

Questions

Probabilistic Models
Decision Problems
Sequential Problems
Model Uncertainty
State Uncertainty

Probabilistic Models

[↑]

Problem 1

Consider the definition of conditional probability:

$$P(A, B) = P(A|B)P(B)$$

Can you come up with a simple explanation in words as to why this works? Use similar reasoning to come up with an expression for $P(A, B|C)$.

Problem 2

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammographies. 9.6% of women without breast cancer will also get positive mammographies. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

Problem 3

There is a 50% chance there is both life and water on Mars, a 25% chance there is life but no water, and a 25% chance there is no life and no water. What is the probability that there is life on Mars given that there is water?

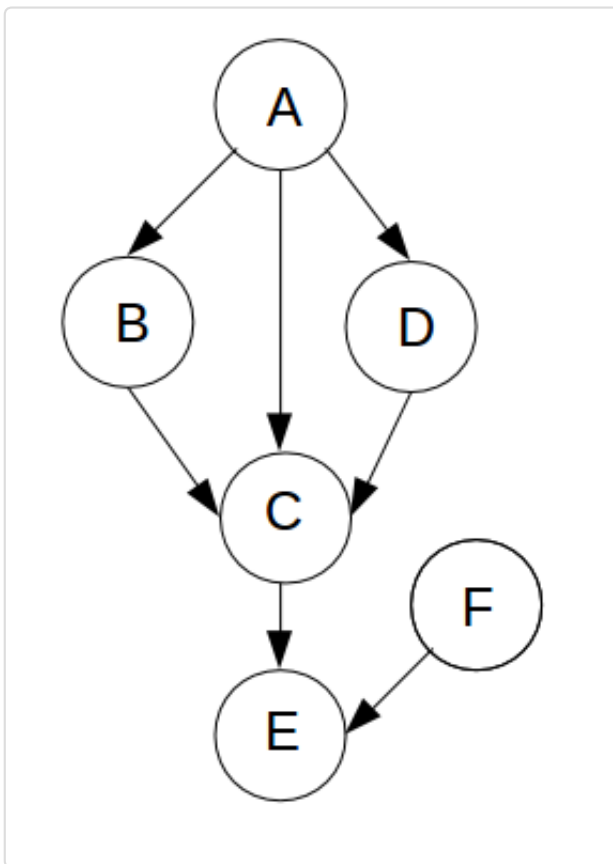
Problem 4

In the textbook, it is stated that if all variables in a Bayesian network are binary, the probability distribution over some variable X with n parents Pa_X can be represented by 2^n independent parameters.

Imagine that X is a binary variable with two parent variables that are not necessarily binary. Imagine that the first parent can assume three different values, and that the second can assume two values. How many independent parameters are needed to represent this distribution ($P(X|\text{Pa}_X)$)? How many would it be if you added another parent that can assume four different values?

Now assume that X itself is not binary, but can assume three different values (and still has the three parents as specified above). How many values are needed to represent this distribution? Can you come up with a general rule for the number of independent parameters needed to represent a distribution over some variable X with parents Pa_X ?

Problem 5



(images/problem3.png)

Given the displayed Bayes net, determine whether the following are true or false:

1. $(B \perp D | A)$
2. $(B \perp D | C)$
3. $(B \perp D | E)$
4. $(B \perp C | A)$

Problem 6

It is known that 80-foot blue whales consume, on average, 3200 kg of krill per day. 100-footers consume on average 3600 kg of krill per day. Assume that the mean daily krill consumption varies linearly with whale length and that the daily consumption for a given whale follows a Gaussian distribution with a standard deviation of 200 kg of krill per day. Define the linear Gaussian distribution, $P(k | l)$, relating the rate of krill consumption k to whale length l .

