11/22
11/29

Solution to Exercise 4 Artificial Intelligence

Problem 4.1: Model, satisfaction relation, and entailment

Notice that False and True are sentence in propositional logic. Symbol **correct** denotes the statement is correct and symbol **incorrect** denotes the statement is incorrect.

1. $\mathsf{False} \models \mathsf{True}$

* Entailment: is the relationship between 2 sentences where the truth of one sentence requires the truth of the other sentence: $\alpha \models \beta$.

$\alpha \models \beta$ iff $M(\alpha) \subseteq M(\beta)$

i.e β has to be true for α to be true.

$\mathsf{False} \models \mathsf{True}$
= ⟨by definition of entailment⟩
$M(\mathsf{False}) \subseteq M(\mathsf{True})$
= ⟨by model of False is empty set⟩
$\emptyset \subseteq M(\mathsf{True})$
⇒ ⟨by $\emptyset$ is subset of any set⟩
correct

False never evaluates to True!

2. $\mathsf{True} \models \mathsf{False}$

$\mathsf{True} \models \mathsf{False}$
= ⟨by definition of entailment⟩
$M(\mathsf{True}) \subseteq M(\mathsf{False})$
= ⟨by model of False is empty set⟩
$M(\mathsf{True}) \subseteq \emptyset$
⇒ ⟨by $M(\mathsf{True})$ is the set of all possible worlds; hence $\neq \emptyset$⟩
incorrect

* $\boxed{\text{Remember that } S \subseteq \emptyset \text{ if and only if } S = \emptyset}$

3. $(A \land B) \models (A \Leftrightarrow B)$

$(A \land B) \models (A \Leftrightarrow B)$
= ⟨by definition of entailment⟩
$M(A \land B) \subseteq M(A \Leftrightarrow B)$
= ⟨by model of $(A \land B)$ is the truth value assignment where $(A \land B)$ is correct. We use tuple $(\text{true}, \text{false})$ to denote the assignment of *true* to A and *false* to B. Therefore, $M(A \land B) = \{(\text{true}, \text{true})\}$⟩
$\{(\textbf{true}, \textbf{true})\} \subseteq M(A \Leftrightarrow B)$
= ⟨by model of $(A \Leftrightarrow B)$ is $\{(\textbf{true}, \textbf{true}), (\textbf{false}, \textbf{false})\}$ ⟩
$\{(\textbf{true}, \textbf{true})\} \subseteq \{(\textbf{true}, \textbf{true}), (\textbf{false}, \textbf{false})\}$
⇒ ⟨by subset definition⟩
correct

Truth Table

A	B	$A \land B$	$A \Leftrightarrow B$
T	T	T ←	T ←
T	F	F	F
F	T	F	F
F	F	F	T ←

4. $A \Leftrightarrow B \models A \lor B$

$$\begin{aligned}
& A \Leftrightarrow B \models A \lor B \\
= \quad & \langle \text{by definition of entailment} \rangle \\
& M(A \Leftrightarrow B) \subseteq M(A \lor B) \\
= \quad & \langle \text{by model of } (A \Leftrightarrow B) \text{ is } \{(\text{true}, \text{true}), (\text{false}, \text{false})\} \rangle \\
& \{(\text{true}, \text{true}), (\text{false}, \text{false})\} \subseteq M(A \lor B) \\
= \quad & \langle \text{by model of } (A \lor B) \text{ is } \{(\text{true}, \text{true}), (\text{true}, \text{false}), (\text{false}, \text{true})\} \rangle \\
& \{(\text{true}, \text{true}), (\text{false}, \text{false})\} \subseteq \{(\text{true}, \text{true}), (\text{true}, \text{false}), (\text{false}, \text{true})\} \\
\Rightarrow \quad & \langle \text{by subset definition} \rangle \\
& \textbf{incorrect}
\end{aligned}$$

