Write T if the statement is always and unconditionally true, and write F otherwise.

1.1) __T__ Perceptrons, arranged with no hidden layer, may express logical NOT.

1.2) __F__ Probability theory and propositional logic make the same epistemological commitments.

1.3) __T__ Bayesian Networks are a compact way in which to represent a joint probability distribution when some conditional independence assumptions hold.

1.4) __T__ Bayes' rule contains both prior and posterior probabilities.

1.5) __F__ BFS requires less memory than DFS.

1.6) __F__ Neural networks can only be used to solve classification problems.

1.7) __T__ SVM can classify data that is not linearly separable.

1.8) __T__ A and B are independent if and only if P(A|B)=P(A)

1.9) __F__ Under the Markov assumption, current state can be determined by the next state.

1.10) __F__ People usually use forward error propagation to train neural networks.

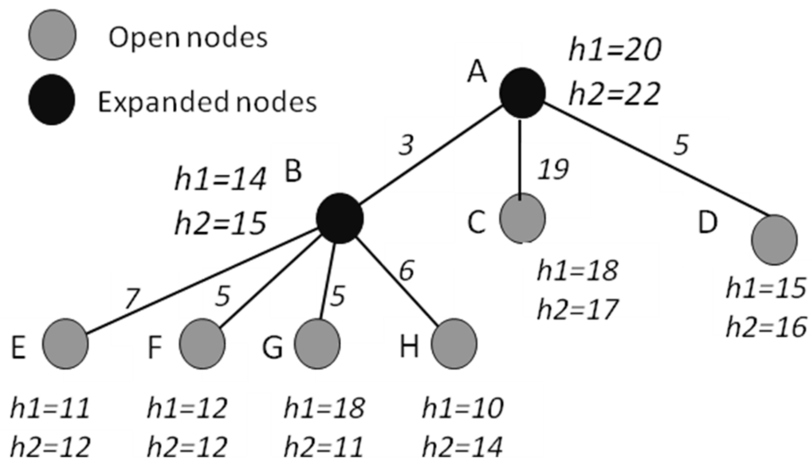
1.11) __T__ An admissible heuristic function never overestimates the actual minimal cost of reaching the goal.

1.12) __F__ When solving MDP, we usually cannot find the optimal policy in an accessible and stochastic environment with known transition model.

1.13) __T__ We can use dynamic programming to solve MDPs.

1.14) __T__ We usually solve CSPs by recursively backtracking.

1.15) __F__ Understanding complicated visual scenes is a just trivial problem of AI.



In the figure above, you are shown a partially expanded search tree with edge costs and two admissible heuristic estimates h1 and h2 for the distance between each node and a goal state that is not shown. Ties between nodes for all algorithms are broken by selecting the node that is first in alphabetic order.

2.1 [4%]: What is the next node that is expanded if only heuristic estimate h2 is used with:

1. [2%] Best First Search: ___________ G
2. [2%] A* Search: ___________ G

2.2 [2%]: Suppose that h2 proves to be an inadmissible heuristic estimate for some reason. What is the potential impact upon the result obtained by the A* algorithm that uses h2?

A* may not yield the optimal solution – but it will find some solution. – 2 points

2.3 [2%]: From the information in the problem statement, can you deduce whether h1 or h2 is the better heuristic estimate for this search application? Why or why not?

No – h1 > h2 for some nodes; and h2 > h1 for other nodes. So, neither heuristic is better in all cases.

2.4 [2%]: Given that both h1 and h2 are admissible heuristics, can you construct a specific way to combine the heuristics h1 and h2 to yield the best heuristic possible, given the information above?

For a node "n" use: max (h1(n), h2(n)) as the heuristic estimate – since h1 and h2 are admissible, both are less than actual cost. So, whichever is closest to the actual without exceeding it will be the best. That is max(h1(n), h2(n)). – 2 points

Consider the following knowledge base:

1. If something is intelligent, it has common sense
2. Not everything that is intelligent understands emotion.
3. If something has common sense, it can play chess.
4. Something can understand emotion and play chess.
5. Siri can not play chess.

