**INTELLIGENT SYSTEMS M – Michela Milano**
**February 15, 2018**

**Exercise 1 (8 points)**
Given the following training set:

| **Attachments** | **SuspiciousWords** | **Folder** |
|---|---|---|
| Morethan2 | Yes | Spam |
| Exactly1 | ? | NoSpam |
| Morethan1 | Yes | Spam |
| Morethan2 | No | Spam |
| Morethan1 | No | NoSpam |
| Exactly1 | Yes | Spam |
| Exactly1 | No | NoSpam |
| Morethan1 | Yes | Spam |
| Exactly1 | ? | Spam |
| Morethan2 | No | NoSpam |
| Morethan2 | No | NoSpam |
| Morethan1 | Yes | Spam |
| Exactly1 | Yes | NoSpam |

a) Compute the entropy of the training set w.r.t. the attribute Folder
b) Compute the gain of the two attributes with respect to these training examples
c) Build the decision tree with one level for the training set, and compute the labels of each leaf.
d) Classify the instance:

| | |
|---|---|
| Exactly1 | ? |

**Exercise 2 (8 points)**
In a chemical experiment, we have five liquid substances $L_1$, …, $L_5$ that have to be stored in an array of 10 recipients $R_1,\dots,R_{10}$.
We have a known amount $Q_i$ of each liquid and each recipients has a given capacity $R\_cap_j$. Each recipients can contain a single liquid which can be stored in more than one tank (depending on its quantity).
Each liquid i should be stored in a tank j at a given temperature $temp_j$ that should stay within $tmin_i$ and $tmax_i$ where $tmin_1$= $tmin_2$= $tmin_3$= 20C, $tmax_1$= $tmax_2$= $tmax_3$= 30C, $tmin_4$= $tmin_5$= 5C, $tmax_4$= $tmax_5$= 15C.

Liquids 1, and 3 cannot be stored in adjacent recipients as they can trigger dangerous chemical reactions. In addition, if two liquids are stored in two recipients that differ of more than 10 degrees they cannot be stored in adjacent recipients.
Every four adjacent recipients can contain the same chemical maximum twice.
We have to define an assignment of liquids to recipients along with their temperature that is consistent with the above constraints. Model the problem as a CSP, with variables, domains and constraints. Make use of reified constraints and global constraints.

**Exercise 3 (8 points)**

Given the following initial state

**robot_at(location_a), handempty, colour(green), colour(yellow), object_at(ball, location_a), object_at(cube, location_b), connected (location_a, location_c), connected (location_c, location_b), connected (location_c, location_a), connected (location_b, location_c)**

We have to reach the goal: **coloured(ball, green), coloured(cube,red)**

Actions are modelled as follows:

**colour(Object, Location, Colour)**
PRECOND: object_at(Object, Location), robot_at(Location), robot_has(Colour)
ADD: coloured(Object, Colour)

**go(X,Y)**
PRECOND: robot_at(X)
DELETE: robot_at(X)
ADD: robot_at(Y)

**loadColour(C)**
PRECOND: handempty
DELETE: handempty
ADD: robot_has(C)

**releaseColour(C)**
PRECOND: robot_has(C)
DELETE: robot_has(C)
ADD: handempty

Solve the problem with the POP algorithm, identifying threats and their solution during the process.

**Exercise 4 (7 points)**

1) Compute the lgg of the two clauses

```
C1= f(q(a))  ←   c(X,a), c(q(X),b).
C2= f(q(Z))  ←   c(a,Z), c(q(Z),Y), c(r(Z),a)
```

2) Build two levels of graph plan for the exercise 3
3) What is the main features of swarm intelligent algorithms.
4) What convolutional networks have in common with other networks and what are the distinguishing features ?
5) What are the main features of inductive logic programming?

## SOLUZIONE

**Esercizio 1**

a) $info(S)=-7/14*\log_2 7/14-7/14*\log_2 7/14= 1$

b)

$info_{Laurea}(S)= 7/14*(-3/7*\log_2 3/7-4/7*\log_2 4/7) + 7/14*(-4/7*\log_2 4/7-3/7*\log_2 3/7) = 0.5*0.985+0.5*0.985=0.985$

$gain(Laurea) =1 -0.985=0.015$

$splitinfo(Laurea)= - 7/14*\log_2(7/14) - 7/14*\log_2(7/14)= 1$

$gainratio(Laurea) =0.015/ 1=0.015$

Per calcolare il guadagno dell'attributo Occupazione non si usa l'entropia calcolata su tutto il training set ma solo sugli esempi che hanno Occupazione noto (insieme F):

$info(F)=-6/12*\log_2 6/12- 6/12*\log_2 6/12=1$

$info_{Occupazione}(F)=4/12*(-1/4*\log_2 1/4-3/4*\log_2 3/4)+4/12(-4/4*\log_2 4/4-0/4*\log_2 0/4) +4/12(-1/4*\log_2 1/4-3/4*\log_2 3/4)=0.333 *0.811 +0.333 *0+0.333 *0.811=0.540$

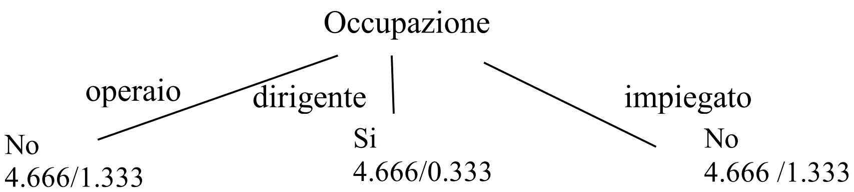
$gain(Occupazione) =12/14*(1-0.540) =0.394$

$splitinfo(Occupazione) =-6/14*\log_2(6/14)-6/14*\log_2(6/14)-2/14*\log_2(2/14)= 1.449$

$gainratio(Occupazione) =0.394/1.449=0.272$

c) L'attributo scelto per la radice dell'albero è Occupazione

.




d) l'istanza viene divisa in tre parti, di peso rispettivamente 4.666/14=0.333, 4.666/14=0.333 e 4.666/14=0.333. La prima parte viene mandata lungo il ramo operaio e classificata come No con probabilità 3.333/4.666=71.4% e come Si con probabilità 1.333/4.666=28.6%. La seconda parte viene mandata lungo il ramo dirigente e classificata come Si con probabilità 4.333/4.666=92.9% e come No con probabilità 0.333/4.666=7.1%. La terza parte viene mandata lungo il ramo impiegato e classificata come No con probabilità 3.333/4.666=71.4% e come Si con probabilità 1.333/4.666=28.6%. Quindi in totale la classificazione dell'istanza è

$P(Si) =0.333*28.6\%+0.333*92.9\%+0.333*28.6\%=0.5$

$P(No) = 0.333*28.6\%+0.333*7.1\%+0.333*28.6\%=0.5$

**Esercizio 2**

C1= samebib(a,b) ←author(a,a(a)),author(b,a(a))
C2= samebib(a,p(b)) ←author(a,a(X)),author(p(b),a(Y))

lgg(C1,C2)=samebib(a,A) ← author(a,a(B)),author(C,a(D)),author(E,a(B)),author(A,a(D)).

A/b,p(b)
B/a,X
C/a,p(b)
D/a,Y
E/b,a
F/b,p(b)