5. $A \Leftrightarrow B \models \neg A \lor B$

$$\begin{aligned}
& A \Leftrightarrow B \models \neg A \lor B \\
= \quad & \langle \text{by definition of entailment} \rangle \\
& M(A \Leftrightarrow B) \subseteq M(\neg A \lor B) \\
= \quad & \langle \text{by model of } (A \Leftrightarrow B) \text{ is } \{(\text{true}, \text{true}), (\text{false}, \text{false})\} \rangle \\
& \{(\text{true}, \text{true}), (\text{false}, \text{false})\} \subseteq M(\neg A \lor B) \\
= \quad & \langle \text{by model of } (\neg A \lor B) \text{ is } \{(\text{true}, \text{true}), (\text{false}, \text{true}), (\text{false}, \text{false})\} \rangle \\
& \{(\text{true}, \text{true}), (\text{false}, \text{false})\} \subseteq \{(\text{true}, \text{true}), (\text{false}, \text{true}), (\text{false}, \text{false})\} \\
\Rightarrow \quad & \langle \text{by subset definition} \rangle \\
& \textbf{correct}
\end{aligned}$$

Problem 4.2: Validity, satisfiability, and unsatisfiability

A sentence α is **valid** if it is true in *all* models. A sentence α is **satisfiable** if it is true in *some* models.

Problem 4.2.1:

1. Sentence α is valid if and only if $\alpha \equiv \text{True}$

Since $\alpha \equiv$ true

hence $(\alpha \models \text{true}) \land (\text{true} \models \alpha)$

$$\begin{aligned}
& \alpha \text{ is valid} \\
= \quad & \langle \text{by definition of validity} \rangle \\
& M(\alpha) = \text{AllPossWorlds} \\
= \quad & \langle \text{by } M(\text{True}) = \text{AllPossWorlds} \rangle \\
& M(\alpha) = M(\text{True}) \\
= \quad & \langle \text{by definition of set equality} \rangle \\
& M(\alpha) \subseteq M(\text{True}) \land M(\text{True}) \subseteq M(\alpha) \\
= \quad & \langle \text{by definition of entailment} \rangle \\
& \alpha \models \text{True} \land \text{True} \models \alpha \\
= \quad & \langle \text{by definition of logical equivalence} \rangle \\
& \alpha \equiv \text{True}
\end{aligned}$$

2. Sentence α is unsatisfiable if and only if $\alpha \equiv \text{False}$

$$
\begin{aligned}
& \alpha \text{ is unsatisfiable} \\
= & \quad \langle \text{by definition of satisfiability} \rangle \\
& M(\alpha) = \emptyset \\
= & \quad \langle \text{by } M(\text{False}) = \emptyset \rangle \\
& M(\alpha) = M(\text{False}) \\
= & \quad \langle \text{by definition of equality} \rangle \\
& M(\alpha) \subseteq M(\text{False}) \wedge M(\text{False}) \subseteq M(\alpha) \\
= & \quad \langle \text{by definition of entailment} \rangle \\
& \alpha \models \text{False} \wedge \text{False} \models \alpha \\
= & \quad \langle \text{by definition of logical equivalence} \rangle \\
& \alpha \equiv \text{False}
\end{aligned}
$$

Since α is unsatisfiable

hence $M(\alpha) = 0$

Problem 4.2.2:

Given $\alpha \vee \neg\alpha \equiv \text{true}$

$\alpha \wedge \alpha \equiv \alpha$

1. Smoke $\Rightarrow$ Smoke

$$
\begin{aligned}
& \text{Smoke} \Rightarrow \text{Smoke} \\
\equiv & \quad \langle \text{by implication elimination } \alpha \Rightarrow \beta \equiv \neg\alpha \vee \beta \rangle \\
& \neg\text{Smoke} \vee \text{Smoke} \\
\equiv & \quad \langle \text{by excluded middle rule } \alpha \vee \neg\alpha \equiv \text{True} \rangle \\
& \text{True}
\end{aligned}
$$

Since we have shown Smoke $\Rightarrow$ Smoke $\equiv$ True, and by property in Problem 4.2.2 above, we conclude it is valid. Since it is valid, it must be satisfiable as well.