3.1. [7.5%] Convert the sentences in the above knowledge base into first order logic expressions, using the following predicates:

- I(x): x is intelligent
- C(x): x can play chess
- S(x): x has common sense
- E(x): x understands emotion

1. □x I(x) □ S(x)
2. □x I(x) □ ~E(x)   or   ~□x I(x) □ E(x)
3. □x S(x) □ C(x)
4. □x E(x) □ C(x)
5. ~C(Siri)

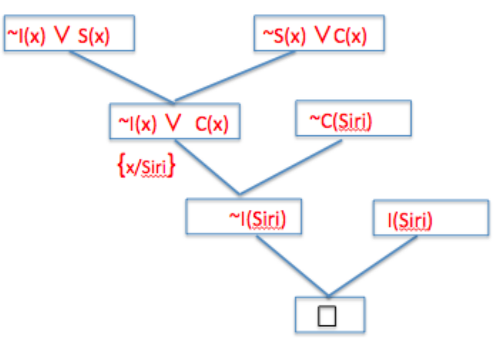
3.2. [7.5%] Use the above knowledge base and proof by contradiction to prove that "Siri is not intelligent".

Convert (relevant) expressions into CNF

- #1:   ~I(x) v S(x)
- #3:   ~S(x) v C(x)

To prove: ~I(Siri), Add I(Siri) to the KB - 2.5 points for these steps

use proof by contradiction: (5 points for this step)



Consider the following STRIPS action schema for planning the end of semester:

LookupExamDate(subject)
    Pre: Enrolled(subject)
    Post: HaveExamDate(subject)

BookFlight()
    Pre: HaveExamDate(Math), HaveExamDate(Physics)
    Post: Have(Ticket)

Study(subject)
    Pre:
    Post: Prepared(subject)

TakeExam(subject)
    Pre: At(School), HaveExamDate(subject), Prepared(subject)
    Post: Passed(subject)

Shop()
    Pre: At(School), Prepared(Math), Prepared(Physics)
    Post: Have(Gift)

Fly()
    Pre: At(School), Have(Ticket)
    Post: ¬At(School), ¬Have(Ticket), At(Home)

