Sample Exam Problems:

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

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

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

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

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

1.5) ____ BFS requires less memory than DFS.

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

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

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

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

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

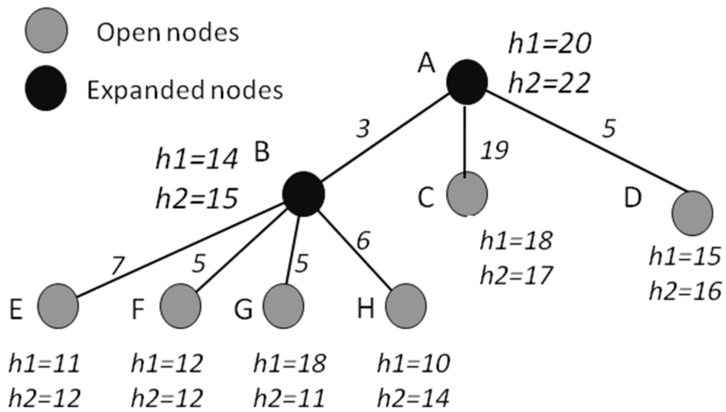
1.11) ____ An admissible heuristic function never overestimates the actual minimal cost of reaching the goal.

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

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

1.14) ____ We usually solve CSPs by recursively backtracking.

1.15) ____ 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: __________
2. [2%] A* Search: __________

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?

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?

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?

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

3.2. [7.5%] Use the above knowledge base and proof by contradiction with Resolution to prove that "Siri is not intelligent". You will lose points if you use anything else than resolution or if you do not clearly indicate all your substitutions.

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.

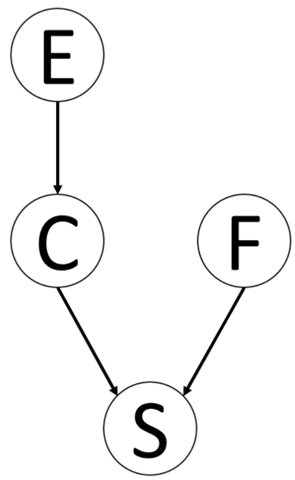
You should draw the partial order plan on the next page.

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 "I is False"). You do not need to calculate the numeric answer. You will lose 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):

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

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

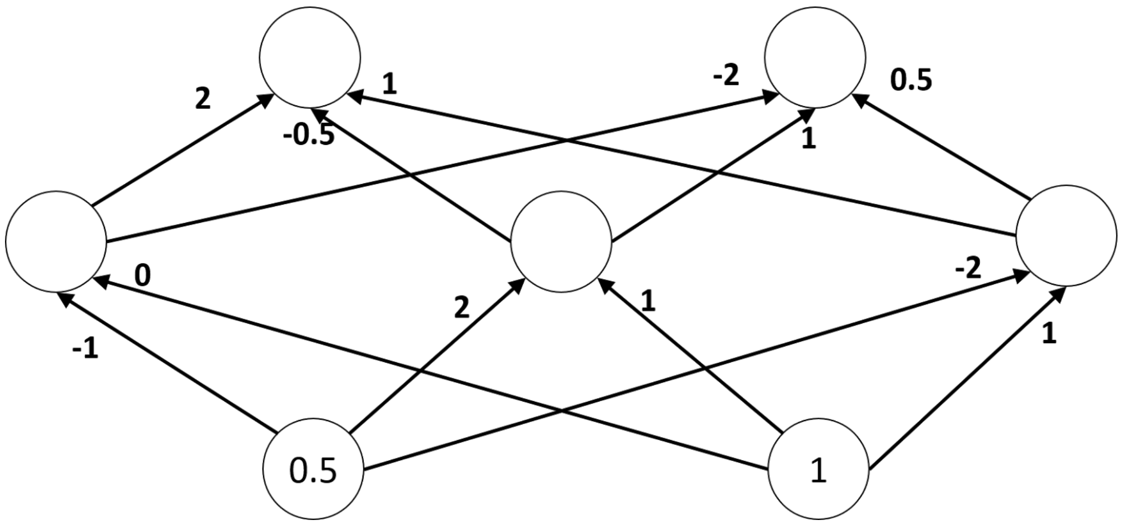
(4) [3%] What is the probability of getting exhausted given a cold, P(+e | +c)?

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

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

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**.



7.2. [3%] 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?

7.3. [2%] To answer 7.2, do you have to do all the network computations once again? Explain why you do or do not have to do this.

