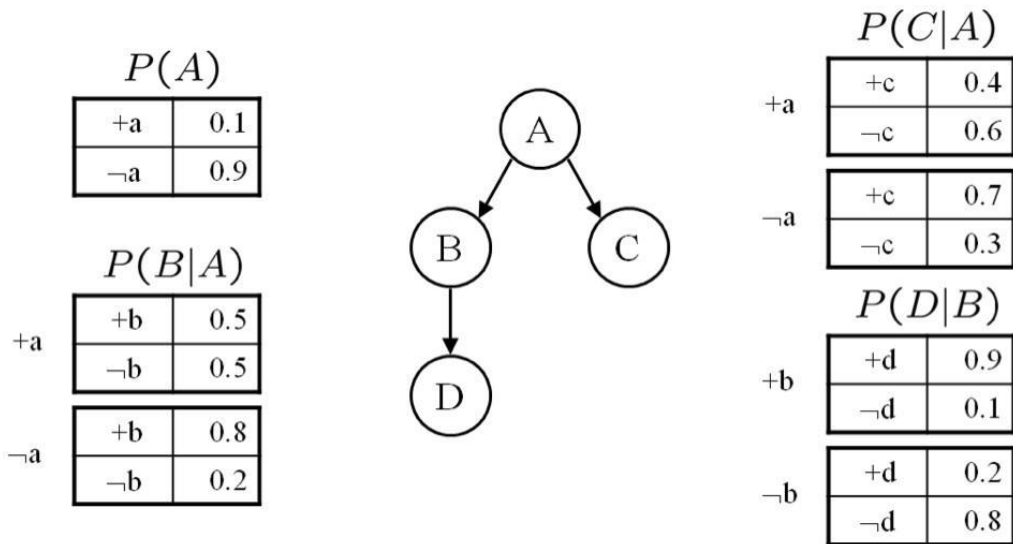




## Reasoning about Uncertainty (30pts)

**Bayesian Networks (20pts)** Given this network calculate the following probabilities. Give both the formula and calculations with values. These questions are designed so that they can be answered with a minimum of computation. If you find yourself doing a copious amount of computation for each part, step back and consider whether there is simpler way to deduce the answer.



1. [4pts]  $P(a, \neg b, c, \neg d)$

$$= P(a)P(\neg b|a)P(c|a)P(\neg d|\neg b) \quad +2\text{pts}$$

$$= 0.1 \times 0.5 \times 0.4 \times 0.8 = 0.016 \quad +2\text{pts}$$

2. [4pts]  $P(b)$

$$= \sum_{A=\{a, \neg a\}} P(A)P(b|A) \quad +2\text{pts}$$

$$= 0.1 \times 0.5 + 0.9 \times 0.8 = 0.77 \quad +2\text{pts}$$

**Bayesian Networks continued**

3. [4pts]  $P(a|b)$

$$= P(a) P(b|a) / P(b) \quad +2pts$$

$$= 0.1 \times 0.5 / .77 = 0.064935 \quad +2pts$$

4. [4pts]  $P(d|a)$

$$= \sum_{B=\{b, \neg b\}} P(d|B)p(B|a) \quad +2pts$$

$$= 0.9 \times 0.5 + 0.2 \times 0.5 = 0.55 \quad +2pts$$

5. [4pts]  $P(d|a,c)$

From the conditional independence properties of the graph,  $D \perp C | \{A\}$ . Hence,  
 $P(d|a,c) = p(d|a)$  +2pts

$$= 0.55 \quad +2pts$$

**Probability (10pts)**

Circle **True** for each of the following statements if it is always true.

1. [2pts] [True/False]  $P(A, B) = P(A)P(B)$

2. [2pts] [True/False]  $P(A|B) = P(A)P(B)$

3. [2pts] [True/False]  $P(A, B) = P(A)P(B) - P(A|B)$

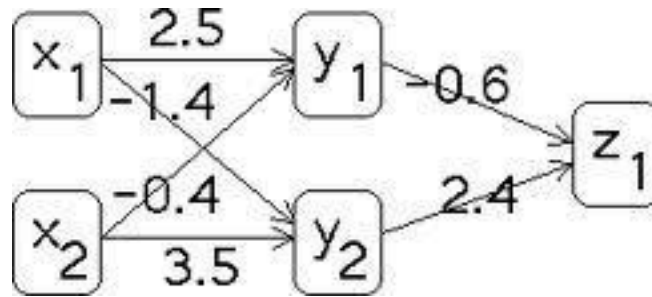
4. [2pts] [True/False]  $P(A, B, C) = P(A|B, C)P(B|C)P(C)$

5. [2pts] [True/False] If  $P(A|B, C) = P(B|A, C)$  then  $P(A|C) = P(B|C)$

## Machine Learning (50pts)

### Neural Networks (10pts)

In this network the input layer consists of units  $X_1$  and  $X_2$ . The hidden layer is  $Y_1$  and  $Y_2$ , and the output layer is  $Z_1$ . Units  $Y_1$ ,  $Y_2$  and  $Z_1$  all have threshold functions where  $t > 0$  for the unit to activate.



