Massachusetts Institute of Technology

6.034 Artificial Intelligence

Handout

Problem 1 True of False (20 Points)

1. Many rule-based systems can answer some questions about their own behavior, such as certain questions focused on "how" and "why," but rule-based systems generally cannot explain the origins of the rules.

   True. Many generate a goal tree, so they can answer some questions about how and why, but rule-based systems generally contain no information about where the rules came from.

2. The main purpose for introducing the rete algorithm was to connect rule based systems to information in relational databases.

   False. The main purpose for introducing the rete algorithm was to increase speed. The rete algorithm is most redily explained in terms of relational operations, but no relational database is necessarily, or even often, involved.

3. If you use a beam search, without backtracking, to a specified depth, you may miss the shortest path from start to goal.

   True. You might miss the shortest path because the the shortest path may not look sufficiently likely at any level. Also, searching to a specified depth may mean terminating search before the shortest path, or any path, is found.

4. If you use alpha-beta search, to a specified depth, you may miss a better move that would be found by mini-max alone, looking ahead to the same specified depth.

   False. Alpha-beta incurs no risk of missing the correct line of play from both the min and max perspectives.

5. If you face an opponent that uses mini-max to a specified depth to play a game, you can reduce the ability of that opponent to win by playing what you believe to be your second-best move, rather than your best move.

   False. Your second best move will enable the opponent to force you to a worse score.

6. The labeled arcs on a semantic transition tree grammar often provide domain specific contraints on sentence interpretation.

   True. Arcs may constrain the search to a word that identifies, say, a kind of block, tool, or geographical place.

7. In semantic transition tree grammars, recursions handle grammatical constructions that would otherwise require loops.

   True. The adjectives in "The big red angry silent block" are handled by recursion, not by a loop.

8. There is no need for backup when using semantic transition tree grammars to analyze English sentences

   False. Semantic transition trees ofter nothing that eliminates backup.

9. Relational databases provide all the functions of frame systems, but frame systems are more commonly used in AI applications because there is rarely a need to handle huge quantities of data.

   False. Relational databases provide no inheritance mechanism.

10. The performance of a nearest-neighbor recognition system will, in general, go down if you add measurements in dimensions that are unrelated to identity.

    True. If something to be identified accidently lies close to the wrong sample in irrelevant dimensions, it may seem closer to that wrong sample than to the sample that provides the proper identity.

Problem 2 Nearest Neighbors (25 points)

One of the key components of the question was the following statement:

> If your answer to the above question was **Yes**, SAW is curious about the "correct" nearest neighbor approach for this file. They ask you to recommend the *simplest* possible nearest neighbor classifier that will work for this data. The different possible approaches are listed below. The simplest technique is first and the most complex last. You need only give the corresponding number.

So in general when there are two approaches which might work, you should pick the *simplest*.

Another key piece of information was the following:

> The customers of SAW are picky about their widgets, if they pay extra for a GREAT widget they want to get a GREAT widget (i.e. **it is important to never classify a GOOD widget as a GREAT widget**). SAW loses the additional profit they might make by charging more for a GREAT widget, but they are not too worried about it (i.e. **you can classify some of the GREAT widgets as GOOD widgets, but fewer is better**).

From this you should conclude that classifying a "PLUS" as a "DIAMOND" is much worse that classifying a "DIAMOND" as a "PLUS".

Given this information the correct answers are:

Data File	Appropriate?	Which Approach?	Weights (If necessary)	
File A	Yes	1	$w1 =$	$w2 =$
File B	Yes	3	$w3 =$	$w4 =$
File C	Yes	2	$w5 = 1$	$w6 = 10000$
File D	No		$w7 =$	$w8 =$
File E	No		$w9 =$	$w10 =$

File A is "Yes" because the data is well separated; any of the nearest neighbor approaches will work.

File B is "Yes" because the GOOD widgets (the PLUSes) are clustered together tightly. In the region containing GOOD widgets there are only a few GREAT ones. In general if you find the 5 nearest neighbors of any PLUS, at least 3 of them are PLUSes. So a 5-Nearest-Neighbor approach will almost always correctly classify the GOOD widgets. The 5-Nearest-Neighbor algorithm will incorrectly classify a few of the GREAT widgets, but we don't care too much about that.

File C is "Yes" because the data is separated, but only by sensor 6. Sensor 5 is ineffective. The range of values that sensor 6 takes on is much smaller than sensor 5, so the Euclidean distance places the PLUSes very near to the DIAMONDs. The solution is to stretch out the distances in the vertical direction. This is done by setting the weight for sensor 6, $w6$, to a very large value (with respect to $w5$). In general anything where $\frac{w6}{w5} > 3$ was accepted.

