{Fields}\}\}$

b) For  $A = \{\text{Age, Sex, LEMS}\}$ ,  $\text{POS}_A(\{d\}) = \{\{\text{Smith}\}, \{\text{Fields}\}, \{\text{Jones}\}, \{\text{Moore, Starr}\}\}$

### 4. Discernability 3

a) Equivalence classes without  $d = \{\{\text{Smith}\}, \{\text{Jones}\}, \{\text{Fields}\}, \{\text{Parker, Hanson}\}, \{\text{Moore, Starr}\}\}$

b)

	Smith	Jones	Parker	Hanson	Moore	Fields	Starr
Smith	$\Theta$						
Jones	LEMS	$\Theta$					
Parker	Age, LEMS	Age, LEMS	$\Theta$				
Hanson	Age, LEMS	Age, LEMS	$\Theta$		$\Theta$		
Moore	Age, Sex, LEMS	$\Theta$					
Fields	Sex	Sex, LEMS	Sex, LEMS	Sex, LEMS	Sex, LEMS	Age, LEMS	$\Theta$
Starr	Age, Sex, LEMS	Age, Sex, LEMS	Age, Sex, LEMS	Age, Sex, LEMS	$\Theta$	Age, LEMS	$\Theta$

c) Boolean discernability function

$$\begin{aligned}
 &= (\text{LEMS})(\text{Age} \vee \text{LEMS})(\text{Age} \vee \text{LEMS})(\text{Age} \vee \text{Sex} \vee \text{LEMS})(\text{Sex})(\text{Age} \vee \text{Sex} \vee \text{LEMS})(\text{Age} \vee \text{LEMS}) \\
 &\quad (\text{Age} \vee \text{LEMS})(\text{Age} \vee \text{Sex} \vee \text{LEMS})(\text{Sex} \vee \text{LEMS})(\text{Age} \vee \text{Sex} \vee \text{LEMS})(\text{Age} \vee \text{Sex} \vee \text{LEMS})(\text{Sex} \vee \text{LEMS}) \\
 &\quad (\text{Age} \vee \text{Sex} \vee \text{LEMS})(\text{Age} \vee \text{Sex} \vee \text{LEMS})(\text{Sex} \vee \text{LEMS})(\text{Age} \vee \text{Sex} \vee \text{LEMS})(\text{Age} \vee \text{LEMS}) \\
 &\quad (\text{Age} \vee \text{LEMS}) \\
 &= (\text{LEMS})(\text{Sex})
 \end{aligned}$$

$\{\text{Smith}\}$  is discernable from  $\{\text{Parker}, \text{Hanson}\}$ , but the latter two are never discernable.

## 5. Rule-based classification

a)

- $\{\text{Smith}\} = \text{Yes}$
- $\{\text{Jones}\} = \text{No}$
- $\{\text{Fields}\} = \text{Yes}$
- $\{\text{Parker}, \text{Hanson}\} = \{\text{Yes}, \text{No}\}$
- $\{\text{Moore}, \text{Starr}\} = \text{No}$

The table is inconsistent due to  $\{\text{Parker}, \text{Moore}\}$ .

b)

	Smith	Jones	Parker	Hanson	Moore	Fields	Starr
Smith	$\Theta$						
Jones	LEMS	$\Theta$					
Parker	Age, LEMS	$\Theta$		$\Theta$			
Hanson	$\Theta$		Age, LEMS	$\Theta$		$\Theta$	
Moore	Age, LEMS	Sex, LEMS	Age, LEMS	Sex, $\Theta$	Age, LEMS	Sex, $\Theta$	
Fields	$\Theta$		$\Theta$		Sex, LEMS	$\Theta$	Age, LEMS
Starr	Age, LEMS	Sex, $\Theta$		$\Theta$	Age, LEMS	Sex, $\Theta$	Age, LEMS

c)

Boolean discernability function

$$\begin{aligned}
 &= (\text{LEMS})(\text{Age} \vee \text{LEMS})(\text{Age} \vee \text{Sex} \vee \text{LEMS})(\text{Age} \vee \text{Sex} \vee \text{LEMS})(\text{Age} \vee \text{LEMS})(\text{Age} \vee \text{Sex} \vee \text{LEMS}) \\
 &\quad (\text{Sex} \vee \text{LEMS})(\text{Age} \vee \text{Sex} \vee \text{LEMS})(\text{Age} \vee \text{Sex} \vee \text{LEMS})(\text{Age} \vee \text{LEMS})(\text{Age} \vee \text{LEMS}) \\
 &= (\text{LEMS})
 \end{aligned}$$

The prime implicant is LEMS.

d) Decision rules based on LEMS:

- IF LEMS = 50 THEN Yes
- IF LEMS = 26-49 THEN {Yes, No}
- IF LEMS = 1-25 THEN No
- IF LEMS = 0 THEN No

e) Accuracy for Walk = No:

- IF LEMS = 50 THEN Yes  $\Rightarrow \frac{0}{2}$
- IF LEMS = 26-49 THEN {Yes, No}  $\Rightarrow \frac{1}{2}$
- IF LEMS = 1-25 THEN No  $\Rightarrow \frac{2}{2}$
- IF LEMS = 0 THEN No  $\Rightarrow \frac{1}{1}$

Coverage for Walk = No:

- IF LEMS = 50 THEN Yes  $\Rightarrow \frac{0}{4}$
- IF LEMS = 26-49 THEN {Yes, No}  $\Rightarrow \frac{1}{4}$
- IF LEMS = 1-25 THEN No  $\Rightarrow \frac{2}{4}$
- IF LEMS = 0 THEN No  $\Rightarrow \frac{1}{4}$

Strength for Walk = No:

- IF LEMS = 50 THEN Yes =>  $\frac{2}{7}$
- IF LEMS = 26-49 THEN {Yes, No} =>  $\frac{2}{7}$
- IF LEMS = 1-25 THEN No =>  $\frac{2}{7}$
- IF LEMS = 0 THEN No =>  $\frac{1}{7}$

Support for Walk = No:

- IF LEMS = 50 THEN Yes =>  $\frac{0}{7}$
- IF LEMS = 26-49 THEN {Yes, No} =>  $\frac{1}{7}$
- IF LEMS = 1-25 THEN No =>  $\frac{2}{7}$
- IF LEMS = 0 THEN No =>  $\frac{1}{7}$

## 6.

	Classifier says “Lung” (1)	Classifier says “Colon” (0)
Real origin “Lung” (1)	34	3
Real origin “Colon” (0)	10	22

- a) True positive = 34
- b) True negative = 22
- c) False negative = 3
- d) False positive = 10
- e) Sensitivity =  $\frac{34}{34+3} = \frac{34}{37} \approx 0.92$
- f) Specificity =  $\frac{22}{22+10} = \frac{22}{32} \approx 0.69$
- g)

ROC curve shows separation of positives from negatives.

x-axis shows false positive rate

y-axis shows true positive rate

h)

This classifier is reasonably strong and could probably be used.

i)

Threshold values:

- 0.8 is good
- 0.9 very good
- Close to diagonal is bad

- If very low, flip how we categorize outcomes