Problem 7

Assuming a hidden Markov model with states $s_{0:t}$ and observations $o_{0:t}$, prove the following:

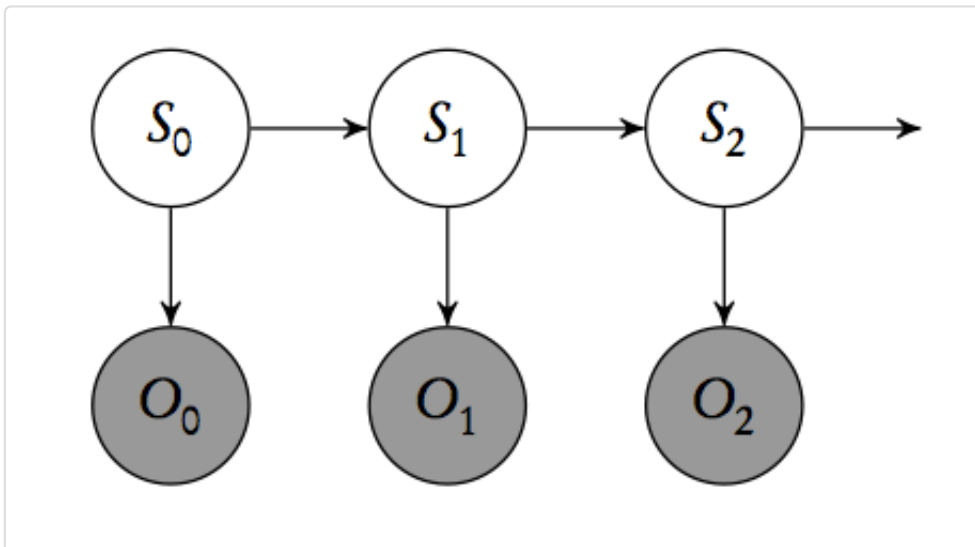
$$P(s_t | o_{0:t}) \propto P(o_t | s_t, o_{0:t-1})P(s_t | o_{0:t-1})$$

Starting from the previous equation, prove the following:

$$P(s_t | o_{0:t}) \propto P(o_t | s_t) \sum_{s_{t-1}} P(s_t | s_{t-1}) P(s_{t-1} | o_{0:t-1})$$

Problem 8

What is the Markov blanket for some node o_t of the hidden Markov model below? Explain why this is so.



(images/hmm.png)

Problem 9

One possible representation of the law of total probability is:

$$P(A) = \sum_{B \in B_{set}} P(A | B)P(B)$$

where B_{set} is a set of mutually exclusive and exhaustive propositions. Can you find a similar expression for $P(A | C)$?

Problem 10

What is a topological sort? Why is it important to perform a topological sort before sampling from a Bayesian network? Does a topological sort always exist? Is a topological sort always unique?

Problem 11

Formulate the following 3SAT problem as a Bayesian network.

$$F(x_1, x_2, x_3, x_4) = (x_1 \vee x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee \neg x_4) \wedge (x_2 \vee x_3 \vee x_4)$$

This shows that inference in Bayesian networks is at least as hard as 3SAT. If 3SAT is NP-complete, what does that make inference in Bayesian networks?

Problem 12

What are the differences between inference, parameter learning, and structure learning? What are you looking for in each case and what is assumed to be known? When might you use each of them?

Problem 13

What is a classification task? Assume you are classifying using a naive Bayes model. What assumptions are you making? Draw a naive Bayes model using the compact representation shown in class. What is the name of this kind of representation?

Problem 14.1

What is an important drawback of maximum likelihood estimation?

Problem 14.2

Bayesian parameter learning estimates a posterior $p(\theta \mid D)$ for the parameters θ given the data D . What are some of its advantages and drawbacks?

Problem 15

What is the gamma function Γ ? What is $\Gamma(5)$?

Problem 16

Imagine that you want to estimate θ , the probability that one baseball team (call them Team A) beats another team (call them Team B). Assume you know nothing else about the two teams. What is a reasonable prior distribution?

Now imagine that you know the two teams well and are confident that they are evenly matched. Would a prior of Beta(9,9) be better than a Beta(2,2) in this case? If so, why?

Now imagine that you know that Team A is more likely to win (maybe they are the Warriors). What kind of prior might you use in this case? Imagine that the teams are going to play many games against each other. What does this mean for the prior you select?

Problem 17

Consider the two Beta distributions Beta(2,2) and Beta(9,9). Beta(9,9) gives much more weight to $\theta = 0.5$. Can you explain intuitively why this is so?

