

DM828 - Introduction to Artificial Intelligence

Exercise Sheet 7, Autumn 2011 [pdf format]

This is a list of exercises for preparation to the written exam on January 19. (Last updated January 3rd.) Questions on the exercises can be sent to the teacher by email. Solutions of exercises for which there has been a request will be discussed on January 16 in U49E at 10. A list of received questions will be posted here.

Exercises

1. Search

Suppose you have two containers, initially empty. One has a capacity of exactly 3 liters; the other has a capacity of 5 liters. You can pour water from one container to another, empty a container, or fill a container at any time. Your problem is to place exactly 4 liters of water in the 5-liter container. Describe how this problem could be framed as a search problem defining the related components. Solve the problem within this framework and report the search performed.

2. Prove that the set of states expanded by algorithm A^* is a subset of those expanded by breadth-first search.

3. Every day that he leaves work, Albert the Absent-minded Professor, toggles his light switch according to the following protocol: (i) if the light is on, he switches it off with probability 0.80; and (ii) if the light is off, he switches it on with probability 0.30. At no other time (other than the end of each day) is the light switch touched.

a) Suppose that on Monday night, Albert's office is equally likely to be light or dark. What is the probability that his office will be lit all the other four nights of the week (Tuesday through Friday)?

b) Suppose that you observe that his office is lit on both Monday and Friday nights. Compute the expected number of nights, from that Monday through Friday, that his office is lit.

Now suppose that Albert has been working for five years (i.e., assume that the Markov chain is in steady state).

(c) Is his light more likely to be on or off at the end of a given workday?

4. Consider a variant of the game *Nim*. A number of tokens are placed on a table between the two opponents; at each move, the player must divide a pile of tokens into two nonempty piles of different sizes. Thus, 6 tokens may be divided into piles of 5 and 1 or 4 and 2, but not 3 and 3. The first player who can no longer make a move loses the game. Explore exhaustively the search space for a number of tokens equal to 7 and represent the game tree. Who will win the game?

5. Design a spam mail classifier by means of k-nearest neighbor and decision trees and report your design choices. What is a lazy classifier and which of the two methods mentioned is such?

6. Suppose that a fair-looking coin is tossed three times and lands heads each time. Show that a classical maximum likelihood estimate of the probability of landing heads would give 1, implying that all future tosses will land heads. By contrast, show that a Bayesian approach with a prior of 0.5 for the probability of heads would lead to a much less extreme conclusion on the posterior probability of observing heads.

7. Bayesian Networks and Learning

Consider a Naive Bayes problem with three features, x_1, x_2, x_3 . Imagine that we have seen a total of 16 training examples, 8 positive (with $y = 1$) and 8 negative (with $y = 0$). In Table 1 you find some of the counts.

What are the values estimated from the data for the following parameters:

	$y = 0$	$y = 1$
$x_1 = 1$	7	7
$x_2 = 1$	1	1
$x_3 = 1$	3	5

Table 1: Count table

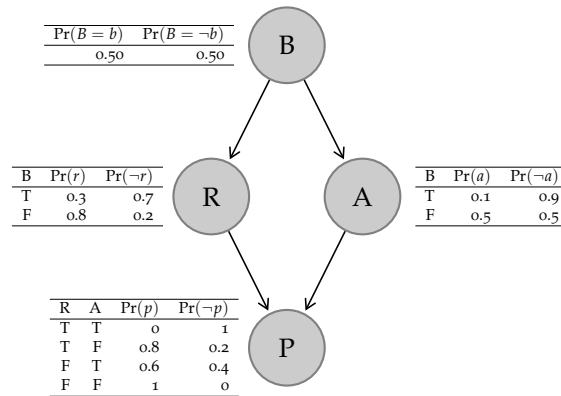


Figure 1: The graphical model of exercise on Bayesian Network Inference. Lower-case letter indicate the outcome that the upper-case letter can take.

- $\Pr(x_1 = 1|y = 0) = \theta_{10}$,
- $\Pr(x_2 = 1|y = 1) = \theta_{21}$,
- $\Pr(x_3 = 0|y = 0) = 1 - \theta_{30}$?

(You need to show the full derivation, answers by intuition without analytical justification do not count.)

8. Bayesian Networks – Inference.

Figure 1 shows a graphical model with conditional probabilities tables about whether or not you will panic at an exam based on whether or not the course was boring (“B”), which was the key factor you used to decide whether or not to attend lectures (“A”) and revise doing the exercises after each lecture (“R”).

You should use the model to make exact *inference* and answer the following queries:

- what is the probability that you will panic or not before the exam given that you attended the lectures and revised after each lecture?
- what is the probability that you will panic or not before the exam?
- Your teacher saw you panicking at the exam and he wants to work out from the model the reason for that. Was it because you did not come to the lecture or because you did not revise?

Describe how stochastic inference methods like Prior-Sample, Rejection-Sampling, Likelihood-weighting and Markov Chain Monte Carlo could be used to answer the queries above.

9. Directed Graphical Models Consider the graph in Figure left.

- Write down the standard factorization for the given graph.
- For what pairs (i, j) does the statement X_i is independent of X_j hold? (Don’t assume any conditioning in this part.)

