CS 4613 **Fall 2023**
HW 4 Solutions **E. K. Wong**

Total # questions = 6. Total # points = 90.

1. [5 points] Any two methods from the list: model checking, truth-table enumeration, resolution, modus ponens, forward chaining, backward chaining.

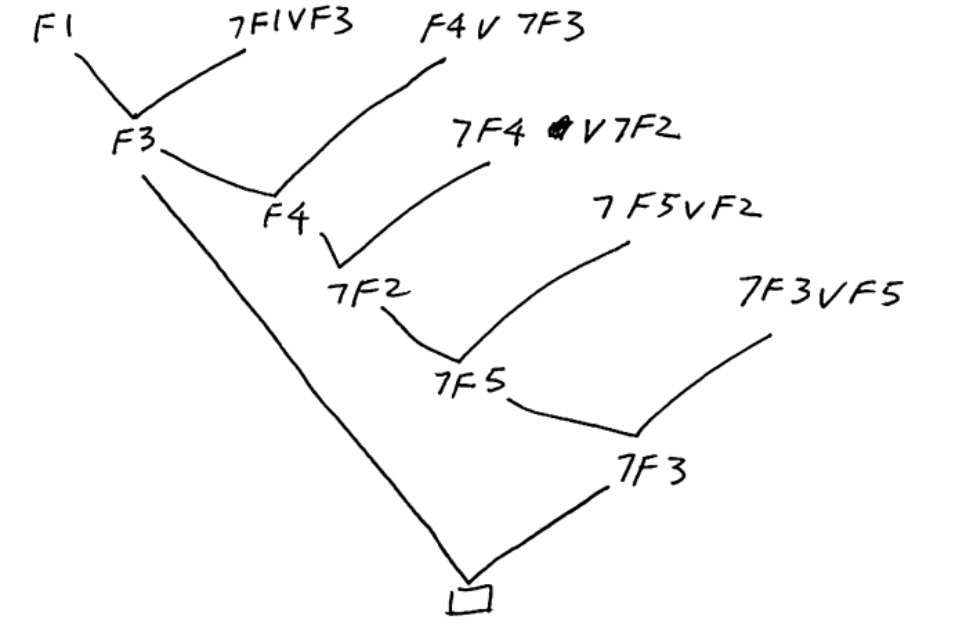
2. [15 points] Conversion to clauses

KB:

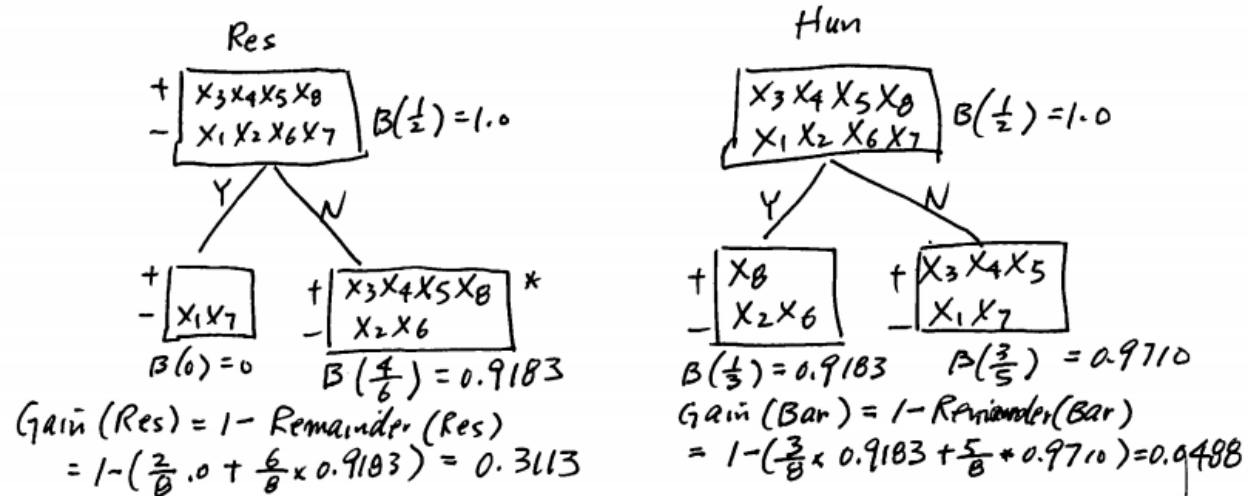
- $F1 \wedge (\neg F4 \Leftrightarrow \neg F3)$
  $F1 \wedge (\neg F4 \Rightarrow \neg F3) \wedge (\neg F3 \Rightarrow \neg F4)$
  $F1 \wedge (F4 \vee \neg F3) \wedge (F3 \vee \neg F4)$
  $F1$, $F4 \vee \neg F3$, $F3 \vee \neg F4$
- $\neg F1 \vee F3$
- $F4 \Rightarrow \neg F2$
  $\neg F4 \vee \neg F2$
- $F3 \Rightarrow F5$
  $\neg F3 \vee F5$