1. [1pt] When input  $X_1 = 0$  and input  $X_2 = 0$ , what does the network output?  
0
2. [1pt] When input  $X_1 = 0$  and input  $X_2 = 1$ , what does the network output?  
1
3. [1pt] When input  $X_1 = 1$  and input  $X_2 = 0$ , what does the network output?  
0
4. [1pt] When input  $X_1 = 1$  and input  $X_2 = 1$ , what does the network output?  
1
5. [3pts] Can a weight be changed for the network to compute the Boolean function OR?  
If yes, which weight and what value should it be changed to?

Yes. +1

-0.6 to number  $> 0$  or -1.4 to number  $> 0$  +2

6. [3pts] Can a single weight be changed for the network to compute the Boolean function XOR? If yes, which weight and what value should it be changed to?

No. +3

### Candidate Elimination (20pts)

Reward Card Domain: This domain is a deck of 52 cards (13 cards of each suit). A subset of the deck are considered “Reward” cards. The goal is given positive and negative examples of Reward cards learn the concept that represents the Reward cards only.

$(r=1) \vee \dots \vee (r=10) \vee (r=J) \vee (r=Q) \vee (r=K) \hat{=} \text{ANY-RANK}(r)$

$(r=1) \vee \dots \vee (r=10) \hat{=} \text{NUM}(r)$

$(r=J) \vee (r=Q) \vee (r=K) \hat{=} \text{FACE}(r)$

$(s=\spadesuit) \vee (s=\heartsuit) \vee (s=\clubsuit) \vee (s=\diamondsuit) \hat{=} \text{ANY-SUIT}(s)$

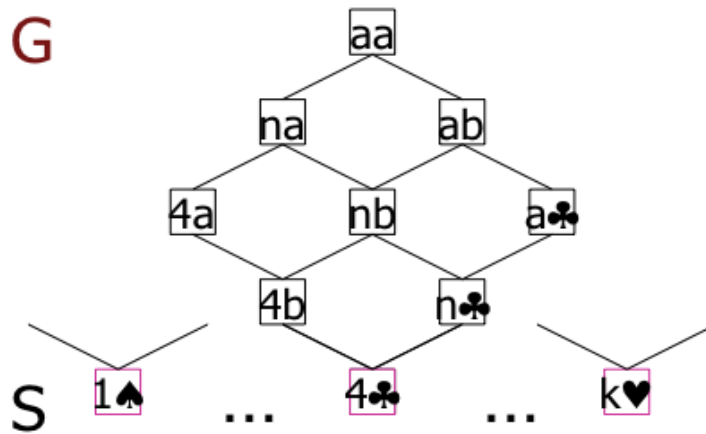
$(s=\spadesuit) \vee (s=\heartsuit) \hat{=} \text{BLACK}(s)$

$(s=\clubsuit) \vee (s=\diamondsuit) \hat{=} \text{RED}(s)$

A hypothesis is represented by  $rs$ , with:

- $r \in \{a, n, f, 1, \dots, 10, j, q, k\}$ , where  $a=\text{ANY-RANK}$ ,  $n=\text{NUM}$ ,  $f=\text{FACE}$
- $s \in \{a, b, r, \spadesuit, \heartsuit, \clubsuit, \diamondsuit\}$ , where  $a=\text{ANY-SUIT}$ ,  $b=\text{BLACK}$ ,  $r=\text{RED}$

The hypothesis space for the Reward Card domain can be seen as a partial order, where  $G$  is the set of most general hypotheses ( $aa = \text{ANY-RANK, ANY-SUIT}$ ), and  $S$  is the set of most specific hypotheses (4♣). Given this subset of the partial order, answer the questions below.



- [2pts] Given 4♣ is a **negative** example of a Reward card, list the hypotheses that would classify 4♣ as a true negative.

1♠ k♥

-1 for extra incorrect hypotheses

### Candidate Elimination continued

2. [2pts] Given 7♣ is a positive example of a Reward card, list the hypotheses that would classify 7♣ as a true positive.

aa, ab, na, nb, ac, nc

+1 for 1-5 correct.  
+2 for all 6

-1 for extra incorrect hypotheses

3. [2pts] Given 5♥ is a negative example of a Reward card, list the hypotheses that would classify 5♥ as a false positive.

aa,+1

na +1

-1 for extra incorrect hypotheses

### Candidate Elimination continued

4. [2pts] Given  $2\spadesuit$  is a positive example of a Reward card, list the hypotheses that would classify  $2\spadesuit$  as a false negative.

