1 First-order logic (10 pts)

1.1 Task

Express the following natural language sentences by means of formulas of first-order logic:

a) *Gauss is a smart mathematician.*

b) *Mathematicians are not smart.*

c) *Not all mathematicians are smart.*

d) *Except for Gauss, no mathematician is smart.*

e) $2x$ *is positive if* x *is positive and negative if* x *is negative.*

In doing so, use exclusively the following predicates/functions:

- g: the constant representing Gauss
- $s(x)$: $\top$ iff x is smart
- $m(x)$: $\top$ iff x is a mathematician
- 0: the constant 0
- 1: the constant 1
- $f < h$: $\top$ iff f is lower than h
- $f = h$: $\top$ iff f is equal to h
- $x + y$: the sum of x and y

1.2 Solution

a) $m(g) \land s(g)$

b) $\forall x(m(x) \rightarrow \neg s(x))$

c) $\exists x(m(x) \land \neg s(x))$

d) $m(g) \land s(g) \land \forall x(m(x) \land s(x) \rightarrow x = g)$

e) $\forall x((0 < x \rightarrow 0 < x + x) \land (x < 0 \rightarrow x + x < 0))$

2 Resolution (12 pts)

2.1 Task

We are given the following knowledge base:

(a) $(B \rightarrow C) \land (\neg C \rightarrow \neg B)$

(b) $A \land B \rightarrow D$

(c) $(A \rightarrow B) \lor (C \rightarrow B)$

Furthermore, we are given the conjecture

(d) $(B \rightarrow A) \lor (\neg C \rightarrow (D \rightarrow \neg D))$

Use the resolution algorithm to prove or disprove (d).

Which of the following statements are true:

1) (d) can be proven using the knowledge base.
2) (d)'s negation can be proven using the knowledge base.
3) The knowledge base is inconsistent.
4) The knowledge base is incomplete.

2.2 Solution

Transforming knowledge base, conjecture, and negated conjecture into CNF:

$$
\begin{aligned}
(a) \quad &\Leftrightarrow \quad (\neg B \vee C) \wedge (C \vee \neg B) \\
&\Leftrightarrow \quad \{\neg B, C\} \\
(b) \quad &\Leftrightarrow \quad \neg(A \wedge B) \vee D \\
&\Leftrightarrow \quad \neg A \vee \neg B \vee D \\
&\Leftrightarrow \quad \{\neg A, \neg B, D\} \\
(c) \quad &\Leftrightarrow \quad (A \rightarrow B) \vee (C \rightarrow B) \\
&\Leftrightarrow \quad (\neg A \vee B) \vee (\neg C \vee B) \\
&\Leftrightarrow \quad \{\neg A, B, \neg C\} \\
(d) \quad &\Leftrightarrow \quad (B \rightarrow A) \vee (\neg C \rightarrow (D \rightarrow \neg D)) \\
&\Leftrightarrow \quad (\neg B \vee A) \vee (C \vee (\neg D \vee \neg D)) \\
&\Leftrightarrow \quad \{A, \neg B, C, \neg D\} \\
\neg(d) \quad &\Leftrightarrow \quad \underbrace{\{\neg A\}}_{(e)}, \underbrace{\{B\}}_{(f)}, \underbrace{\{\neg C\}}_{(g)}, \underbrace{\{D\}}_{(h)}
\end{aligned}
$$

Trying to prove (d) by joining knowledge base and (d)'s negation and attempting to derive contradiction by means of the resolution algorithm:

$$
\begin{aligned}
(a) \wedge (f) \rightarrow \{C\} \quad & (i) \\
(g) \wedge (i) \rightarrow \{\} \quad & \Box
\end{aligned}
$$

Trying to prove (d)'s negation by joining knowledge base and (d) and attempting to derive contradiction by means of the resolution algorithm:

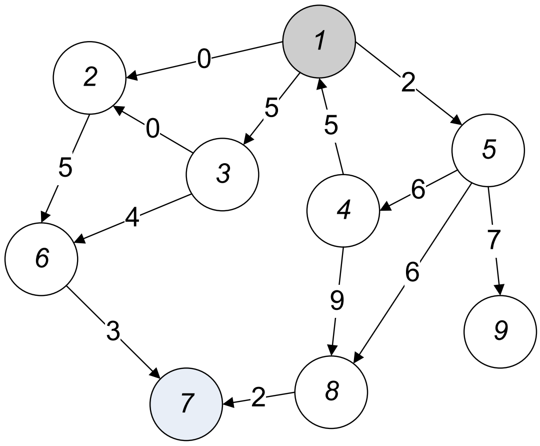
$$
(b) \wedge (c) \rightarrow \{\neg A, \neg C, D\} \quad (j)
$$

This time, no contradiction can be derived. Hence, we can conclude that the knowledge base is consistent and complete and that (d) is true, i.e., 1) is the right answer.

3 A* (12 pts)

3.1 Task

We are given the following graph:



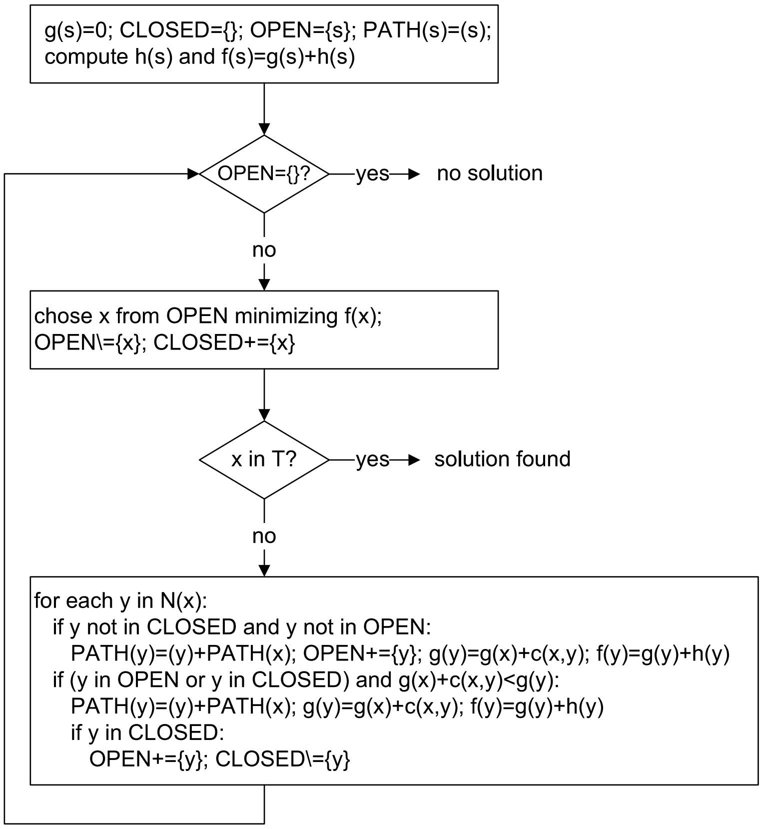
From the figure, we can derive the starting node (1) and the set of target nodes ($T = \{7\}$). Furthermore, we can derive the set of neighbors $N(x)$ for each node x as well as the cost $c(x, y)$ for changing from node x to node y being the arc labels.

Perform an A* search using the following table providing the heuristic estimate $h(x)$:

x	1	2	3	4	5	6	7	8	9
$h(x)$	4	9	0	8	5	2	2	8	8

How do OPEN and CLOSED sets as well as PATH(7) look like when hitting the target?