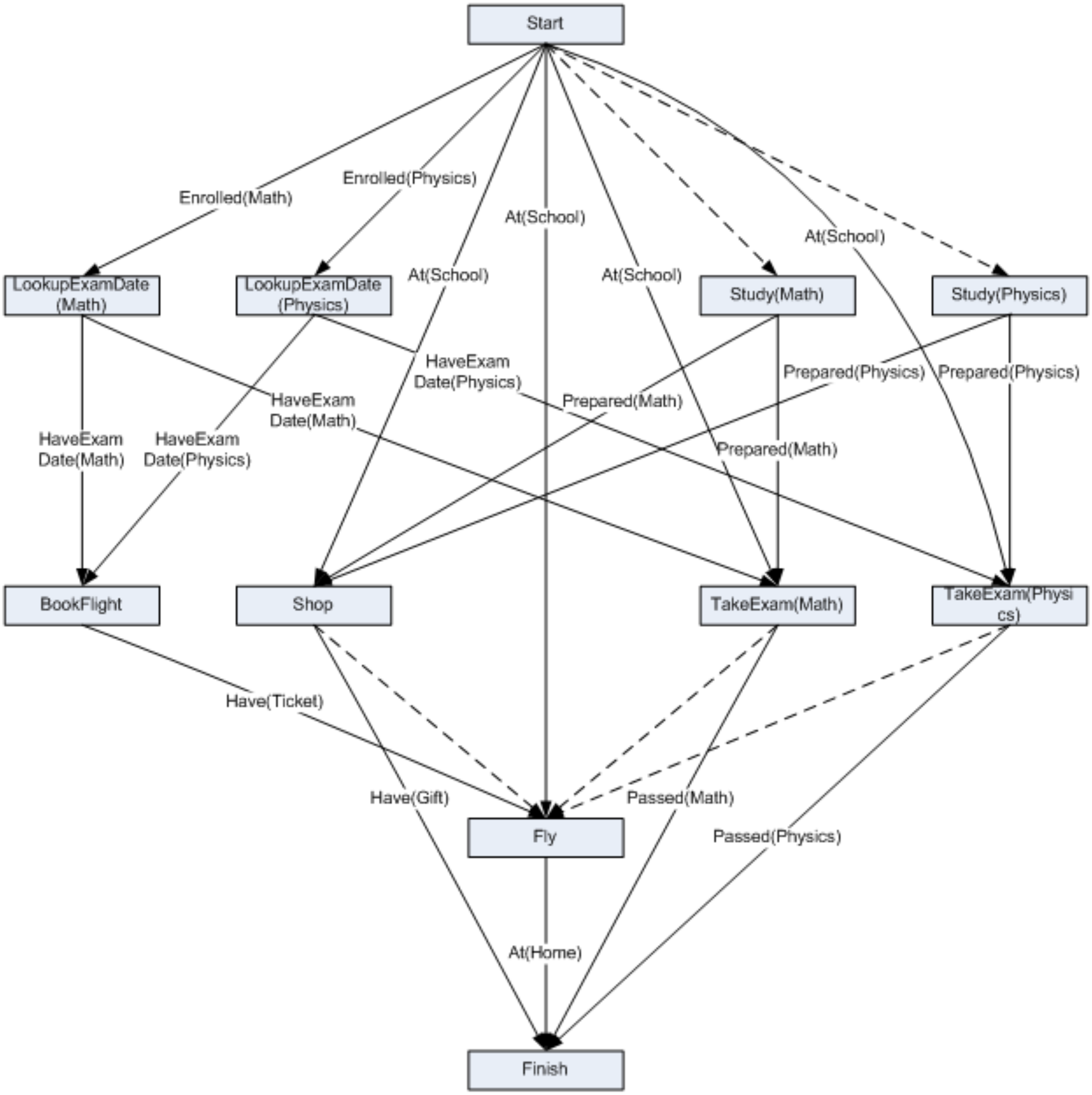
Given the following:

Start: At(School), Enrolled(Math), Enrolled(Physics)
Final: At(Home), Have(Gift), Passed(Math), Passed(Physics)

Show the partial order plan for planning the end of semester. Your answer must be a complete plan and in the form of a graph showing the actions from the Start state to the Final state using the actions above. Your answer must clearly show the precedence relationships between actions and causal links as applicable, allowing the plan executor to have as much choice in the sequencing of actions without violating any causal links.

You should draw the partial order plan on the next page.

Problem 4 (Planning) answer:


Start
Enrolled(Math)
Enrolled(Physics)
At(School)
At(School)
At(School)
At(School)
LookupExamDate (Math)
LookupExamDate (Physics)
Study(Math)
Study(Physics)
HaveExam Date(Math)
HaveExam Date(Physics)
HaveExam Date(Math)
HaveExam Date(Physics)
Prepared(Math)
Prepared(Physics)
Prepared(Math)
Prepared(Physics)
BookFlight
Shop
TakeExam(Math)
TakeExam(Physics)
Have(Ticket)
Have(Gift)
Passed(Math)
Passed(Physics)
Fly
At(Home)
Finish