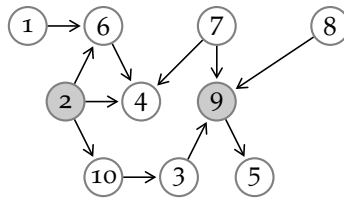


Figure 2: A directed graph.

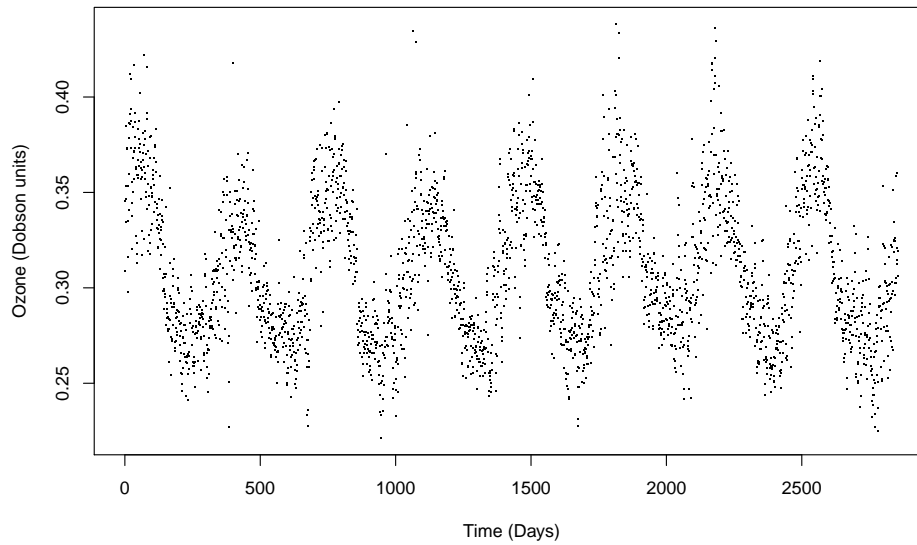


Figure 3: Time series

- Suppose that we condition on $\{X_2, X_9\}$, shown shaded in the graph. What is the largest set A for which the statement X_1 is conditionally independent of X_A given $\{X_2, X_9\}$ holds?
- What is the largest set B for which X_8 is conditionally independent of X_B given $\{X_2, X_9\}$ holds?

10. Neural Networks

Express the output of a neural network with one single hidden layer as a function of the input parameters when the activation function at the units is a linear function. Assume the same linear function at each unit. Would it be possible to simplify the network to a one layer perceptron?

11. A common data analysis task is *time series prediction*, where we have a set of data that show something varying over time, and we want to predict how the data will vary in the future. Examples are stock markets, river levels and house prices.

Suppose we collected the daily measurement of the thickness of the ozone layer above Palmerston North in New Zealand between 1996 and 2004. Ozone thickness is measured in Dobson units, which are 0.01 mm thickness at 0 degree Celsius and 1 atmosphere pressure. The reduction in stratospheric ozone is partly responsible for global warming and the increased incidence of skin cancer. The thickness of the ozone varies naturally over the year as you can see from Figure 3.

Design an application of multi-layer perceptron to predict the ozone levels into the future and report your design choices specifying inputs and outputs for the problem and consequently the input and output nodes for the network.

12. Naive Bayes.

Consider the binary classification problem of spam email in which a binary label $Y \in \{0, 1\}$ is to be predicted from a feature vector $X = (X_1, X_2, \dots, X_n)$, where $X_i = 1$ if the word i is present

in the email and 0 otherwise. Consider a naive Bayes model, in which the components X_i are assumed mutually conditionally independent given the class label Y .

- a Draw a directed graphical model corresponding to the naive Bayes model.
 - b Find a mathematical expression for the posterior class probability $p(Y = 1|x)$, in terms of the prior class probability $p(Y = 1)$ and the class-conditional densities $p(x_i|y)$.
 - c Make now explicit the hyperparameters of the Bernoulli distributions for Y and X_i . Call them, μ and θ_i , respectively. Assume a beta distribution for the prior of these hyperparameters and show how to learn the hyperparameters from a set of training data $\mathbf{d} = (y^j, \vec{x}^j)_{j=1}^m$ using a Bayesian approach. Compare this solution with the one developed in class via maximum likelihood.
13. Consider again a naive Bayes for spam classification. How things would change wrt the previous exercise if the feature vector for a an email j were now $X^j = (X_1^j, X_2^j, \dots, X_{n_j}^j)$, with n_j the number of words in the email and $X_i^j \in \{1, \dots, N\}$ no longer binary but indicating a number that corresponds to the word that appears in the i th position.
 14. **Neural Networks** Make a McCulloch-Pitts neuron or a multilayer network that can calculate the logic function double implication: \Leftrightarrow .
 15. A joint probability table for the binary variables A , B , and C is given below.

A / B	b_1	b_2
a_1	(0.006, 0.054)	(0.048, 0.432)
a_2	(0.014, 0.126)	(0.032, 0.288)

Table 2: Joint probability distribution $P(A, B, C)$

- Calculate $P(B, C)$ and $P(B)$.
- Are A and C independent given B ? (Remember to report the justification of your answer.)