CS360 Homework 4– Solution

Semantic Networks

1) Exactly how can a computer determine that the sentence "I keep my money in the bank" refers to a financial institution rather than the land alongside a river. In this context, explain also what spreading activation is.

Answer:
If the semantic network has nodes for both meanings of the word 'bank,' they get both activated with an initial weight, along with any other nodes that correspond to concepts from the given sentence. Any active node then activates its neighbors with a lower weight, essentially propagating its activation throughout the network. The activation of a node is the higher, the more highly activated nodes are in close distance to it in the semantic network. The node that corresponds to the 'financial institution' meaning of 'bank' will thus be more highly activated than the node that corresponds to the 'land alongside a river' meaning since it is closer to the node corresponding to 'money.'

2) What are advantages and disadvantages of using semantic networks for knowledge representation over using first-order logic?

Answer:
Advantages:

- Semantic networks are easier to read and understand by humans.
- They are easier to implement and can be more efficient (since they can use special purpose procedures).
- They can be more expressive than first-order logic in some regards (for instance, inheritance with exceptions).

Disadvantages:

- They are less expressive than first-order logic in some regards (for instance, negation and disjunction are problems).
- Their semantics is often not well defined.
- They have problems with multiple inheritance of incompatible properties.

Decision Trees

3) Assume that you have the following training examples available:

	F1	F2	F3	F4	F5	Class
Example 1	t	t	f	f	f	p
Example 2	f	f	t	t	f	p
Example 3	t	f	f	t	f	p
Example 4	t	f	t	f	t	p
Example 5	f	t	f	f	f	n
Example 6	t	t	f	t	t	n
Example 7	f	t	t	t	t	n

Use all of the training examples to construct a decision tree. In case of ties between features, break ties in favor of features with smaller numbers (for example, favor F1 over F2, F2 over F3, and so on).

How does the resulting decision tree classify the following example:

	F1	F2	F3	F4	F5	Class
Example 8	f	f	f	t	t	?

Answer:
Let $H(x, y)$ denote the entropy of a data set with x examples belonging to class p and y examples belonging to class n. That is,

$$H(x, y) = -\frac{x}{x+y} \ log_2 \ \frac{x}{x+y} - \frac{y}{x+y} \ log_2 \ \frac{y}{x+y}.$$

We precompute the values for some $H(x, y)$, which will be useful when constructing the decision tree.

$H(0, x) = H(x, 0) = -1 \ log_2 \ 1 - 0 \ log_2 \ 0 = 0$

$H(x, x) = -(1/2) \ log_2 \ (1/2) - (1/2) \ log_2 \ (1/2) = 1$

$H(1, 2) = H(2, 1) = -(1/3) \ log_2 \ (1/3) - (2/3) \ log_2 \ (2/3) = 0.918$

$H(1, 3) = H(3, 1) = -(1/4) \ log_2 \ (1/4) - (3/4) \ log_2 \ (3/4) = 0.811$

First, we choose from { F1, F2, F3, F4, F5 } to become the root.

	F1	F2	F3	F4	F5	Class
Example 1	t	t	f	f	f	p
Example 2	f	f	t	t	f	p
Example 3	t	f	f	t	f	p
Example 4	t	f	t	f	t	p
Example 5	f	t	f	f	f	n
Example 6	t	t	f	t	t	n
Example 7	f	t	t	t	t	n

Let $E(F)$ denote the average entropy resulting from choosing feature F as the root. For instance, let F1 be the root. Then, examples 1, 3, 4 and 6 are grouped in F1's t-branch (since F1 = t for these examples) and examples 2, 5 and 7 are grouped in F1's f-branch. The entropy of F1's t-branch is $H(3, 1)$ (since there are three examples that belong to class p and one example that belongs to class

n) and the entropy of F1's f-branch is $H(1,2)$. Taking the weighted average of these entropies, we get $E(\text{F1})$. Below, we compute $E(F)$ for all features F:

