Probabilistic Artificial Intelligence
Final Exam
Jan 28, 2019

Time limit:	120 minutes
Number of pages:	20
Total points:	100

You can use the back of the pages if you run out of space. Collaboration on the exam is strictly forbidden. Please write your answers with a *pen*.

(1 point) Please fill in your student ID and full name (LASTNAME, FIRSTNAME) in capital letters.

Please leave the table below empty.

Problem	**Maximum points**	**Obtained**
0.	1	
1.	20	
2.	17	
3.	17	
4.	15	
5.	20	
6.	10	
Total	100	

1. Short Questions (20 points)

(20 points) For each of the statements below, state whether they are true or false. Each correct answer gives +1 point, each incorrect answer gives -1 point. You cannot get less than 0 points.

(a) For all valid probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$, it holds that $\Omega \subseteq \mathcal{F}$.

☐ True ☐ False

(b) Given prior and likelihood probabilities, Bayes' Rule can be derived from the product rule only.

☐ True ☐ False

(c) For random variables X, Y, and Z, the statement $X \perp Y | Z$ implies that, for all x, y, and z, $P(X = x, Y = y | Z = z) = P(X = x | Z = z) \cdot P(Y = y | Z = z)$.

☐ True ☐ False

(d) In a sound Bayesian Network, d-separation implies conditional independence but conditional independence does not necessarily imply d-separation.

☐ True ☐ False

(e) Exact inference in Bayesian Networks is never tractable.

☐ True ☐ False

(f) Variable elimination is correct in polytrees only.

☐ True ☐ False

(g) If belief propagation converges on a Bayesian Network, all the marginals are computed correctly.

☐ True ☐ False

(h) Monte Carlo approximations always converge to the true expectations in loopy Bayesian Networks.

☐ True ☐ False

(i) When using the Gibbs Sampling variant where the variables are chosen in a deterministic order, the corresponding Markov Chain has the correct stationary distribution.

☐ True ☐ False

(j) In an ergodic Markov Chain, the stationary distribution is independent of the initial distribution.

☐ True ☐ False

(k) Suppose in a temporal model of a random variable X, the variable at time t, denoted by X_t, is marginally dependent on X_{t-2}. This model cannot be a first order Markov Chain.

☐ True ☐ False

(l) In Hidden Markov Models, the belief propagation algorithm converges to the correct marginal distributions.

☐ True ☐ False

(m) When there are no missing measurements, the Kalman Filter and the Kalman Smoother estimate the same most likely hidden states.

☐ True ☐ False

(n) Policy Iteration converges exactly to an optimal policy after a finite number of steps in finite MDPs.

☐ True ☐ False

(o) A policy might be optimal even if it is not greedy with respect to its induced value function.

☐ True ☐ False

(p) The optimal Bayesian Network structure that is learned when using the Maximum-Likelihood scoring criterion is the maximum spanning tree of the fully connected graph.

☐ True ☐ False

(q) If the Bayesian Information Criterion is used to learn the structure of a Bayesian Network from data, then, as the number of examples goes to infinity, the correct structure is identified.

☐ True ☐ False

(r) The R-Max algorithm uses epsilon-greedy exploration.

☐ True ☐ False

(s) Q-learning is a model-based Reinforcement Learning algorithm.

☐ True ☐ False

(t) For episodic tasks, a policy can be evaluated by rollouts of this policy in the environment.

☐ True ☐ False

2. D-Separation and Exact Inference (17 points)

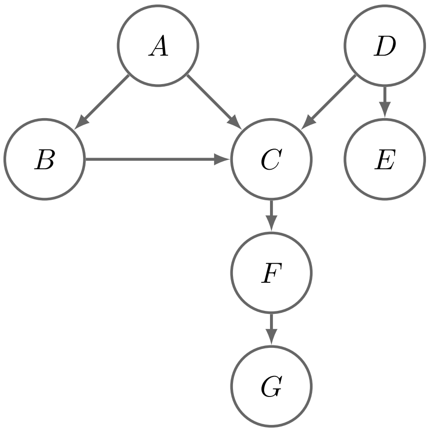


Figure 1: The Bayesian Network has binary random variables A, B, C, D, E, F and G.