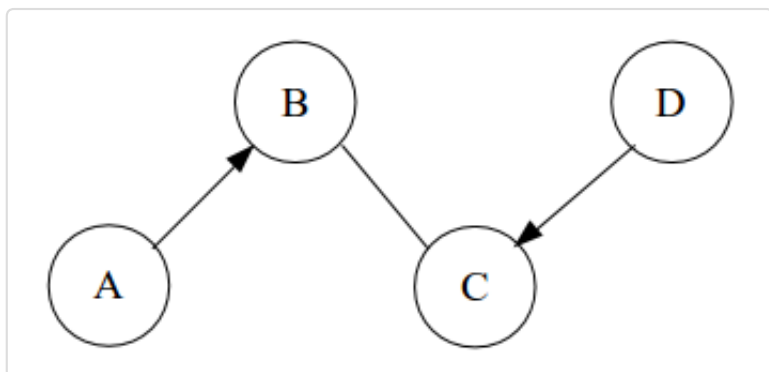
Problem 18

Suppose you have a lot of data and are trying to learn the structure of a Bayesian network that fits this data. Consider two arbitrary Bayesian network designs. One is relatively sparse, whereas the other has many connections between its nodes.

Imagine that your data consists of very few samples. Which Bayesian network would you expect to achieve a better Bayesian score? How would this change if there were many samples?

Problem 19

How many members are there in the Markov equivalence class represented by the partially directed graph shown below?



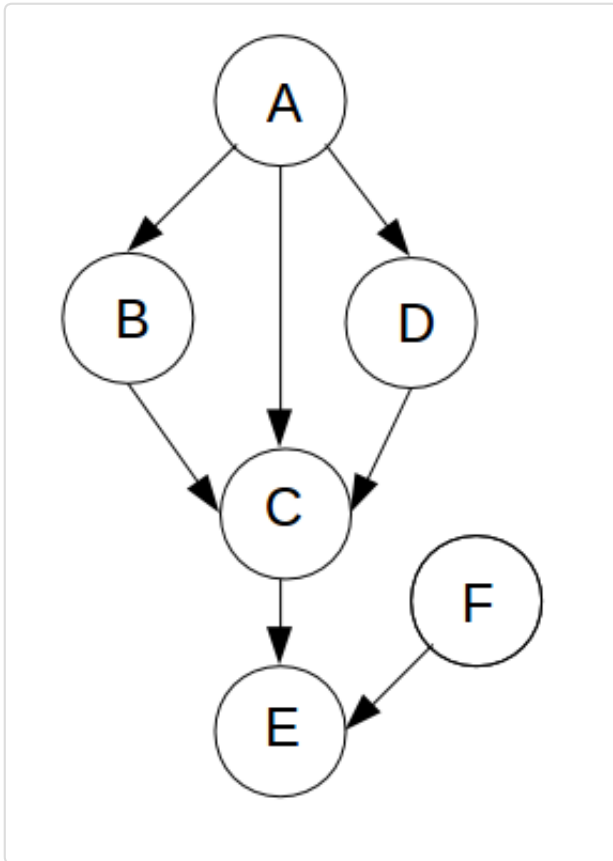
(images/equivalence_class.png)

Problem 20

Gibbs sampling offers a fast way to produce samples with which to estimate a distribution. What are some downsides of Gibbs sampling and how are they handled?

Problem 21

What is a topological sorting of the nodes shown in Question 5 (recreated below)?



(images/problem3.png)

Decision Problems

[↑]

Problem 1

What does it mean to be rational?

Problem 2

Explain the value of information in words. What is the value of information of an observation that does not change the optimal action? Imagine that the optimal action changes after an observation. What does this say about the value of information of that observation?

Problem 3

The prisoners dilemma is an example of a game with a dominant strategy equilibrium. Imagine that the game is modified so that if one prisoner testifies the other only gets four years of prison instead of ten. Does this game still have a dominant strategy equilibrium? Are there any other equilibria?

Problem 4

Explain why the traveler's dilemma has a unique Nash equilibrium of \$2. Draw the utility matrix and use it to show the equilibrium.