2. (Smoke $\Rightarrow$ Fire) $\Rightarrow$ ($\neg$Smoke $\Rightarrow$ $\neg$Fire)

$$
\begin{aligned}
& (\text{Smoke} \Rightarrow \text{Fire}) \Rightarrow (\neg\text{Smoke} \Rightarrow \neg\text{Fire}) \\
\equiv & \quad \langle \text{by implication elimination } \alpha \Rightarrow \beta \equiv \neg\alpha \vee \beta \rangle \\
& \neg(\text{Smoke} \Rightarrow \text{Fire}) \vee (\neg\text{Smoke} \Rightarrow \neg\text{Fire}) \\
\equiv & \quad \langle \text{by implication elimination } \alpha \Rightarrow \beta \equiv \neg\alpha \vee \beta \rangle \\
& \neg(\neg\text{Smoke} \vee \text{Fire}) \vee (\text{Smoke} \vee \neg\text{Fire}) \\
\equiv & \quad \langle \text{by De Morgan rule's } \neg(\alpha \vee \beta) \equiv \neg\alpha \wedge \neg\beta \rangle \\
& (\text{Smoke} \wedge \neg\text{Fire}) \vee (\text{Smoke} \vee \neg\text{Fire}) \\
\equiv & \quad \langle \text{by commutativity of } \vee \rangle \\
& (\text{Smoke} \vee \neg\text{Fire}) \vee (\text{Smoke} \wedge \neg\text{Fire}) \\
\equiv & \quad \langle \text{by distributivity of } \vee \text{ over } \wedge \text{ that is } \alpha \vee (\beta \wedge \gamma) \equiv (\alpha \vee \beta) \wedge (\alpha \vee \gamma) \text{ with } \alpha := (\text{Smoke} \vee \neg\text{Fire}) \rangle \\
& (\text{Smoke} \vee \neg\text{Fire}) \wedge (\text{Smoke} \vee \neg\text{Fire}) \\
\equiv & \quad \langle \text{by rule } \alpha \wedge \alpha \equiv \alpha \rangle \\
& \text{Smoke} \vee \neg\text{Fire}
\end{aligned}
$$

$(S \wedge \neg F) \vee S \vee \neg F$

$(S \vee S \vee \neg F) \wedge (\neg F \vee S \vee \neg F)$

$(S \vee \neg F) \wedge (S \vee \neg F)$

Smoke $\vee \neg$ Fire

It is not valid and not unsatisfiable because Smoke $\vee \neg$Fire $\not\equiv$ True and Smoke $\vee \neg$Fire $\not\equiv$ False, respectively. Since it is not valid and not unsatisfiable, then Smoke $\vee \neg$Fire is satisfiable.

there could be models where it's true & models where it's false

3. Smoke $\vee$ Fire $\vee \neg$Fire

Smoke $\vee$ (fire $\vee \neg$fire) $\equiv$
Smoke $\vee$ true $\equiv$
true

$$\begin{aligned}
& \text{Smoke} \vee \text{Fire} \vee \neg \text{Fire} \\
\equiv\ & \langle \text{by excluded middle rule } \alpha \vee \neg\alpha \equiv \text{True} \rangle \\
& \text{Smoke} \vee \text{True} \\
\equiv\ & \langle \text{by true rule } \alpha \vee \text{True} \equiv \text{True} \rangle \\
& \text{True}
\end{aligned}$$

Since we have shown Smoke $\vee$ Fire $\vee \neg$Fire $\equiv$ True, then the sentence is valid. Since it is valid, it is also satisfiable.

4. (Fire $\Rightarrow$ Smoke) $\wedge$ Fire $\wedge \neg$Smoke

$$\begin{aligned}
& (\text{Fire} \Rightarrow \text{Smoke}) \wedge \text{Fire} \wedge \neg\text{Smoke} \\
\equiv\ & \langle \text{by implication elimination } \alpha \Rightarrow \beta \equiv \neg\alpha \vee \beta \rangle \\
& (\neg\text{Fire} \vee \text{Smoke}) \wedge \text{Fire} \wedge \neg\text{Smoke} \\
\equiv\ & \langle \text{by distributivity of } \wedge \text{ over } \vee \text{ that is } \alpha \wedge (\beta \vee \gamma) \equiv (\alpha\wedge\beta) \vee (\alpha \wedge \gamma) \rangle \\
& ((\neg\text{Fire} \wedge \text{Fire}) \vee (\text{Smoke} \wedge \text{Fire})) \wedge \neg\text{Smoke} \\
\equiv\ & \langle \text{by false rule } \alpha \wedge \neg\alpha \equiv \text{False} \rangle \\
& (\text{False} \vee (\text{Smoke} \wedge \text{Fire})) \wedge \neg\text{Smoke} \\
\equiv\ & \langle \text{by false rule } \alpha \vee \text{False} \equiv \alpha \rangle \\
& \text{Smoke} \wedge \text{Fire} \wedge \neg\text{Smoke} \\
\equiv\ & \langle \text{by false rule } \alpha \wedge \neg\alpha \equiv \text{False} \rangle \\
& \text{False} \wedge \text{Fire} \\
\equiv\ & \langle \text{by false rule } \alpha \wedge \text{False} \equiv \text{False} \rangle \\
& \text{False}
\end{aligned}$$