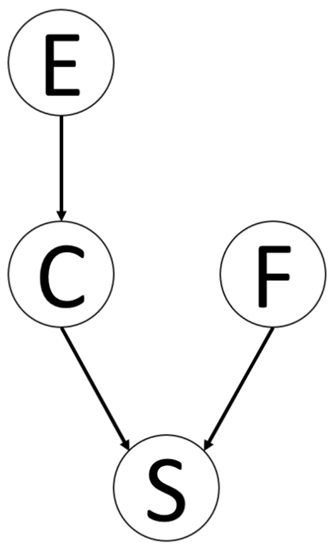
Bayes Net

Suppose John can be very sick (S) that can be caused by two different reasons: getting the seasonal flu and catching a cold (F and C). John thinks that he sometimes catches a cold because he works so hard in his final project that he is exhausted (E) and feels sick. The Bayes' Net and corresponding conditional probability tables are shown below.

For each question, write down the symbolic answer using combinations of P(...) probabilities that are given to you below, e.g., "P(+e) x P(-f)", where "+e" means "E is True" and "-f" means "F is False"). You do not need to calculate the numeric answer. You will use marks if you do not fully develop your answer so that it only uses the probabilities given below.

$P(E)$	
$+e$	0.1
$-e$	0.9

$P(C\|E)$		
$+e$	$+c$	1.0
$+e$	$-c$	0.0
$-e$	$+c$	0.1
$-e$	$-c$	0.9



$P(F)$	
$+f$	0.4
$-f$	0.6

$P(S\|C,F)$			
$+c$	$+f$	$+s$	1.0
$+c$	$+f$	$-s$	0.0
$+c$	$-f$	$+s$	0.9
$+c$	$-f$	$-s$	0.1
$-c$	$+f$	$+s$	0.8
$-c$	$+f$	$-s$	0.2
$-c$	$-f$	$+s$	0.1
$-c$	$-f$	$-s$	0.9

(1) [3%] Please write down the joint probability $P(+s, +c, +f, +e)$.
FORMULA IS OK, NO NEED FOR FINAL NUMBER

$$P(+s,+c,+f,+e) = P(+e)P(+f)P(+c|+e)P(+s|+c,+f) = 0.1 \times 0.4 \times 1.0 \times 1.0 = 0.04$$

so here, P(+e)P(+f)P(+c|+e)P(+s|+c,+f) is ok. FORMULA MUST ONLY USE the P(...) that are in the tables above.

(2) [3%] What is the probability of catching a cold? (Hint: Do not refer to that John feels sick)

$$P(+c) = P(+c|+e)P(+e) + P(+c|-e)P(-e) = 1.0 \times 0.1 + 0.1 \times 0.9 = 0.19$$

(3) [3%] What is the probability that John catches a cold given that he got the seasonal flu?

$$P(+c|+f) = P(+c) = 0.19$$

(4) [3%] What is the probability of working so hard to get a cold $P(+e \mid +c)$?

$$P(+e|+c) = \frac{P(+e,+c)}{P(+c|+e) + P(+c|-e)} = \frac{P(+e)P(+c|+e)}{P(+c|+e) + P(+c|-e)} = \frac{0.1 \times 1.0}{0.1 \times 1.0 + 0.9 \times 0.09} = 0.5263$$

(5) [3%] What is the probability $P(+e \mid +f)$?

$$P(+e|+f) = P(+e) = 0.1$$

(6) [5%] What is the the conditional probability $P(+c \mid +s, +f)$?