File D is "No". Let's examine cases: 1) Nearest neighbor places some DIAMONDS near PLUSES; 2) 5-Nearest-Neighbor does not help, the DIAMONDs are about equally far from a set of DIAMONDS as they are from a set of PLUSes; 3) There are no positive weights that bring DIAMONDS closer together, while keeping them away from the PLUSes – weights allow you to increase the sensitivity to variations in one of the sensors, but in this case neither sensor does the trick; 4) The combination of 2 and 3 does not help either.

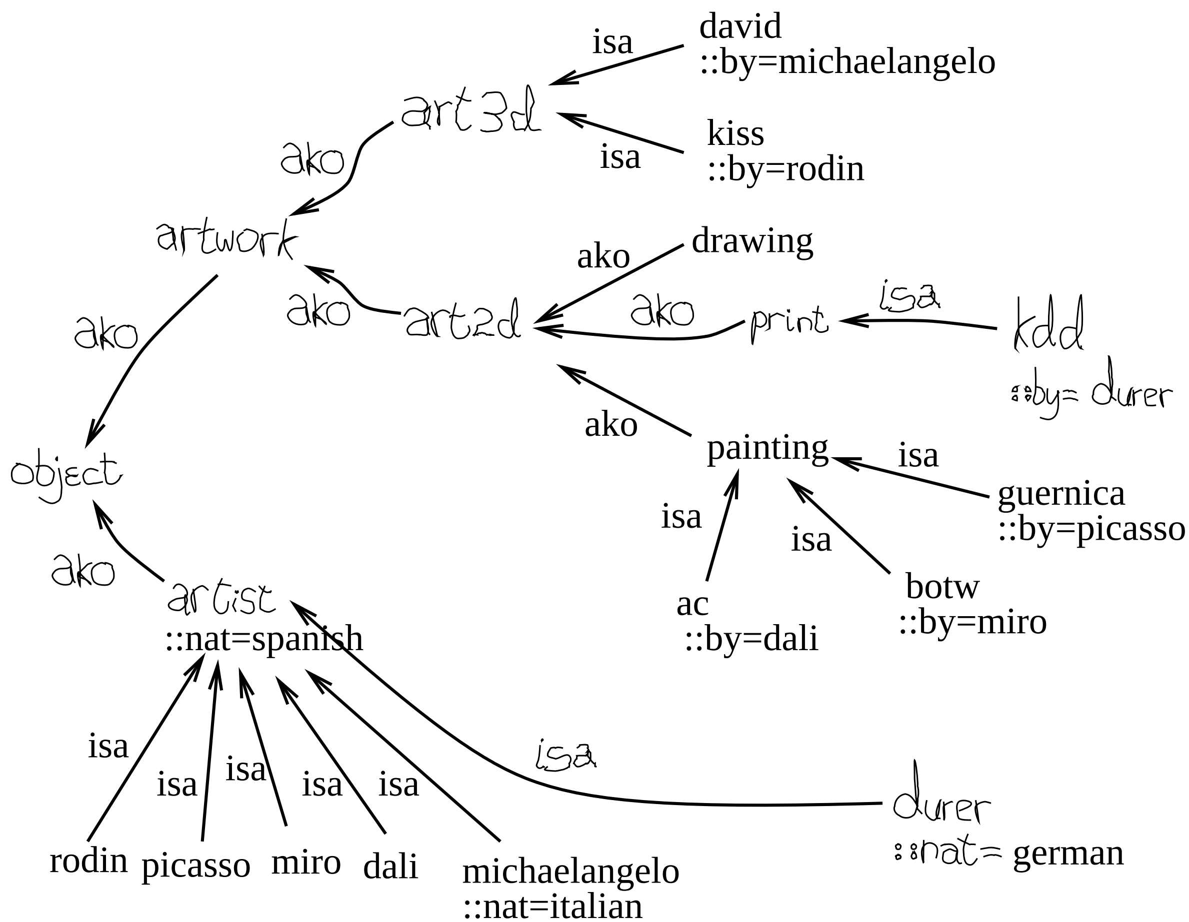
File E is "No". The data are completely intermingled. PLUSes and DIAMONDS are not separated in any way.

(Grading: Each "File" was worth 5 points. If the answer was "No" you had to correctly determine that to get credit. If the answer was "Yes" you were given 2 points for picking the correct nearest neighbor technique.)