Sequential Problems

[↑]

Problem 1

What is the Markov assumption? What does a Markov decision process consist of? What is a stationary MDP? Draw a compact representation of a stationary MDP.

Problem 2

What is the purpose of the discount factor in infinite horizon problems? What is an alternative to using a discount factor in infinite horizon problems? What effect does a small discount factor have? What about a large one? When is one preferable to the other?

Problem 3

Does the optimal policy have to be unique? Does the optimal value for each state have to be unique?

Problem 4

What is the Bellman equation? How does it simplify if transitions are deterministic?

Problem 5

The policy evaluation equation in matrix form is

$$\mathbf{U}^\pi = (\mathbf{I} - \gamma \mathbf{T}^\pi)^{-1} \mathbf{R}^\pi$$

where \mathbf{U}^π and \mathbf{R}^π are the utility and reward functions represented as vectors. What is the meaning of \mathbf{T}^π ? How does this relate Markov decision processes and Markov chains?

Problem 6

What is dynamic programming? Can you give an example? Why is dynamic programming more efficient than brute force methods for solving MDPs?

Problem 7

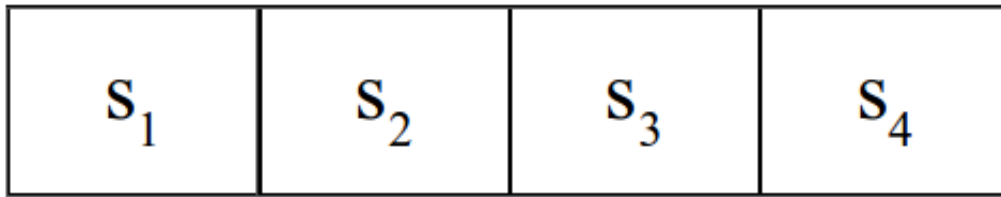
Can you explain what policy iteration and value iteration are? What are their similarities and differences?

Problem 8

What is the difference between open and closed-loop planning?

Problem 9

Consider the simple gridworld shown below. An agent in this world can move to the cell to its immediate left or to the cell to its immediate right, and the transitions are deterministic. Moving left in s_1 gives a reward of 100 and terminates the game. Moving right in s_4 does nothing. Perform value iteration and determine the utility of being in each state assuming a discount factor of 0.9.



(images/simplegrid.png)

Problem 10

How does asynchronous value iteration differ from standard value iteration? What is the importance of the state ordering?

Apply Gauss-Seidel value iteration to the simple gridworld from the previous problem. First, use a state ordering of s_1, s_2, s_3, s_4 . Then use an ordering of s_4, s_3, s_2, s_1 . How many iterations did each ordering take to converge?

Problem 11

In what cases would you prefer to use dynamic programming? Approximate dynamic programming? Online methods?

Model Uncertainty

[↑]

Problem 1

For what types of problems do we use reinforcement learning? What are the two main approaches?

Problem 2

Why is the concept of exploration versus exploitation so important in reinforcement learning?

What is a multi-armed bandit? Describe the various parameters involved in a multi-armed bandit problem. Imagine you have a two-armed bandit and are convinced that one of the levers yields a payout of \$1 with probability 0.9. You have never pulled the other lever, and are unsure if it has any payout. Relate this to the problem of exploration and exploitation.

Problem 3

Suppose we have a two-armed bandit. Our estimate of the payout rate of the first lever is 0.7, and our estimate of the payout rate for the second lever is 0.6. That is, $\rho_1 = 0.7$ and $\rho_2 = 0.6$. Our 95% confidence intervals for θ_1 and θ_2 are (0.6, 0.8) and (0.3, 0.9), respectively.

What is the difference between θ_i and ρ_i ? Suppose you used an ϵ -greedy strategy with $\epsilon = 0.5$. How might you decide what lever to pull? Suppose you used an interval exploration strategy with 95% confidence intervals. What lever would you pull?

Problem 4

What are Q-values and how do they differ from utility values U ? Imagine you have a model of the reward and transition functions. If you were to run a value iteration using the Q-values instead of the utility values U , what would be the update equation?