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

(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

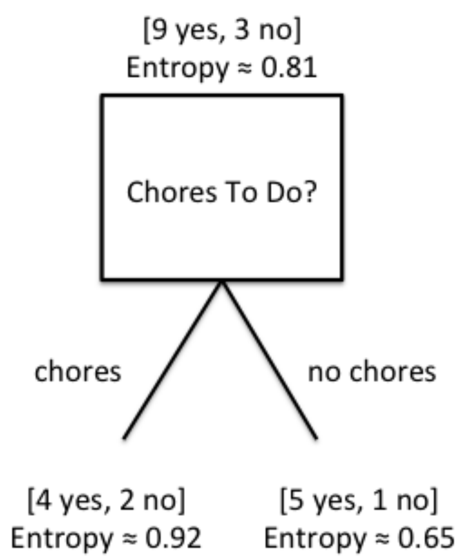
(d) [2%] For efficient real world applications, Knowledge Engineering is:
a. An inexpensive process
b. Not allowed to use entailment
c. Must be both Neat and Scruffy
d. All of the above
e. None of the above

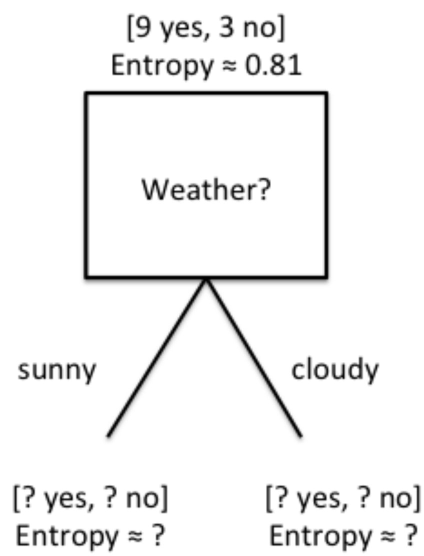
(e) [2%]: Bayesian Networks:
a. Use probability to quantify ignorance
b. Have a bias for the simplest hypothesis
c. Calculate full joint probabilities for rational actions
d. All of the above
e. None of the above

2. In trying to decide whether or not to go hiking, some of the factors I consider are the weather and whether or not I have chores to do at home. There are other factors as well of course. Some example decisions are summarized in the table below.

Weather	ChoresToDo	Hike
Sunny	chores	yes
Sunny	chores	yes
Sunny	no chores	yes
Sunny	no chores	yes
Sunny	no chores	yes
Sunny	no chores	yes
Cloudy	chores	no
Cloudy	chores	no
Cloudy	no chores	yes
Cloudy	no chores	no
Cloudy	chores	yes
Cloudy	chores	yes

With 9 positive examples of hiking and 3 negatives, the entropy or information of this decision in bits is 0.81. Consider the following decision trees, reflecting splitting on the attributes of Weather or ChoresToDo:





a) The ChoresToDo tree has been filled out. Complete the values for the Weather tree, including entropy. (Show your work) (6 points)

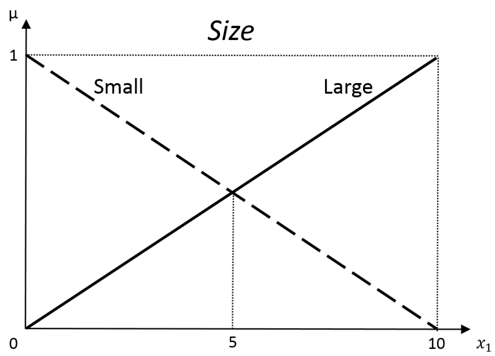
b) Calculate the information gain *IG* from splitting on ChoresToDo and Weather. Please show formulas used and steps clearly. You do not need to compute a final numerical result. Plug in values to the formulas so that the result could be calculated with a calculator. (4 points)