$1\clubsuit$   $k\heartsuit$   $4\clubsuit$   $4b$   $n\clubsuit$   $a\clubsuit$   $4a$

+1 for 1-6 correct.

+2 for all 7

-1 for extra incorrect hypotheses

5. [4pts] Given this set of training examples for Reward card, ( $4\clubsuit$  is a positive example,  $7\clubsuit$  is a positive example,  $5\heartsuit$  is a negative example,  $2\spadesuit$  is a positive example), list the hypotheses that would correctly classify all examples.

$nb, +2$

$ab +2$

-1 for extra incorrect hypotheses

### Candidate Elimination continued

6. [4pts] The example  $j\spadesuit$  is added to the training set for Reward card. If it is a positive example, list the hypotheses that would correctly classify all examples. How many of the 52 cards are Reward cards?

ab, +2  
26 cards +2

-2 for extra incorrect hypotheses

7. [4pts] The example  $j\spadesuit$  is added to the training set for Reward card. If it is a negative example, list the hypotheses that would correctly classify all examples. How many of the 52 cards are Reward cards?

nb, +2  
20 cards +2

-2 for extra incorrect hypotheses

### Reinforcement Learning (20pts)



Given a Gridworld domain, where terminal states (1,3), (4,3), and (4,2) have rewards 50, 500, and -50 respectively, the set of possible actions are {N,E,S,W, or X for terminal states}, the agent moves deterministically, all V and Q values for non terminal states have been initialized to 0, answer the questions below.

3	50			500
2				-50
1				
	1	2	3	4

Circle the letter that corresponds to the best answer for the question:

1. [4pts] What are the optimal values,  $V^*$  of each state in the above grid if  $\gamma = 0.5$ ,  $c(a)=0$ ,  $R(s)=0$  for non terminal states?

(Remember  $V_{t+1}(s) = R(s) + \text{Max}_{a \in A} \{c(a) + \gamma \sum_{s' \in S} \text{Pr}(s'|a,s) V_t(s')\}$  )

- a.  $V_{(1,1)}=15.75$ ,  $V_{(1,2)}=25$ ,  $V_{(2,1)}=31.25$ ,  $V_{(2,3)}=125$ ,  $V_{(3,1)}=62.5$ ,  $V_{(3,2)}=125$ ,  $V_{(3,3)}=250$ ,  $V_{(4,1)}=25$
- b.  $V_{(1,1)}=12.5$ ,  $V_{(1,2)}=25$ ,  $V_{(2,1)}=31.25$ ,  $V_{(2,3)}=125$ ,  $V_{(3,1)}=62.5$ ,  $V_{(3,2)}=125$ ,  $V_{(3,3)}=250$ ,  $V_{(4,1)}=31.25$
- c.  $V_{(1,1)}=15.75$ ,  $V_{(1,2)}=25$ ,  $V_{(2,1)}=31.25$ ,  $V_{(2,3)}=125$ ,  $V_{(3,1)}=62.5$ ,  $V_{(3,2)}=125$ ,  $V_{(3,3)}=250$ ,  $V_{(4,1)}=31.25$
- d.  $V_{(1,1)}=12.5$ ,  $V_{(1,2)}=25$ ,  $V_{(2,1)}=25$ ,  $V_{(2,3)}=25$ ,  $V_{(3,1)}=50$ ,  $V_{(3,2)}=100$ ,  $V_{(3,3)}=250$ ,  $V_{(4,1)}=25$
- e. None of the above

E or C correct

Answer:  $V_{(1,1)}=15.625$ ,  $V_{(1,2)}=25$ ,  $V_{(2,1)}=31.25$ ,  $V_{(2,3)}=125$ ,  $V_{(3,1)}=62.5$ ,  $V_{(3,2)}=125$ ,  $V_{(3,3)}=250$ ,  $V_{(4,1)}=31.25$

### Reinforcement Learning (continued)

2. [4pts] What are the Q values of state (3,2) in the above grid if  $\gamma = 0.5$ ,  $c(a)=0$ ,  $R(s)=-2$  for non terminal states?

(Remember  $Q_{t+1}(a,s) = R(s) + c(a) + \gamma \sum_{s' \in S} \Pr(s'|a,s) Q_t(a',s')$  )

- a.  $Q_{((3,2),N)}=120$ ,  $Q_{((3,2),E)}=-27$ ,  $Q_{((3,2),S)}=58$
- b.  $Q_{((3,2),N)}=120$ ,  $Q_{((3,2),E)}=-27$ ,  $Q_{((3,2),S)}=27$
- c.  $Q_{((3,2),N)}=125$ ,  $Q_{((3,2),E)}=-25$ ,  $Q_{((3,2),S)}=62.5$
- d.  $Q_{((3,2),N)}=120$ ,  $Q_{((3,2),E)}=-27$ ,  $Q_{((3,2),S)}=31.5$
- e. None of the above

