

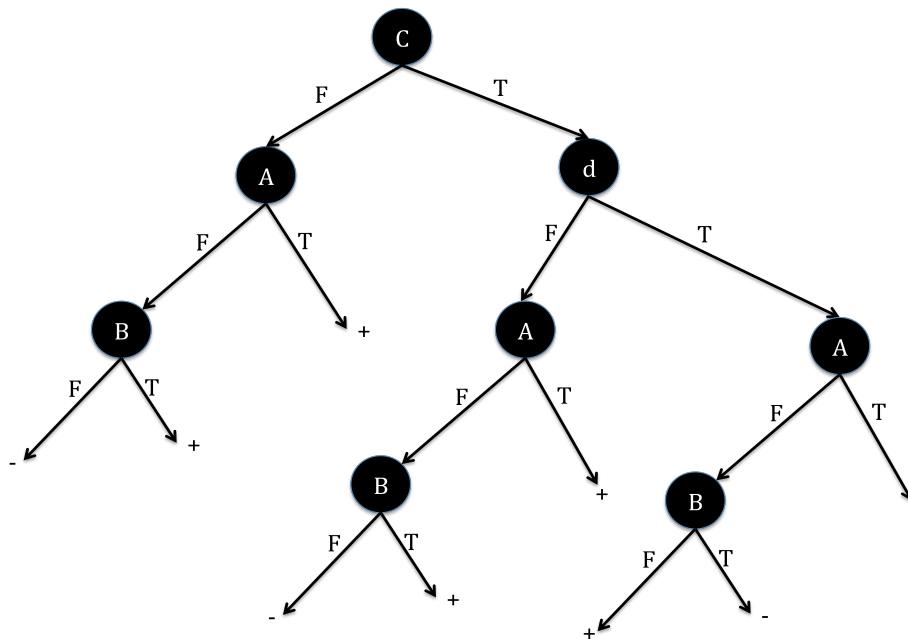
Final Exam

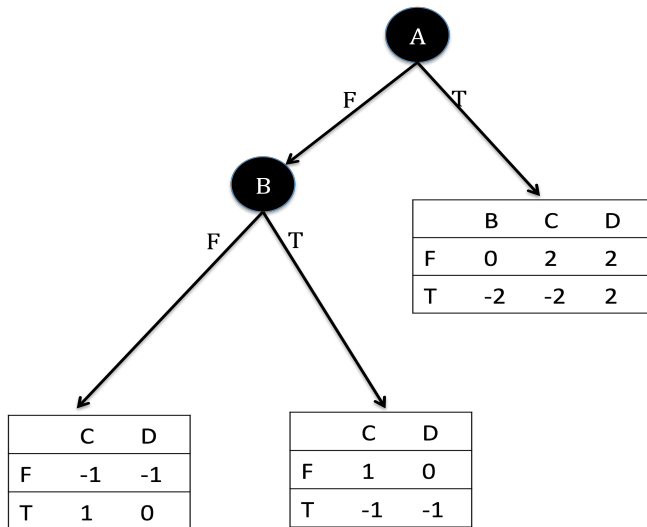
120 minutes, open book. For each question, explain your answer clearly and concisely.

Question 1 (18 points): Decision Tree and Perceptrons:

As an alternative to decision trees for classification, an approach called perceptron trees was developed. It combined together features of a decision tree with a perceptron.

The perceptron tree was similar to a decision tree except that the leaf nodes were perceptrons. It worked liked a decision tree where the splitting on an attribute occurred unit a perceptron was reached. The rest of the attributed that were not used in the higher-level decision nodes were used as an input to the perceptron. An example decision tree and the corresponding perceptron are shown below:





In the diagram above, A and B are decision nodes. The leaf nodes are perceptrons. The learned weights are shown in the boxes. So, the weight for attribute C being F is 1 when A is F and B is T.

1a (12 points) Explain why the perceptron tree is better for classification than either a decision tree or a perceptron by itself.

1b (6 points) In building the perceptron tree, the decision about whether to make a node a perceptron node or a decision node was based on whether the set of attributes (and the corresponding training instances) associated with this decision could be linearly separated. If they could not, then a decision node was used based on a specific attribute from the set, and the tree was further expanded. Explain why this is a reasonable strategy.

Question 2 (18 points) Decision Networks:

The following is a modification of one of the examples presented in class and in Homework 4a.

Suppose an oil company is hoping to buy one of 10 blocks of ocean drilling rights.

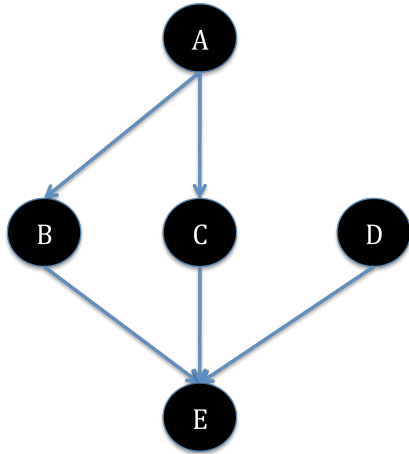
- ⇒ Exactly one block contains oil worth 1000 dollars.
- ⇒ The price of each block is 100 dollars.
- ⇒ TWO seismologists offer the company a survey.
- ⇒ Seismologist A can indicate whether block #3 contains oil. However A is only 70% accurate for the case there is oil and the case there is no oil.
- ⇒ Seismologist B can ALSO indicate whether block #3 contains oil. However B is only 85% accurate for the case there is oil and the case there is no oil.

2a (12 points): What is the value of information for combining the response from both Seismologists?

2b (6 points): Draw a decision network diagram for this scenario.

Question 3 (10 points): Bayesian Inference

An admissions committee for a college is trying to determine the probability that an admitted candidate is really qualified; the relevant probabilities are given in the Bayesian Network shown here.



A = Applicant is qualified

B = Applicant has a high grade point average

C = Applicant has excellent recommendations

D = Applicant has a parent who is an alumni of the college

E = Applicant is admitted.

Represent $p(A|D,E)$ in-terms of the conditional probabilities given in the Bayesian Network above. Make sure even the “constant” can be represented using the Bayesian Network.

Question 4 (18 points) Reinforcement Learning

Consider the deterministic world below. Arrows show allowable moves. There is a reward of 12 units for entering the top-right most state. There is no reward for being in any other state. Given the $Q(s,a)$ values in the figure below,

4a) (12 points) show the changes in the $Q(s,a)$ estimates for the first two steps in the path shown by the dotted line (the agent starts in the lower left cell) when discount factor = 0.5.

4b) (6 points): Explain why the new Q value computed in the third step is not affected by the new Q value computed in the second step.