$E(\text{F1}) = 4/7\ H(3,1) + 3/7\ H(1,2) = 4/7 \times 0.811 + 3/7 \times 0.918 = 0.857$

$E(\text{F2}) = 4/7\ H(1,3) + 3/7\ H(3,0) = 4/7 \times 0.811 + 3/7 \times 0 = 0.463$

$E(\text{F3}) = 3/7\ H(2,1) + 4/7\ H(2,2) = 3/7 \times 0.918 + 4/7 \times 1 = 0.965$

$E(\text{F4}) = 4/7\ H(2,2) + 3/7\ H(2,1) = 4/7 \times 1 + 3/7 \times 0.918 = 0.965$

$E(\text{F5}) = 3/7\ H(1,2) + 4/7\ H(3,1) = 3/7 \times 0.918 + 4/7 \times 0.811 = 0.857$

Since F2 minimizes the average entropy, F2 becomes the root.

Now, we choose F2's f-child.

	F1	F2	F3	F4	F5	Class
Example 2	f	f	t	t	f	p
Example 3	t	f	f	t	f	p
Example 4	t	f	t	f	t	p

Since all examples are in class p, class p becomes F2's f-child.

Next, we choose from { F1, F3, F4, F5 } to be F2's t-child.

	F1	F2	F3	F4	F5	Class
Example 1	t	t	f	f	f	p
Example 5	f	t	f	f	f	n
Example 6	t	t	f	t	t	n
Example 7	f	t	t	t	t	n

$E(\text{F1}) = 2/4\ H(1,1) + 2/4\ H(0,2) = 2/4 \times 1 + 2/4 \times 0 = 0.5$

$E(\text{F3}) = 1/4\ H(0,1) + 3/4\ H(1,2) = 1/4 \times 0 + 3/4 \times 0.918 = 0.6885$

$E(\text{F4}) = 2/4\ H(0,2) + 2/4\ H(1,1) = 2/4 \times 0 + 2/4 \times 1 = 0.5$

$E(\text{F5}) = 2/4\ H(0,2) + 2/4\ H(1,1) = 2/4 \times 0 + 2/4 \times 1 = 0.5$

F1, F4 and F5 minimize the average entropy. Since we break ties in favor of features with smaller numbers, F1 becomes F2's t-child.

Now, we choose F1's f-child.

	F1	F2	F3	F4	F5	Class
Example 5	f	t	f	f	f	n
Example 7	f	t	t	t	t	n

Since all examples are in class n, class n becomes F1's f-child.

Next, we choose from { F3, F4, F5 } to be F1's t-child.

	F1	F2	F3	F4	F5	Class
Example 1	t	t	f	f	f	p
Example 6	t	t	f	t	t	n

$E(\text{F3}) = 0/2\ H(0,0) + 2/2\ H(1,1) = 0/2 \times 0 + 2/2 \times 1 = 1$

$E(\text{F4}) = 1/2\ H(0,1) + 1/2\ H(1,0) = 1/2 \times 0 + 1/2 \times 0 = 0$

$E(\text{F5}) = 1/2\ H(0,1) + 1/2\ H(1,0) = 1/2 \times 0 + 1/2 \times 0 = 0$

F4 and F5 minimize the average entropy. Since we break ties in favor of features with smaller numbers, F4 becomes F1's t-child.

Now, we choose from { F3, F5 } to be F4's f-child.

	F1	F2	F3	F4	F5	Class
Example 1	t	t	f	f	f	p

Since the only example has class p, class n becomes F4's f-child.

Next, we choose from { F3, F5 } to be F4's t-child.

	F1	F2	F3	F4	F5	Class
Example 6	t	t	f	t	t	n

Since the only example has class n, class n becomes F4's t-child.

The final decision tree is shown below. It classifies example 8 (f, f, f, t, t) as belonging to class p.