E or B correct

Answer:  $Q_{((3,2),N)}=122$ ,  $Q_{((3,2),E)}=-27$ ,  $Q_{((3,2),S)}=27.5$

### For questions 3-5:

When the agent is not given the map with the reward function, it starts Q-learning with these 3 training examples or episodes/epochs:

E1: (1,1), N, 0, (1,2), N, 0, (1,3), X, 50

E2: (1,1), N, 0, (1,2), N, 0, (1,3), X, 50

E3: (1,1), N, 0, (1,2), S, 0, (1,1), E, 0, (2,1), E, 0, (3,1), E, 0, (4,1), N, 0, (4,2), X, -50

During Q-Learning each example is processed in the order the states are observed, for example in E1 the Q values are calculated for the observed states in this order (1,1), (1,2) (1,3).

Remember that the Q-Learning equation is:

$$Q(a,s) \leftarrow Q(a,s) + \alpha (R(s) + \gamma \max_{a' \in A} Q(a',s') - Q(a,s))$$

3. [4pts] Given  $\alpha=1$  and  $\gamma = 0.5$ , what are the non-zero Q values after observing all 3 of these examples?

- a.  $Q_{((1,1),N)}=50$ ,  $Q_{((1,2),N)}=50$ ,  $Q_{((1,3),X)}=50$ ,  $Q_{((4,2),X)}=-50$
- b.  $Q_{((1,1),N)}=12.5$ ,  $Q_{((1,2),N)}=25$ ,  $Q_{((1,3),X)}=50$ ,  $Q_{((4,2),X)}=-50$
- c.  $Q_{((1,1),N)}=12.5$ ,  $Q_{((1,2),S)}=6.25$ ,  $Q_{((1,2),N)}=25$ ,  $Q_{((1,3),X)}=50$ ,  $Q_{((4,2),X)}=-50$
- d.  $Q_{((1,1),N)}=12.5$ ,  $Q_{((1,2),S)}=12.5$ ,  $Q_{((1,2),N)}=25$ ,  $Q_{((1,3),X)}=50$ ,  $Q_{((4,2),X)}=-50$
- e. None of the above

### Reinforcement Learning (continued)

3 2 1	3	50	W=0 S=0	W=0 E=250 S=125	500
	2	N=25 S= 7.875		N=125 E=-25 S=31.5	-50
	1	N=12.5 E=15.75	E=31.25 W=7.875	N=62.5 W=3.9375 E= -12.5	N=-25 W=31.5
		1	2	3	4

4. [4pts] After Q-Learning with many more examples, the Q values are shown in the grid above. What is the optimal policy given these values?
- (1,1) N, (1,2) N, (2,1) E, (3,1) N, (4,1) W, (3,2) N, (3,3) E
  - (1,1) E, (1,2) S, (2,1) E, (3,1) N, (4,1) W, (3,2) N, (3,3) E
  - (1,1) E, (1,2) N, (2,1) E, (3,1) N, (4,1) W, (3,2) N, (3,3) E
  - All of the above
  - None of the above
5. [4pts] If now there is a “living reward” of -5 for all non terminal states ( $R(s) = -5$ ), so that all new examples would see a -5 instead of 0 at non terminal states. How would this affect the agents policy?
- It would prefer actions providing the shortest path to state (4,2)
  - It would prefer actions providing the shortest path to state (4,3)
  - It would prefer actions providing the shortest path to state (1,2)
  - All of the above
  - None of the above

## Short Answer (20pts)

**Keep your answers brief, one or two sentences.**

1. [4pts] List 2 reasons why machine learning is needed?

Description of 2 of the following:

Unknown environments

Adaptability

Lazy programmers

Autonomous

Human Cognition

See also Lecture 10 slide 4

-2 for additional incorrect answer

2. [4pts] What is Ockham's razor? How is it used in decision tree learning?

Bias for simplest hypothesis +2

Using information gain metric to create smaller decision trees +2

-2 for additional incorrect answer

3. [4pts] What does the size of a hypothesis class have to do with generalization and overfitting?

Many possible answers - this is one example:

Smaller trees allow for more generalization of the concept +2

Larger trees allow for overfitting to the training data +2

-2 for additional incorrect answer

4. [4pts] In reinforcement learning, why can it be useful to sometimes act in a way that is believed to be suboptimal?

To avoid being stuck in a local maxima +2

choosing a suboptimal action may lead to a more rewarding state and improve the agent's policy and performance. +2

-2 for additional incorrect answer

5. [4pts] Choose 2 research topics from Lecture 12 on Communicating, Perceiving, and Acting and describe each topic, including the limitations/pitfalls and future directions.

Choose any 2 research topics from lecture 12 ( See lecture 12 slides 3-6):

Description of research topic +1

Description of its limitations/pitfalls and future directions +1