$(\neg F \wedge F \wedge \neg S) \vee (S \wedge F \wedge \neg S)$

(false $\wedge \neg$Smoke) $\vee$ (false $\wedge$ fire)
false $\vee$ false
false

Since we have shown that (Fire $\Rightarrow$ Smoke) $\wedge$ Fire $\wedge \neg$Smoke $\equiv$ False, then the sentence is unsatisfiable.

Problem 4.3: Knights and Knaves

proof by inference rule

Problem 4.3.1: If A is a knight (A is True), then A must tell the truth. Therefore, B must a knight too (B is True). If A is a knave (A is False), then A lies. Therefore, B must be a knave too (B is False). It is true when both are true or both are false (when A and B are of the same type). The proper model for this sentence is

$$A \Leftrightarrow B. \tag{1}$$

Problem 4.3.2: Sentence Remark has the interpretation of either true or false, and so does propositional variable P. If prop. variable P is interpreted as true, it has the intended meaning that P is a knight; hence Remark must be interpreted as true too, since knights only tell the truth. Similarly, if prop. variable P is interpreted as false, then P is a knave and the Remark must be false; knaves only tell the lies. Hence, as a generalisation whenever P says Remark, we model it as

$$P \Leftrightarrow \text{Remark}. \tag{2}$$

Problem 4.3.3:

1. $B \Leftrightarrow (A \Leftrightarrow \neg A)$
2. $C \Leftrightarrow \neg B$

Problem 4.3.4:

1. The identity of B can be deduced from the first sentence:

$$B \Leftrightarrow (A \Leftrightarrow \neg A) \equiv B \Leftrightarrow \text{False} \equiv \neg B \tag{3}$$

   From the model of B, this means that **B is a knave**.

2. With this knowledge, the identity of C can be deduced from the second sentence:

$$\frac{C \Leftrightarrow \neg B \qquad \neg B}{C}$$

   From the model of C, this means that **C is a knight**.

Problem 4.4: Superman does not exist.

Proof by resolution

Problem 4.4.1:

* if-then statements are usually modelled with implication

1. If Superman were able and willing to prevent evil, he would do so.

$$A \wedge W \Rightarrow P \tag{4}$$

2. If Superman were unable to prevent evil, he would be impotent.

$$\neg A \Rightarrow I \tag{5}$$

3. If he were unwilling to prevent evil, he would be malevolent.

$$\neg W \Rightarrow M \tag{6}$$

4. Superman does not prevent evil.

$$\neg P \tag{7}$$

5. If Superman exists, he is neither impotent nor malevolent.

$$E \Rightarrow \neg I \wedge \neg M \tag{8}$$

Problem 4.4.2: What we want to prove is $\alpha \equiv \neg E$ and the knowledge base KB is the conjunction of all sentences above. The resolution principle is based on the following theorem.

Theorem 1. *For any two propositional sentence α and β, $\alpha \models \beta$ if and only if $\alpha \wedge \neg\beta$ is unsatisfiable.*

Notice that this is exactly proof by contradiction (*reductio ad absurdum*)!!

What we should do next is to find the clausal representation (CNF) of $KB \wedge \neg\alpha \equiv KB \wedge E$.