(10 points) (i) For each of the statements below, state whether they are true or false. Each correct answer gives +1 point, each incorrect answer gives -1 point. You cannot get less than 0 points.

(a) There is at least one active trail from A to D with respect to no observations.

☐ True ☐ False

(b) There is at least one active trail from A to D given F.

☐ True ☐ False

(c) There is at least one active trail from A to E given F and D.

☐ True ☐ False

(d) d-sep$(A, F|C)$.

☐ True ☐ False

(e) d-sep$(A, E|G)$.

☐ True ☐ False

(f) d-sep$(B, E|A, F, G)$.

☐ True ☐ False

(g) The Bayesian Network in Figure 1 is a polytree.

☐ True ☐ False

(h) Belief Propagation can be used to perform exact inference in polytrees.

☐ True ☐ False

(i) Belief Propagation in graphs with cycles always converges, but the resulting marginals can be incorrect.

☐ True ☐ False

(j) At every iteration of Belief Propagation, the marginal probability of every variable X being equal to realization x is always proportional to the sum of messages from all neighboring factors.

☐ True ☐ False

(2 points) (ii) Write down the factorization of the joint probability density induced by the Bayesian Network.

(iii) Suppose you perform variable elimination in the Bayesian Network in Fig. 1. Determine the number of variables in the largest factor(s) for the orderings:

(2 points) (a) G, E, F, D, B, C, A

(2 points) (b) C, D, F, E, G, B, A

(1 point) (c) For which of the two orderings is variable elimination computationally cheaper?

3. Approximate Inference (17 points)

You are planning to submit the results of your Master Thesis to a top Computer Science conference. You want to estimate the probability of your paper being rejected ($R = 1$). The paper is rejected in the following ways: (1) when there are too many good papers submitted this year ($M = 1$), your paper will be rejected with probability p_m; (2) if the paper is novel ($N = 1$) it will still be rejected with probability p_n; (3) the paper can be rejected for others causes such as the conference not being the correct venue, not having a reviewer that is familiar with your topic, you exceeded the maximum number of pages, etc., with probability p_o. All three types of causes are assumed to be independent of each other.

Finally, you believe that the probability of the conference being interesting ($I = 1$) depends only on how many good papers (M) are submitted. You will always attend the conference ($A = 1$) if your paper is accepted. If rejected and the conference is interesting, you still attend with probability p_a.

I.e., the model has the following binary random variables (valued 1 if the statement is true, 0 if false):
(R) – your paper is rejected;
(N) – paper is novel;
(M) – too many good papers submitted this year;
(I) – conference is interesting;
(A) – you attend the conference.

(3 points) (i) Draw the Bayesian Network corresponding to the text above.

(2 points) (ii) Express in terms of p_m, p_n, and p_o the probability of the paper being accepted given that the paper is novel but too many good papers where submitted this year.

(iii) You now want to answer the following query: What is the probability of the paper being accepted given that the paper is novel and the conference is interesting? Instead of computing this probability in closed form, you decide to use a Monte Carlo approximation. For this, you collect T samples from the Bayesian Network via Monte Carlo Sampling.

(2 points) (a) Which variables have to be fixed and which variables have to be sampled from when using Monte Carlo Sampling to solve this query?

(2 points) (b) How do you estimate this query from the samples using Rejection Sampling?

(2 points) (c) What is the problem with Rejection Sampling to solve this query?

(iv) To solve this problem, you decide to run the variant of Gibbs Sampling where the variables are chosen in random order.

(2 points) (a) Which variables have to be fixed and which variables have to be sampled from when using the Gibbs sampler to solve the query of question (iii)?

(2 points) (b) Given that you collected T samples with the Gibbs sampler, how would you estimate the query of question (iii)?

(2 points) (c) In iteration t, the Markov Chain state of the sampler is $(R = 1, N = 1, M = 1, I = 1, A = 0)$. Compute the probability that, at iteration $t + 1$, the state is $(R = 0, N = 1, M = 1, I = 1, A = 0)$.

