Homework 5

- *Homework deadline: 11:59am on December 7*

- *Please hand in a hard copy of this homework. You need not submit any code by e-mail.*
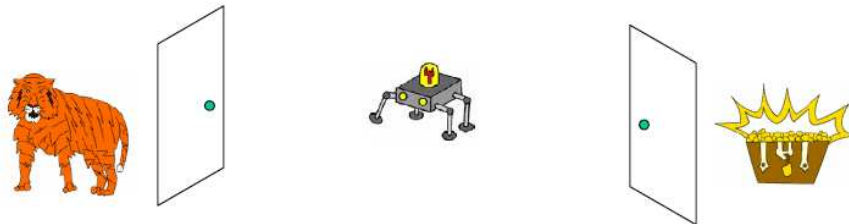
1. **POMDPs (50 pts)**

Figure 1: Tiger and pot-of-gold POMDP

Consider the popular "tiger and pot-of-gold" problem from the supplementary course reading (Kaelbling et al.). A robot is faced with the choice of opening two doors (left and right). Behind one door sits a tiger, behind the other door is a pot of gold, and the robot is uncertain which door is which. The robot can open the left door, open the right door, or listen to try to determine which door the tiger is behind. There is a cost of listening, and there is a chance that the robot will hear the tiger behind the wrong door. If the robot opens the door with the tiger, it receives a negative reward. If it opens the door with the pot-of-gold, it receives a positive reward. The world resets after the robot opens a door. The rewards are discounted after each action.

Use the following values for your formulation:

- +10 reward for opening pot-of-gold door. -100 for opening tiger door. -1 for listening.
- Opening a door resets the tiger to a 50-50 chance of being behind either door.
- When listening, the robot will hear the tiger behind the correct door with a 75% chance. It will hear the incorrect door with a 25% chance.
- The discount factor is $\gamma = 0.9$.

(a) Formulate the tiger and pot-of-gold problem as a POMDP. Describe the states, actions, transition probabilities, observations, observation probabilities, and rewards. Also draw the underlying MDP (you need not include observations on the drawing). (12 pts)

Hint: In class, we associated rewards with states. Here, you should associate rewards with performing an action while in a state. This is a more general model for POMDP and MDP reward functions.

(b) Draw the horizon 1 value function (immediate rewards) for this POMDP putting the one-dimensional belief space on the x-axis. (7 pts)

(c) Suppose you believe the tiger is behind the right door with probability 0.75, then perform the listen action and hear that the tiger is on the left. What is the new probability that the tiger is behind the right door? (8 pts total)

Now answer the same question in each of the following circumstances:

- Assuming you heard the tiger on the right instead of on the left
- Assuming you performed the open-right action instead of listening
- Assuming the initial belief was 0.3333 instead of 0.75 (and you heard the tiger on the left)

(d) Describe what happens to the value function as the horizon is extended. Draw a qualitative diagram of the value function as horizon increases. You need not do the exact calculations. (5 pts)

(e) This problem will invite you to explore the performance of greedy POMDP solving methods on the tiger and pot-of-gold POMDP. If you need more information about greedy methods, they are described in detail in Tony Cassandra's PhD document (pp 257-262):

`http://www.cs.brown.edu/research/ai/pomdp/papers/thesis-dbl.ps.gz`

i. What will happen if you use the "most-likely" heuristic to determine your actions in this POMDP (break ties randomly)? Discuss the behavior for several different starting belief states (not just 50-50). Are these actions optimal? Why or why not? (5 pts)

ii. What will happen if you use the "voting" heuristic to determine your actions in this POMDP? Discuss the behavior for several different starting belief states. Do these actions provide a higher reward than the "most-likely" heursitic? (5 pts)

iii. Describe which actions QMDP will choose for this POMDP for several different belief states. Are these actions optimal? If not, could you use the dual-mode heuristic to improve performance? Describe why or why not. (8 pts)

Hint: Think about if/how QMDP will utilize the listening action.

2. **Computational biology (50 pts)**

This question is about the Viterbi algorithm on a profile hidden Markov Model (HMM). We use the notations in the book *Biological Sequence Analysis* by Durbin et. al, available as electronic reserve (pp.100-122) from CMU library, if interested. The sequence is $X = (x_1, \cdots, x_L)$, $x_i \in \{A, C, G, T\}$. Match, insert, and delete states are denoted as M_j, I_j, D_j, respectively. Transition and emission probabilities are $a_{S_1 S_2}$ and $e_{M_j}(x_i)$.
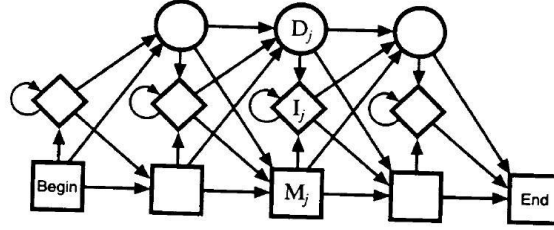
Figure 2: A profile HMM of three match states