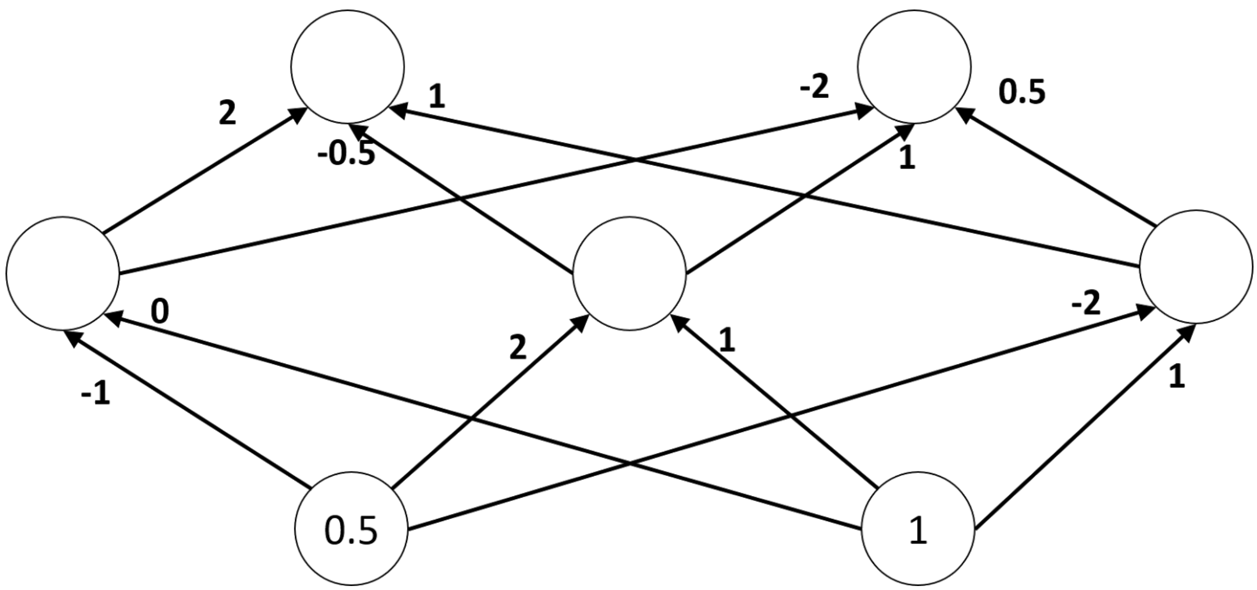
$$P(+c|+s,+f) = \frac{P(+c,+s,+f)}{P(+c,+s,+f) + P(-c,+s,+f)} = \frac{P(+c)P(+f)P(+s|+c,+f)}{P(+c)P(+f)P(+s|+c,+f) + P(-c)P(+f)P(+s|-c,+f)}$$

$$= \frac{0.19 \times 0.4 \times 1.0}{0.19 \times 0.4 \times 1.0 + 0.81 \times 0.4 \times 0.8} = \frac{0.076}{0.076 + 0.2592} = 0.2267$$