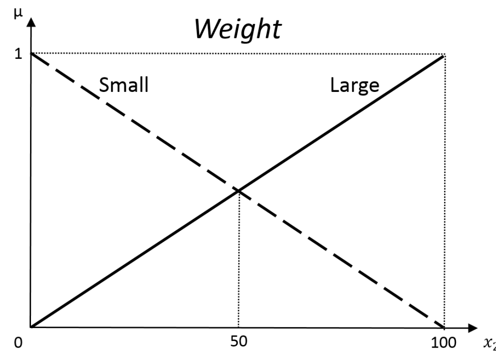
IG(ChoresToDo) =

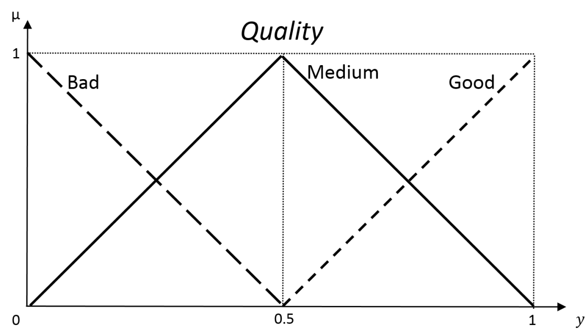
IG(Weather) =

c) Which of the two attributes is a better choice in constructing a decision tree? Why? (2 points)

We want to use fuzzy inference system to inspect the quality of apples. In the inspection, we check **Size** and **Weight** to decide **Quality**. Thus, consider three variables, x_1 for size, x_2 for weight, and y for quality. The membership functions for Small or Large size, Small or Large weight, and Bad, Medium or Good quality have triangular shapes as shown below:







Rule 1: If Size is Small and Weight is Small, then Quality is Bad.

Rule 2: If Size is Small and Weight is Large, then Quality is Medium.

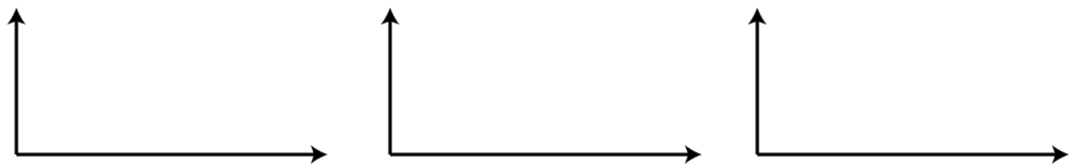
Rule 3: If Size is Large and Weight is Small, then Quality is Medium.

Rule 4: If Size is Large and Weight is Large, then Quality is Good.

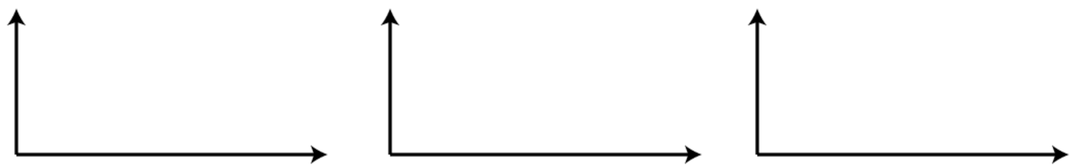
Now, given an apple such that $\{ x_1, x_2 \} = \{ 2, 25 \}$,

4.1. [16%] Draw graphs to show how the fuzzy rules evaluate for the given apple.
Show the clipped values in your drawings. 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!

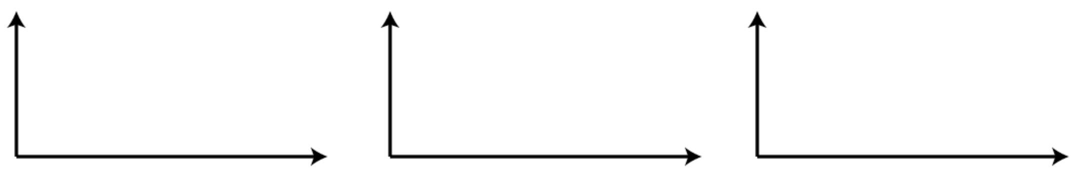
Rule 1:



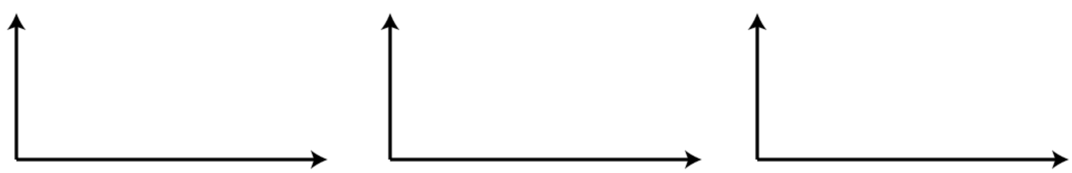
Rule 2:



Rule 3:



Rule 4:



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

4.3. [1%] Using center of mass, approximately show the defuzzified quality of the given apple, on the same graph as you draw for question 4.2.