Problem 5

What is the central equation behind incremental estimation? Identify the temporal difference error and the learning rate. Imagine you have an estimate of some random variable X . Imagine that this estimate is $\hat{x} = 3$. If the learning rate is 0.1, what happens to your estimate after observing a new sample $x = 7$? What happens if the learning rate is 0.5? Comment on the effect that learning rate has on incremental estimation.

Problem 6

What are the similarities and differences between Q-learning and Sarsa?

Problem 7

Use Q-values, the Bellman equation, and the incremental update equation to derive the update equations for Q-learning and Sarsa.

Problem 8

What is the difference between Sarsa and Sarsa(λ)? What types of problems can be solved more efficiently using eligibility traces?

Problem 9

What are the differences between model-based reinforcement learning and model-free reinforcement learning in terms of the quality of the learned policy and computational cost?

State Uncertainty

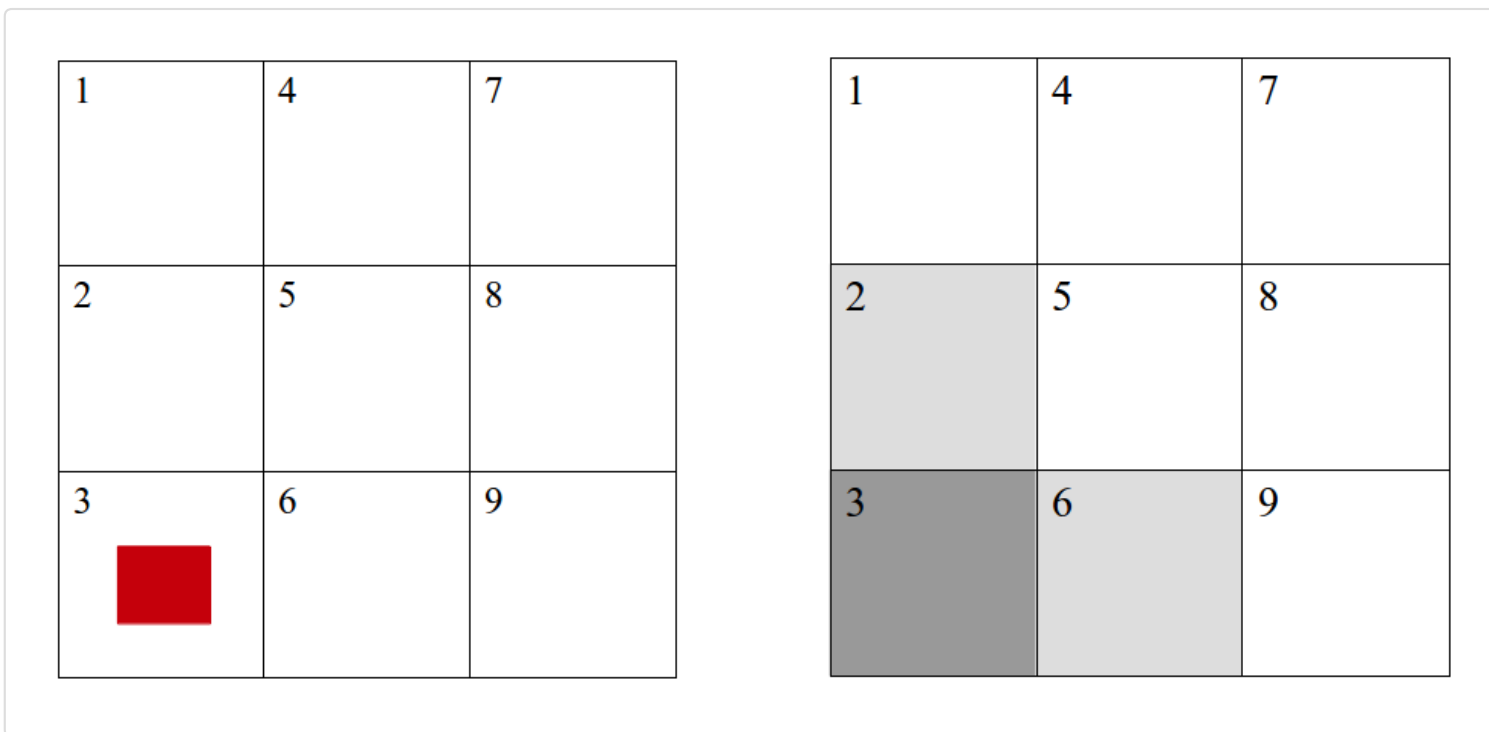
[↑]

Problem 1

What is a POMDP and how does it differ from an MDP? Draw the structure of a POMDP and compare it to that of an MDP.