We use the Viterbi algorithm to maximize the *likelihood ratio* of a sequence given the best alignment versus a random model R, defined as $P(x|R) = \prod_i q_{x_i}$. q_{x_i} is the background frequency, and also the emission probability of the insert state. Let $V_j^M(i), V_j^I(i), V_j^D(i)$ be log of the Viterbi dynamic programming score matrix, calculated recursively by:

$$V_0^M(0) = 0 \ , \ V_0^I(0) = -\infty \ , \ V_0^D(0) = -\infty$$

$$V_j^M(i) = \log \frac{e_{M_j}(x_i)}{q_{x_i}} + \max \begin{cases} V_{j-1}^M(i-1) + \log a_{M_{j-1}M_j} \\ V_{j-1}^I(i-1) + \log a_{I_{j-1}M_j} \\ V_{j-1}^D(i-1) + \log a_{D_{j-1}M_j} \end{cases}$$

$$V_j^I(i) = \max \begin{cases} V_j^M(i-1) + \log a_{M_j I_j} \\ V_j^I(i-1) + \log a_{I_j I_j} \\ V_j^D(i-1) + \log a_{D_j I_j} \end{cases}$$

$$V_j^D(i) = \max \begin{cases} V_{j-1}^M(i) + \log a_{M_{j-1}D_j} \\ V_{j-1}^I(i) + \log a_{I_{j-1}D_j} \\ V_{j-1}^D(i) + \log a_{D_{j-1}D_j} \end{cases}$$

(a) The Needham-Wuncsh (NW) algorithm can align a target sequence to a consensus sequence, while the Viterbi algorithm aligns a target sequence to the *conserved sites* (match states). The NW-style diagram below has the target sequence on top and the conserved sites (character heights proportional to emission probabilities of match states) on left. For NW algorithm, the dynamic programming matrix $F(i,j)$ would correspond to the nodes. What would the Viterbi matrix update correspond to in this diagram? Compare the Viterbi update formula above to the update formula of $F(i,j)$, and indicate the three Viterbi matrices $V_j^M(i), V_j^I(i), V_j^D(i), 1 \leq i, j \leq 2$ on the diagram. Hint: Try to map the match, insert, and delete operations to horizontal, vertical, and diagonal movements. (20 pts)

(b) Based on your answer above, describe at least one difference between profile HMM alignment and NW algorithm. Give an intuitive motivation or advantage of this difference. (10 pts)
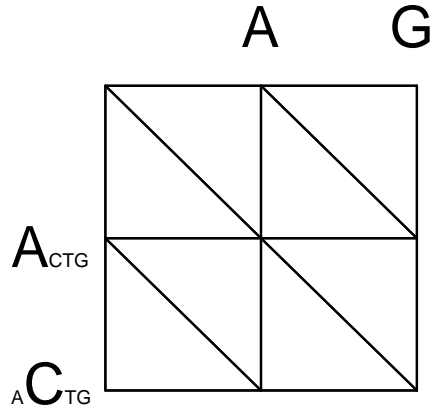
Figure 3: NW-style diagram corresponding to Viterbi matrices

(c) Consider a profile HMM of 3 match states:

$$a_{M_{j-1}M_j} = a_{D_{j-1}D_j} = a_{D_{j-1}M_j} = 1/2 \ , \ a_{M_jI_j} = a_{M_{j-1}D_j} = a_{I_{j-1}M_j} = 1/4$$

$$a_{I_jI_j} = 3/4 \ , \ a_{I_{j-1}D_j} = a_{D_jI_j} = 0$$

$$e_{M_1}(A) = e_{M_2}(C) = e_{M_3}(T) = 1/2 \ , \ e_{M_1}(T) = e_{M_2}(G) = e_{M_3}(A) = 1/4$$

$$e_{M_1}(C) = e_{M_1}(G) = e_{M_2}(A) = e_{M_2}(T) = e_{M_3}(C) = e_{M_3}(G) = 1/8$$

$$q_N = 1/4, N \in \{A, C, G, T\}$$

For simplicity, use only $\log_2$ in calculations (assume $\log_2 3/4 \approx -.4$). Also consider a NW algorithm that aligns a target sequence to the consensus sequence "ACT", using the gap penalty $d = -1$ and the substitution matrix $S(x,y) = 1$ if $x = y$, $S(A,T) = S(T,A) = S(C,G) = S(G,C) = 0$, and -1 otherwise. For example, the sequence "ATGT" will result in the same alignment by both algorithm:

```
NW algorithm target:    ATGT
NW algorithm consensus: A-CT

Profile HMM target:     ATGT
Profile HMM consensus:  A-CT
```

Write a short target sequence that results in **different** alignment by the above profile HMM and NW algorithm. Show the alignments by both algorithm, and give an intuitive explanation. You may wish to compute the Viterbi matrix for this alignment to make sure that you answer is correct, though you are not required to hand it in. Hint: Your answer above could be a clue. (20 pts)