4. Hidden Markov Models (15 points)

Its 1865 and Atlantic Telegraph Company is sending messages over the Atlantic cable from London to New York using a telegraph. The communication is inefficient, and the noise in the transmission often causes messages to be wrongly interpreted. Imagine you are working at the New York office as an engineer, and your manager tasks you to design an error correcting system for the current communication model.

Only *bits* are communicated over the telegraph i.e. 0 or 1. Due to the noise in the system, 0 can be corrupted to 1 with probability 0.1, and 1 to 0 with probability 0.2.

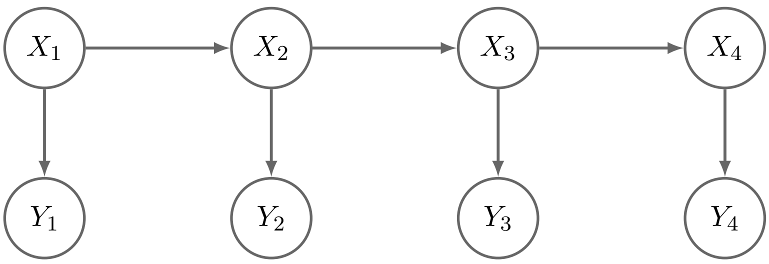
The communication proceeds with messages where each message consists of 4 bits. In this exercise, we will consider a vocabulary of **equally likely** words:

$$V = \{(1010), (0101), (1001)\}.$$

Each message is separated with a pause to mark separation.

As you just finished your PAI class (in 1865!), you decide to use a Hidden Markov Model to design this error correcting system. You decide to build a Hidden Markov Model that explains the transitions and noisy observations. Finally, you use the most probable word from within the vocabulary given the observations to return the error corrected word.

The Hidden Markov Model uses binary random variables X_k, for $k = 1, 2, 3, 4$ to represent the k-th bit of the message to be transmitted from London. In New York, a corrupted version of it is observed. This is denoted with Y_k. Finally, the bit transitions from k to $k + 1$ are modelled as Markovian. The resulting graphical model is:



(1 point) (i) Given the vocabulary V, what is the Maximum Likelihood estimate of $P(X_1 = x)$, for $x \in \{0, 1\}$? *(Hint: Treat elements in V as identically, independently distributed samples of $P(X_{1:4})$. We use the notation $X_{1:4}$ to denote the variables (X_1, X_2, X_3, X_4)).*

x	$P(X_1 = x)$
0	
1	

(2 points) (ii) Given the vocabulary V, what is the Maximum Likelihood estimate of $P(X_{k+1} = x'|X_k = x)$, for $x, x' \in \{0, 1\}$? Assume the HMM is stationary, i.e., the above probabilities do not depend on k. *(Hint: Treat elements in V as identically, independently distributed samples of $P(X_{1:4})$. We use the notation $X_{1:4}$ to denote the variables (X_1, X_2, X_3, X_4))*

x	x'	$P(X_{k+1} = x'\|X_k = x)$
0	0	
0	1	
1	0	
1	1	

(6 points) (iii) Using the Hidden Markov Model, and the probabilities in points (i) and (ii), which of the messages in V is the most probable message given the observation 1101?

(iv) The time between the bits of a message is too long and your manager does not want to wait for the whole message to arrive. Your colleague proposes to use Bayesian Filtering on the same Hidden Markov Model to estimate the probability of each of the hidden states as the message arrives.

(2 points) (a) What is $P(X_1 = x|Y_1 = 1)$, for $x \in \{0, 1\}$?

(4 points) (b) What is $P(X_2 = x|Y_1 = 1)$, for $x \in \{0, 1\}$?

5. Push Your Luck (20 points)

Consider the following game played over multiple rounds. In each round, you have the option to either push your luck (P) or leave (L). If you leave, you obtain 0 reward and the game ends. Otherwise, you draw a card uniformly at random from a deck of cards. In the deck of cards, you have lucky and unlucky cards. When a lucky card is drawn, you obtain a positive reward shown on the card, the card is removed from the deck, and you continue to the next round. When an unlucky card is drawn, you obtain a negative reward shown on the card, and the game ends immediately.