$\neg G$: $\neg(\neg F2 \wedge F5)$
$F2 \vee \neg F5$ (or $\neg F5 \vee F2$)

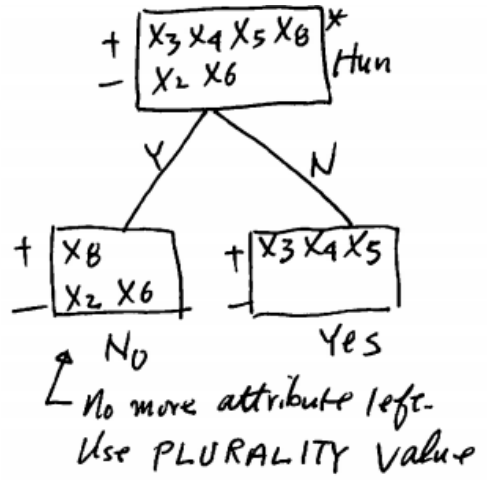
Resolution tree. The tree below is not unique.



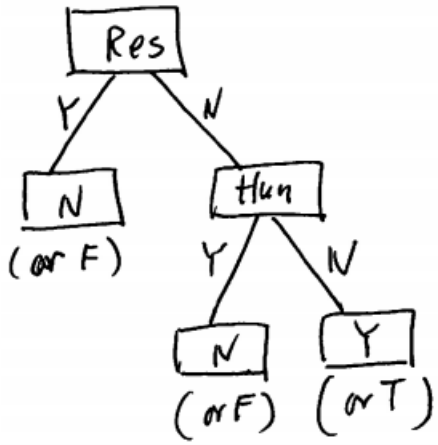
3. [20 points] For the root node: compute Gain of *Res* and *Hun*:



Res has a higher gain. Use *Res* at the root node.
For the right child node* in the partial tree above, use *Hun*:



The final decision tree is therefore:



4. [15 points]

$(a_1, a_2, a_3, a_4) = (1.0, 1.5, -2.0, -2.5)$

$a_5 = \overset{ReLu}{\wedge}(1.0 \times 0.1 + 1.5 \times 0.1) = ReLu(0.25) = 0.25$

$a_6 = ReLu(-2.0 \times 0.1 - 2.5 \times 0.1) = ReLu(-0.45) = 0$

$a_7 = \delta(0.25 \times 0.2 + 0 \times 0.2) = \delta(0.05)$

$= \frac{1}{1 + e^{-0.05}} = \frac{1}{1 + 0.9512} = 0.5125$

5. [20 points]

$y_7 = 1.0$

$Err_7 = y_7 - a_7 = 1.0 - 0.5125 = 0.4875$

$\Delta_7 = Err_7 \times g'(in_7) = Err_7 \times g(in_7)(1 - g(in_7))$

$= Err_7 \times a_7 (1 - a_7)$

$= 0.4875 \times 0.5125 \times (1 - 0.5125)$

$= 0.1218$

$w_{57} = w_{57} + \alpha \times \Delta_7 \times a_5$

$= 0.2 + 0.1 \times 0.1218 \times 0.25$

$= 0.2 + 0.0030$

$= 0.2030$

$w_{67} = w_{67} + \alpha \times \Delta_7 \times a_6$

$= 0.2 + 0.1 \times 0.1218 \times 0$

$= 0.2$

$\Delta_5 = g'(in_5) w_{57} \Delta_7$

$= 1 \times 0.2 \times 0.1218$

$= 0.02436$

$\Delta_6 = g'(in_6) \times w_{67} \Delta_7$

$= 0 \times 0.2 \times 0.1218$

$= 0$

$w_{15} = w_{15} + \alpha \times \Delta_5 \times a_1$

$= 0.1 + 0.1 \times 0.02436 \times 1.0 = 0.10024$

$w_{25} = w_{25} + \alpha \times \Delta_5 \times a_2$

$= 0.1 + 0.1 \times 0.02436 \times 1.5 = 0.1037$

$w_{36} = w_{36} + \alpha \times \Delta_6 \times a_3$

$= 0.1 + 0.1 \times 0 \times -2.10 = 0.1$

$w_{46} = w_{46} + \alpha \times \Delta_6 \times a_4$

$= 0.1 + 0.1 \times 0 \times -2.5 = 0.1$

6. [15 points]

(a) Each neuron in layer 2 has 5 x 5 = 25 connections but the parameters are shared. The number of parameters that need to be trained is 25.

(b) If there are no parameter sharing in layer 2, all 25 connections to each neuron would need to be trained. The total number of parameters therefore equals $124 \times 124 \times 25 = 384{,}400$.

(c) The size of layer 4 = $60 \times 60 = 3{,}600$.

(d) $32 \times 32 \times 32 \times 32 = 1{,}048{,}576$.