For this task, use the A^* convention introduced in the class as depicted in the following diagram:



3.2 Solution

ID	step	x	$N(x)$	PATH(x)	OPEN	CLOSED	$g(x)$	$f(x)$
1	init	1	{2,3,5}	(1)	{1}	{}	0	4
2	min	1	{2,3,5}	(1)	{}	{1}	0	4
3	iter	2	{6}	(2,1)	{2}	{1}	0	9
4		3	{2,6}	(3,1)	{2,3}	{1}	5	5
5		3	{4,8,9}	(5,1)	{2,3,5}	{1}	2	7
6	min	3	{2,6}	(3,1)	{2,5}	{1,3}	5	5
7	iter	2	$g(x) + c(x, y) \geq g(y) \rightarrow$ skipped					
8		6	{7}	(6,3,1)	{2,5,6}	{1,3}	9	11
9	min	5	{4,8,9}	(5,1)	{2,6}	{1,3,5}	2	7
10	iter	4	{1,8}	(4,5,1)	{2,4,6}	{1,3,5}	8	16
11		8	{7}	(8,5,1)	{2,4,6,8}	{1,3,5}	8	16
12		9	{}	(9,5,1)	{2,4,6,8,9}	{1,3,5}	9	17
13	min	2	{6}	(2,1)	{4,6,8,9}	{1,2,3,5}	0	9
14	iter	6	$g(x) + c(x, y) < g(y) \rightarrow$ reroute					
			{7}	(6,2,1)	{4,6,8,9}	{1,2,3,5}	5	7
15	min	6	{7}	(6,2,1)	{4,8,9}	{1,2,3,5,6}	5	7
16	iter	7	{}	(7,6,2,1)	{4,7,8,9}	{1,2,3,5,6}	8	10
17	min	7	{}	(7,6,2,1)	{4,8,9}	{1,2,3,5,6,7}	8	10

4 Expert and dialog systems (6 pts)

4.1 Task

The interactive voice response system of a power company is to be optimized. Two designs are considered:

I. The system asks

> *What can I help you with: Your bill, tech support, or new services?*

If callers respond *bill* or *new services*, they get escalated to an agent. In case of *tech support*, they get asked whether they experience a *power outage* before getting escalated.

II. The system asks

> *Are you calling because of a power outage?*

If callers say *yes*, they get escalated to a human agent, otherwise the system continues

> *Then, what can I help you with: Your bill, tech support, or new services?*

After collecting the callers' response, the system escalates to an agent.

This is the distribution of some of the symptoms among the caller population:

symptom	ratio
bill	40%
new services	20%
outage	20%

a) What is the average number of questions asked for each of the designs before escalating to an agent?

b) How does the situation in a) change during an outage situation where 80% of the callers call because of their power being down and 10% for other technical difficulties?

c) For which percentage of outage callers in b) would both of the designs perform on par with each other (all other parameters remaining the same)?

4.2 Solution

Terminology: $p(k)$ is the probability that k questions are asked.

a) In Design I, *bill* and *new services* result in one question, i.e., $p_I(1) = 40\% + 20\% = 0.6$. *Tech support* requires two questions and accounts for the rest of the caller population: $p_I(2) = 0.4$. The average number of questions is the expected value

$$E_I = \sum_k kp_I(k) = p_I(1) + 2p_I(2) = 1.4.$$

In Design II, only when there is an *outage* callers get escalated after one question, i.e., $p_{II}(1) = 0.2$. The rest of the cases require two questions: $p_{II}(2) = 80\% = 0.8$. The average number of questions is $E_{II} = 1.8$.

b) In Design I, *outage* and other *tech support* symptoms require two questions: $p'_I(2) = 80\% + 10\% = 0.9$. The rest of the cases require only one question: $p'_I(1) = 0.1$. The average number of questions is $E'_I = 1.9$.

The average number of questions for Design II can be calculated the same way as in a): $E'_{II} = 1.2$.

c) Using the generic variable x for the fraction of outage callers, we know for Design I that $p^x_I(2) = x + 0.1$ and $p^x_I(1) = 1 - p^x_I(2)$, so we have

$$E^x_I = 1 - (x + 0.1) + 2 \cdot (x + 0.1) = x + 1.1.$$

For Design II, we have $p^x_{II}(1) = x$ and $p^x_{II}(2) = 1 - p^x_{II}(1)$ resulting in

$$E^x_{II} = x + 2 \cdot (1 - x) = 2 - x.$$

Setting $E^x_I = E^x_{II}$, we get the solution $x = 0.45$. I.e., for 45% outage callers, both designs would perform the same.