For all the following questions, we consider only two rounds of the game, that is, the game ends after (at most) two choices of P or L. Furthermore, the deck consists of three cards, two lucky and one unlucky card, with rewards g_1, g_2, and $-c$, respectively, where $g_1, g_2, c > 0$.

(6 points) (i) Draw an MDP diagram that encodes this game, annotating the transitions with the corresponding actions, transition probabilities, and associated rewards.

(7 points) (ii) Assume that $g_1 > c$ and $g_2 < c$. For the undiscounted version ($\gamma = 1$), compute the optimal values of all (non-terminal) states by running value iteration until convergence. Initialize all values to 0.

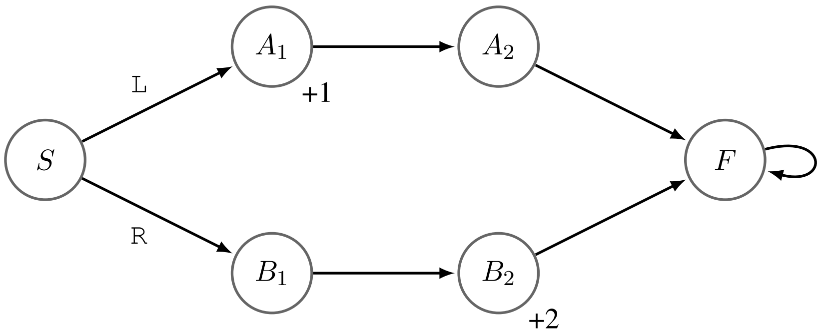
(Hint: The Value Iteration update is $V_{k+1}(x) = \max_a \left\{\sum_{x'} P(x' \mid x, a)(r(x, a, x') + V_k(x'))\right\}$*).*

(4 points) (iii) Compute the optimal policy using the result of the previous question.

(3 points) (iv) If $g_1 < c$ and $g_2 < c$, but $g_1 + g_2 > c$, briefly show how the optimal values and the optimal policy computed in the previous questions change.

6. Reinforcement Learning (10 points)

Consider the MDP represented by the transition graph shown below. The agent starts in state S, where two actions, `L` and `R`, are available. All other states have a single default action that leads to the next state as indicated. There is a +1 reward in state A_1, a +2 reward in state B_2 and a reward of 0 in all other states.



In the following, we consider two policies π_L and π_R, which take action `L` and `R` in state S respectively.

(2 points) (i) Compute the value function $V_\gamma(S)$ for both policies π_L and π_R in state S, with arbitrary discount factor $0 < \gamma \leq 1$.

(1 point) (ii) For what values of γ is the π_L policy optimal?

We would like to find the optimal policy for the MDP but we don't know the transition model nor the reward function. We do know the state and action space, and that the transition model and the reward function are deterministic. Furthermore, we know that the reward function only depends on the current state. For training, we use episodes of length 4. In each episode, the agent starts in state S and finishes in state F. The rewards do not accumulate from episode to episode.

(4 points) (iii) Assume that Q-learning is used to estimate the optimal $Q^\star$-function. Suppose we initialize our estimate $\hat{Q} = 0$ for all state-action pairs and use π_L as exploratory policy.

Does the estimate converge to the optimal Q-values? Does this change if the exploratory policy was π_R? If it does not, how would you solve it? Justify your answers.

(iv) Assume the transition model of the MDP is now known and the reward function is still unknown. However, we know that the rewards are bounded between 0 and 2, are deterministic, and depend only on the current state. We apply a variant of the Rmax algorithm where before each episode the MDP is solved exactly with the correct transition model of the MDP. On each state visited, we observe the associated reward, which allows us to immediately update $\hat{R}$ to the correct value.

(1 point) (a) How should the algorithm be initialized?

(2 points) (b) How many episodes does this algorithm need to find the optimal policy in the MDP above? Justify your answers.