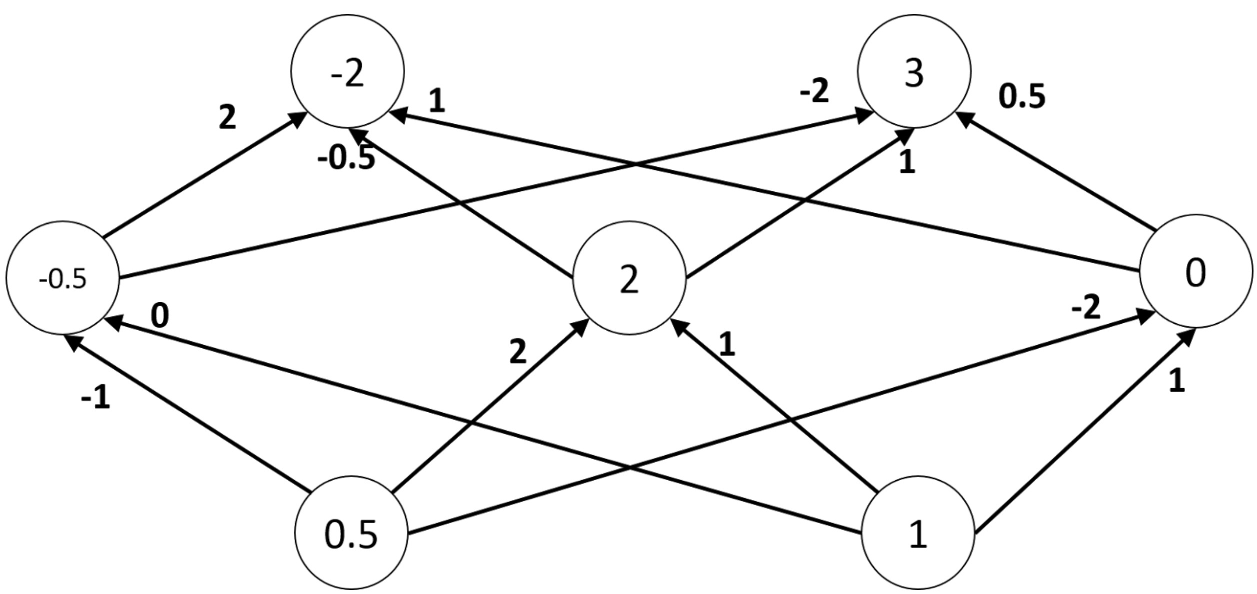
Neural Networks

The following is a network of linear neurons, that is, neurons whose output is identical to their net input. The numbers in the circles indicate the output of a neuron, and the numbers at connections indicate the value of the corresponding weight.

7.1. [5%] Compute the output of the hidden-layer and the output-layer neurons for the given input (0.5, 1) and enter those values into the corresponding circles.



Answer:



7.2. [5%] What is the output of the network for the input (1, 2), i.e. the left input neuron having the value 1 and the right one having the value 2? Do you have to do all the network computations once again in order to answer this question? Explain why you do or do not have to do this.

Answer:

It's (-4, 6). We do not have to do all the computations, because every neuron computes a **linear** function on its inputs, which means that the entire network computes a **linear** function. For such a function, if we double the input, we simply double the output as well.

1. **Decision Tree Learning**

1. I(P) = I (1216 , 416) = -34 log(34) - 14 log(14) = 0.81 (not wanted)
2. a) weather: sunny → 6 yes , 0 no   entropy = -1 * log(1) - 0*log(0) = 0

   weather: cloudy → 3 yes , 3 no   entropy = -0.5*log(0.5) – 0.5*log(0.5) = 1

   b) IG (choresToDo) = I (Hike) – remainder (choresToDo) =

   0.81 – [12*I(46, 26) + 12*I(56, 16)] = 0.81 – [0.5*0.92 + 0.5*0.65] = 0.025

   IG(weather) = I (Hike) – remainder(weather) =

   0.81 – [12*I(1,0) + 12*I(0.5,0.5)] = 0.81 – 0.5 = 0.31

   c) weather is the better choice. Because the information gain (IG) from it is more.

Circle the best choice for each question:

(a) [2%] Markov Decision Processes (MDP) are:

a. A form of knowledge sharing
b. Non-monotonic
c. An ontology for multiple inheritance
d. All of the above
e. None of the above (both answers ok)

(b) [2%] Occam's Razor:

a. Passes the Turing Test
b. Prefers smaller Decision Trees that generalize
c. Causes overfitting in Neural Networks
d. All of the above
e. None of the above

(c) [2%] Machine learning is:

