Knowledge based systems

Lab 2

1. Formal definitions

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

A = {Age, Sex, LEMS}

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

V_{Sex} = {Male, Female}

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

2. Discernability 1

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

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

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

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

Good model? No.

3. Discernability 2

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

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

4. Discernability 3

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

b)

	Smith	Jones	Parker	Hanson	Moore	Fields	Starr
Smith	Θ						
Jones	LEMS	Θ					
Parker	Age, LEMS	Age, LEMS	Θ				
Hanson	Age, LEMS	Age, LEMS	Θ	Θ			
Moore	Age, Sex, LEMS	Age, Sex, LEMS	Age, Sex, LEMS	Age, Sex, LEMS	Θ		
Fields	Sex	Sex, LEMS	Sex, LEMS	Sex, LEMS	Age, LEMS	Θ	
Starr	Age, Sex, LEMS	Age, Sex, LEMS	Age, Sex, LEMS	Age, Sex, LEMS	Θ	Age, LEMS	Θ

c) Boolean discenability function

= (LEMS)(Age ∨ LEMS)(Age ∨ LEMS)(Age ∨ Sex ∨ LEMS)(Sex)(Age ∨ Sex ∨ LEMS)(Age ∨ LEMS)(Age ∨ LEMS)(Age ∨ Sex ∨ LEMS)(Sex ∨ LEMS)(Age ∨ Sex ∨ LEMS)(Age ∨ Sex ∨ LEMS)(Sex ∨ LEMS)(Age ∨ Sex ∨ LEMS)(Age ∨ Sex ∨ LEMS)(Sex ∨ LEMS)(Age ∨ Sex ∨ LEMS)(Age ∨ LEMS)(Age ∨ LEMS)

= (LEMS)(Sex)

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

5. Rule-based classification

a)

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

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

b)

	Smith	Jones	Parker	Hanson	Moore	Fields	Starr
Smith	Θ						
Jones	LEMS	Θ					
Parker	Age, LEMS	Θ	Θ				
Hanson	Θ	Age, LEMS	Θ	Θ			
Moore	Age, Sex, LEMS	Age, Sex, LEMS	Θ	Age, Sex, LEMS	Θ		
Fields	Θ	Θ	Sex, LEMS	Θ	Age, LEMS	Θ	
Starr	Age, Sex, LEMS	Θ	Θ	Age, Sex, LEMS	Θ	Age, LEMS	Θ

c)

Boolean discenability function

= (LEMS)(Age ∨ LEMS)(Age ∨ Sex ∨ LEMS)(Age ∨ Sex ∨ LEMS)(Age ∨ LEMS)(Age ∨ Sex ∨ LEMS)(Sex ∨ LEMS)(Age ∨ Sex ∨ LEMS)(Age ∨ Sex ∨ LEMS)(Age ∨ LEMS)(Age ∨ LEMS)

= (LEMS)

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

Coverage for Walk = No:

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

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