Problem 2

Examine the two gridworlds shown below. In the left-most gridworld, you know the position of the agent, represented by the red square. In the right-most gridworld, you only have a probability distribution over possible states. How might you represent the "state" for each case? Use this to explain why POMDPs are sometimes called "belief-state MDPs" and are generally intractable.



(images/twogrid.png)

Problem 3

A key to solving POMDPs is the ability to maintain a belief, or probability distribution, over states. What methods can be used to update beliefs? When might one be preferred over the others?

Problem 4

Derive the following equation for a discrete state filter:

$$b'(s') \propto O(o | s', a) \sum_s T(s' | s, a) b(s)$$

from the definition of a belief update $b'(s') = P(s' | o, a, b)$.

Problem 5

Why would you use a particle filter with rejection? Why would you use a particle filter without rejection? Why is it better to use a larger number of particles in your particle filter? What is particle deprivation, and how can you prevent it?

Problem 6

Work through the crying baby example presented in the textbook. Work through the math, updating your belief with the actions and observations given. Verify that your numbers match those in the text.

Problem 7

In MDPs, the policy is a mapping from states to actions. What does a POMDP policy look like? How do you use this policy to find the utility of a belief state?

Problem 8

Imagine you have an exam tomorrow, but there is a non-negligible chance the professor and TAs forgot about the exam. You have a choice: you can study, or you can take the evening off. If you study and there is an exam, you get a reward of 100. If you study and there is no exam, you receive no reward. If you take the evening off and there is no exam, the enjoyment of not studying gives you a reward of 100. But if you take the evening off and there is an exam, the certain F and associated stress give you a reward of -100.

Write down the alpha vectors for this problem. How sure should you be that there will be no exam before you take the evening off? Imagine you have a third option, which is to drop out of school and live in the wilderness. This simple lifestyle would give you a reward of 30, regardless of whether the exam takes place or not. What can you say about this option? Would you ever take it?

Problem 9

Imagine that you have already solved for the policy of a 3-state POMDP, and you have the following alpha vectors:

$$\begin{pmatrix} 300 \\ 100 \\ 0 \end{pmatrix}, \begin{pmatrix} 167 \\ 10 \\ 100 \end{pmatrix}, \begin{pmatrix} 27 \\ 50 \\ 50 \end{pmatrix}$$

The first and third alpha vectors correspond to action 1, and the second alpha vector corresponds to action 2.

Is this even a valid policy? Can you have multiple alpha vectors per action? If the policy is valid, determine the action you would take given you have the following belief: 0% chance in state 1, 70% chance in state 2, 30% chance in state 3.

Problem 10

What does it mean to solve a POMDP offline versus solving it online? What are the advantages and disadvantages of each? How do QMDP, FIB, and point-based value iteration work? What are the advantages and disadvantages of each?

Problem 11

The update equation for QMDP is shown below.

$$\alpha_a^{(k+1)}(s) = R(s, a) + \gamma \sum_{s'} T(s' | s, a) \max_{a'} \alpha_{a'}^{(k)}(s')$$

How many operations does each iteration take? Compare this with the number of operations required per iteration required for FIB, whose update equation is shown below.

$$\alpha_a^{(k+1)}(s) = R(s, a) + \gamma \sum_o \max_{a'} \sum_{s'} O(o | s', a) T(s' | s, a) \alpha_{a'}^{(k)}(s')$$

Write code that applies both QMDP and FIB to the crying baby problem.

Website generated with MDwiki (<http://www.mdwiki.info>) © Timo Dörr and contributors.