a. Science
b. Neat
c. Rational
d. All of the above
e. None of the above

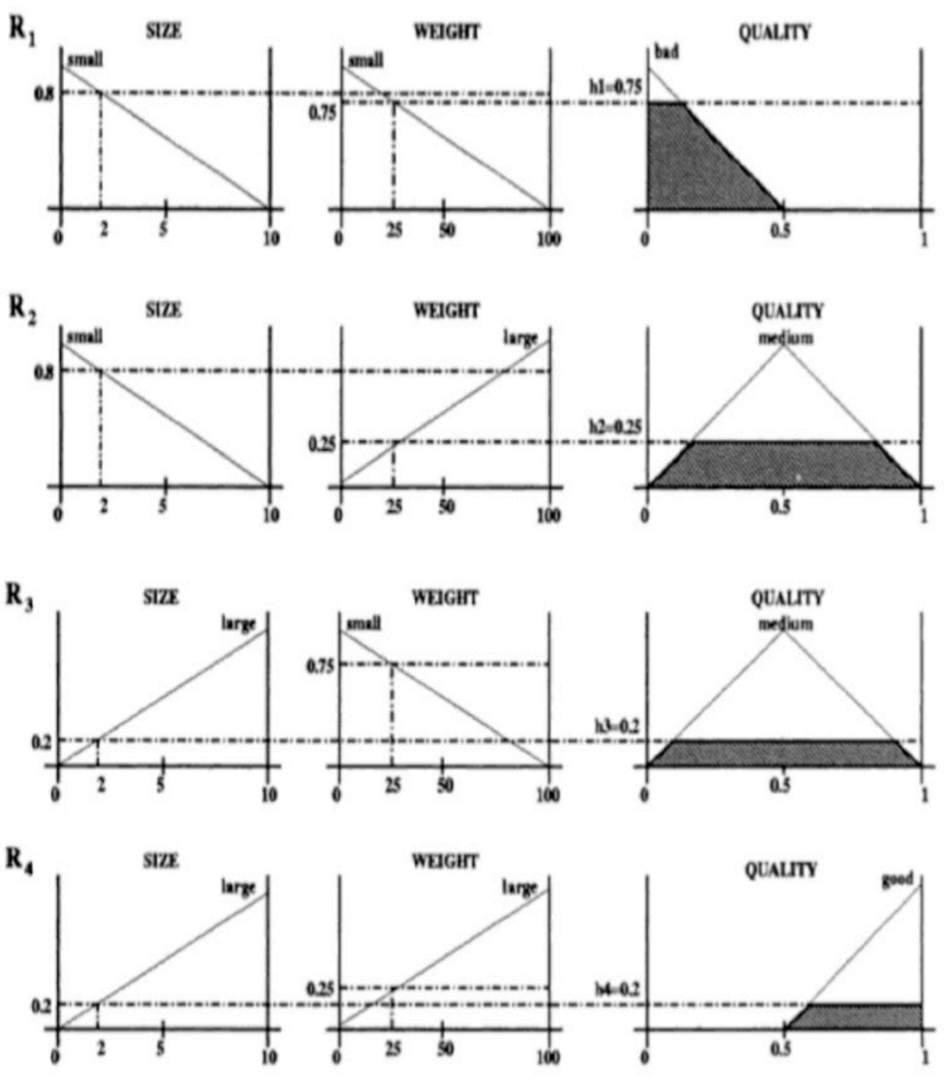
(d) [2%] For efficient real world applications, Knowledge Engineering is:

a. An inexpensive process
b. Not allowed to use entailment
c. Must be both Neat and Scruffy
d. All of the above
e. None of the above

(e) [2%]: Bayesian Networks:

a. Use probability to quantify ignorance
b. Have a bias for the simplest hypothesis
c. Calculate full joint probabilities for rational actions
d. All of the above
e. None of the above

4.1. [16%] Draw graphs to show how the fuzzy rules evaluate for the given apple.
Show the clipped values in your drawings. <u>You do not have to be very precise as long as you indicate the important values on your axes. Be sure to label your axes!</u>



4 points each rule, as follows:

1pt fuzzify size=2 onto correct size fuzzy concept mentioned in rule (small or large). 1 vertical line at size=2 and 1 horizontal line at intersect between the vertical line and the concept

1pt to fuzzify weight=25 in the same way

1pt to take the AND (min) between both horizontal lines

1pt to clip the correct quality concept (bad, medium or good) mentioned in the rule

4.2. [3%] From the clipped values, approximately draw the fuzzy aggregate:

Shape should roughly look like the union of the 4 clipped areas from 4.1