Problem 3 Classy Objects d'Art (25 Points)

Part A (6 points)

A sample complete figure is shown below. Points were deducted for omitting information. A common error was missing the fact that Spanish nationality is the class default for artist.



david
::by=michaelangelo
kiss
::by=rodin
art3d
artwork
drawing
print
kdd
::by= durer
art2d
painting
guernica
::by=picasso
ac
::by=dali
botw
::by=miro
object
artist
::nat=spanish
durer
::nat= german
rodin
picasso
miro
dali
michaelangelo
::nat=italian
isa
ako

Part B (7 points)

Using `'nat` for *nationality* and `'by` for *was created by*, and the same class and object labels as in the diagram, show the `frames.scm` expressions you would write to:

B1 (1 point)

Represent the fact that Picasso was an artist.

```
(define-instance 'picasso 'artist)
```

B2 (1 point)

Represent the fact that Picasso painted *Guernica*.

```
(assert-property 'guernica 'by 'picasso)
```

B3 (1 point)

Represent the fact that Autumnal Cannibalism is a painting.

```
(define-instance 'ac 'painting)
```

B4 (1 point)

Represent the fact that Dürer was German.

```
(assert-property 'durer 'nat 'german)
```

B5 (1 points)

Represent the default assumption that artists in the collection are Spanish.

`(assert-property 'artist 'nat 'spanish)` or
`(define-class-default 'artist 'nat 'spanish)`

B6 (2 points)

Represent the fact that artworks are either 2d or 3d.

```
(define-class 'artwork 'art2D)
(define-class 'artwork 'art3d)
```

Part C (11 points)

C1 (1 point)

Who painted *Guernica*?
`(get-property 'guernica 'by) => picasso`

C2 (1 point)

What is *David*?
`(get-class 'david) => art3d`

C3 (1 point)

Are drawings artworks?
`(subclass?  'drawing 'artwork) => #t`

C4 (1 point)

What nationality was Rodin?
`(get-property 'rodin 'nat) => spanish`

C5 (2 points)

Is *Autumnal Cannibalism* an artwork?
`(subclass?  (get-class 'ac) 'artwork) => #t` or
`(class-member?  'ac 'artwork) => #t`

C6 (2 points)

Was Dürer German?
`(equals?  (get-property 'durer 'nat) 'german) => #t`

C7 (3 points)

Was the sculptor of *The Kiss* Spanish?
`(equals?  'spanish (get-property (get-property 'kiss 'by) 'nat))) => #t`

Part D (1 point)

The answer was item 3, Relational databases, because counting and listing are primitive operations in databases but not in the other technologies listed.

Problem 4 Icy Road Networks at Bay (30 points)

Part A Warm-up!(4 pts)

Why is there *no* directed arc between H and W in the network?

Answer, either:

1. H and W are *conditionally* independent.
2. H and W do not causally interact (or some variant thereof).

Part B The Inspector Calls (10 pts)

B.1 3 pts

What is the numerical value of $P(W = t, I = t)$?

$$P(W, I) = P(W|I) \cdot P(I) = 0.8 \times 0.7 = 0.56$$

B.2 3 pts

Calculate the probability of icy roads *given* the knowledge that Watson has crashed, that is, $P(I = t|W = t)$. Write down the calculation symbolically, a formula in terms of probabilities such as $P(I = t)$ that are specified in the probability tables in the figure above. Then write down the formula for $P(W = t)$ also in terms of the entries in the tables.

Symbolic formula for $P(I = t|W = t) =$

$$\begin{aligned} P(I|W) = &\quad P(W|I) \cdot P(I)/P(W) = P(W = t|I = t) \cdot P(I = t)/P(W = t) \\ = 0.56/0.59 & \end{aligned}$$

where $P(W = t)$ is given immediately below. (You were not asked to provide the numerical answer.)

Symbolic formula for $P(W = t) =$

$$
\begin{aligned}
P(W=t) = \quad & P(W=t|I=t)\cdot P(I=t) + P(W=t|I=f)\cdot P(I=f) \\
= \quad & 0.56 + 0.03 \\
= \quad & 0.59
\end{aligned}
$$

B.3 4 pts

In terms of probabilities of fully specified states, write down the *simplest* symbolic formula needed to calculate the probability that Holmes has crashed, conditional on W, that is, incorporating the fact that Watson has crashed, but that we do not yet know whether or not the roads are icy.

There are several ways to write this answer.

$$P(H|W) = P(H=t, I=t, W=t) + P(H=t, I=f, W=t)/P(H=*, I=*, W=t)$$

so the denominator must be a sum over four terms, ranging over the various values for I and H:

Denominator is:

$$\Sigma_{x\in t,f}\Sigma_{y\in t,f}P(W=t, H=x, I=y)$$

or alternatively, written out,

$$P(H=t, I=t, W=t) + P(H=f, I=t, W=t) + P(H=t, I=f, W=t) + P(H=f, I=f, W=t)$$

Part C Connecting nets (6 pts)

C.1 5 pts

Suppose that besides being a bad driver and crashing on icy roads, Watson also sometimes drives a Saab with bad tires, denoted by the binary-valued variable BS. We set the value of $P(BS)=t$ to be 0.8. We assume for this problem that BS and I are independent random variables. We may also assume that when Watson drives a Saab with bad tires on *non*-icy roads, there is a 0.85 probability that he will crash (and so a 0.15 probability that he will not crash when driving a Saab with bad tires on non-icy roads). (Note that this means that BS and I are independently causative factors for Watson crashing.)