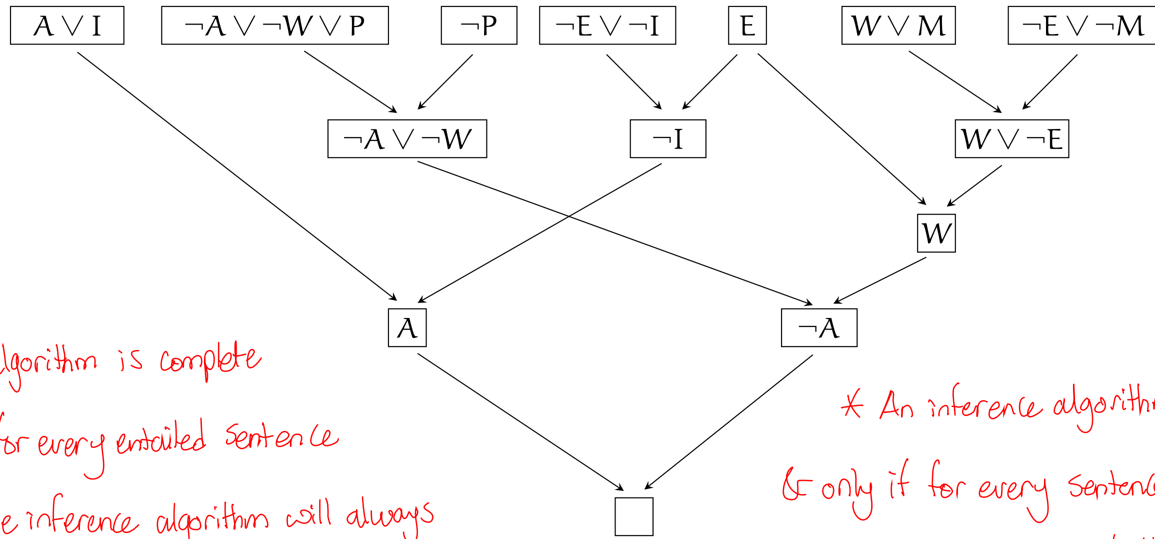
You can do every clause seperately [exe_4_notes, page 31]

$$\begin{aligned}
& (A \wedge W \Rightarrow P) \wedge (\neg A \Rightarrow I) \wedge (\neg W \Rightarrow M) \wedge (\neg P) \wedge (E \Rightarrow (\neg I \wedge \neg M)) \wedge (E) \\
= \quad & \langle \text{by Eliminate } \alpha \Rightarrow \beta \text{ with } \neg\alpha \vee \beta \rangle \\
& (\neg(A \wedge W) \vee P) \wedge (A \vee I) \wedge (W \vee M) \wedge (\neg P) \wedge (\neg E \vee (\neg I \wedge \neg M)) \wedge (E) \\
= \quad & \langle \text{by Eliminate } \neg(\alpha \wedge \beta) \text{ with } \neg\alpha \vee \neg\beta \rangle \\
& (\neg A \vee \neg W \vee P) \wedge (A \vee I) \wedge (W \vee M) \wedge (\neg P) \wedge (\neg E \vee (\neg I \wedge \neg M)) \wedge (E) \\
= \quad & \langle \text{by Eliminate } \alpha \vee (\beta \wedge \gamma) \text{ with } (\alpha \vee \beta) \wedge (\alpha \vee \gamma) \rangle \\
& (\neg A \vee \neg W \vee P) \wedge (A \vee I) \wedge (W \vee M) \wedge (\neg P) \wedge (\neg E \vee \neg I) \wedge (\neg E \vee \neg M) \wedge (E)
\end{aligned}$$

The set of clause are:

1. $\neg A \vee \neg W \vee P$
2. $A \vee I$
3. $W \vee M$
4. $\neg P$
5. $\neg E \vee \neg I$
6. $\neg E \vee \neg M$
7. E

Problem 4.4.3:



* An inference algorithm is complete if & only if for every entailed sentence $KB \models \alpha$, the inference algorithm will always be able to derive it $KB \vdash \alpha$.

* An inference algorithm is sound if & only if for every sentence it derives $KB \vdash \alpha$, it is guaranteed that the sentence is entailed $KB \models \alpha$.

Problem 4.5: Completeness and soundness

Problem 4.5.1: This inference algorithm is **complete**. It is because for every entailed sentence, this algorithm will always be able to derive it. However, this inference algorithm is **unsound**, because it can derive a sentence that is not entailed.

Problem 4.5.2: This inference algorithm is **incomplete**. It is because for every entailed sentence, this algorithm will always be unable to derive it. However, this inference algorithm is **sound** since it never derive any sentence.

* Soundness is the property of only being able to prove "true" things.
A system is sound if & only if the inference rules of the system admit only valid formulas.
Or another way, if we start with valid premises, the inference rules do not allow an invalid conclusion to be drawn.

* Completeness is the property of being able to prove "all" true things.
A system is valid if & only if all valid formula can be derived from the axioms & the inference rules. So there are no valid formula that we can't prove.