Show how to update the existing Bayes network to include this new BS factor, using the space below. Add/delete links as appropriate. Naturally, there will have to be some probability table(s) associated with the new node as well as some potential changes to the existing tables; please draw in those tables as well. Fill in those entries for which you can pick values from the original network or from the text above. If a value is not known, leave it blank.

Answer:

The straightforward answer adds a directed arc from B to W, a new table for BS, with $P(BS)=t$=0.8, and a new conditional probability table for W that looks like this:

f	f	0.1
f	t	0.8
t	f	0.85
t	t	??

Two more sophisticated and heroic answers are possible. First, one can calculate the t, t case, $P(W|BS, I)$, carrying out the noisy-or calculation. The correct answer here is 0.96 (*not* 0.965 or 0.97). Second, if one assumes real probabilistic independence between BS and I, then one can multiply out the probabilities to get values for all the entries — it turns out that the numbers provided are inconsistent with these.

C.2 1 pt

The Inspector's assistant told her that the roads are, in fact, *not* icy. That is, after the assistant speaks, the values for W and I are now both known: $W = t$ (Watson has crashed) and $I = f$ (the roads are not icy). As we saw in the earthquake/alarm question in problem set 7, this means that our initial estimate of the likelihood that Watson drives a Saab with bad tires, $P(BS = t)$, might have to be re-estimated, given our new knowledge.

Given that the roads are now known to be *not* icy ($I = f$), and that Watson *has* crashed ($W = t$) which of the following statements best characterizes the probabilistic relationship between $P(BS = t)$ and $P(BS = t|I = f, W = t)$? Circle *one* answer below.

- The two will have the **same value**
- $P(BS = t|I = f, W = t)$ will be **smaller** than $P(BS = t)$
- $P(BS = t|I = f, W = t)$ will be **greater** than $P(BS = t)$
- There is no way to tell the relation between the two, given the information provided

The correct answer is **greater**.

Part D Noisy networks (10 pts)

Of course, we can imagine many possible causes for Watson crashing in addition to icy roads and Watson driving a Saab with bad tires. For instance, we could add a new possible cause, such as E, indicating whether Watson is Exhausted or not. In such cases, as we have seen, it often becomes useful to posit a *noisy-or* model to reduce the number of probability parameters we need to use for calculations. In the following problem, assume that just E and I cause crashes, neglecting BS.

D.1 3 pts

If the "causal probability" from I to W is 0.7, and the causal probability from E to W is 0.8, then what is the numerical value of the probability $P(W = t|I = t, E = t)$ under the noisy-or assumptions, assuming no leak term is present?

$P(W = t|I = t, E = t)$=

$$1 - (1 - P_c(W|I))(1 - P_c(W|E)) = (1 - (0.2)(0.3)) = 1 - 0.06 = 0.94$$

D.2 3 pts

Now suppose we also add a *Leak* variable, L, to cover the possibility where the "symptom" W appears (Watson crashes), but neither I nor E is true. That is, the $P_c(L)$ is like a probability in that it is always positive.

Please write out symbolically the formula for the noisy-or calculation using the *Leak* binary-valued variable, L. Use $P_c(X|Y)$ to denote the "causal probability" estimates that Y causes X for each conditional probability.

Symbolic formula for $P(W = t|I = t, E = t) =$

$$(1 - (1 - P_c(W|I))(1 - P_c(W|E))(1 - P_c(L))$$

D.3 4 pts

Using the same causal probabilities for I and E as in Part (i) above, suppose the probability $P(W = t|I = t, E = t)$ has been calculated and found to be equal to 0.90 (when Watson is both exhausted and the roads are icy, there is a higher probability of Watson crashing).

Is this probability of 0.90 consistent with the noisy-or assumptions, given a leak variable? Explain why or why not.

Why or why not consistent with noisy-or:

Answer: It is not consistent with noisy-or.

If one carries out the noisy-or calculation, we would have:

$$0.9 =?(1 - (1 - P_c(W|I))(1 - P_c(W|E))(1 - P_c(L))$$

and if we solve for $P_c(L)$:

$$0.9 =?1 - (1 - 0.7)(1 - 0.8)(1 - P_c(L)) =$$

we get $-4/6$, an invalid "probability" under our assumptions.