Advanced AI Techniques (WS05/06)

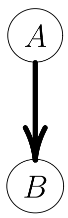
Exercise sheet 13 (Exam Preparation)
Deadline: There is no deadline. It will not be discussed.

Exercise 1 (Section Lars Schmidt-Thieme (15 credit points))

a) (2 credits) Describe two ways for checking d-separation.

b) (2 credits) Give a simple example that demonstrate a d-separation (i.e. 2 vertices being d-separated by another vertex). Explain your example.

c) (6 credits) Given the following bayesian network structure.



Assume the following data on two binary variables A and B is observed (dots indicate missing values).

A	*B*
0	*0*
0	*1*
0	*1*
1	*0*
1	*1*
.	*0*

Setup an EM learning scheme to learn the parameters of this network. Conduct the first two iterations of the algorithm (including the third M-Step).

d) (2 credits) Give a criterion that allows to check two Bayesian networks for Markov equivalence.

e) (3 credits) What are PC algorithms used for?

Exercise 2 (Section Luc de Raedt (15 credit points))

Mary wants to conduct a statistical experiment tossing an unbiased coin several times. Indeed, she has a coin which comes up head or tail with a probability of 0.5 each. However, after tossing the coin, she might accidentally (and without noticing it) lose it and instead pick up a bent coin also lying on the ground, which she will use in the remaining tosses and which comes up head with a probability of 0.2 and tail with a probability of 0.8. The probability of losing her coin and picking up the other after a toss is 0.3 .

1. *Draw a HMM for this problem with states {Normal,Bent}, possible observation symbols {Head,Tail} and corresponding parameters. (5 points)*
2. *Suppose Mary tosses three times, observing (Head,Tail,Tail). Go through the Viterbi algorithm for this observation sequence, by filling out the table below. What is the most likely state sequence? (10 points)*

	t=1	*t=2*	*t=3*
$\delta_{Normal}(t)$			
$\delta_{Bent}(t)$			
$\psi_{Normal}(t)$			
$\psi_{Bent}(t)$			
$\hat{X}_t \in \{N, B\}$			

Reminder:

$$\delta_S(t) = \max_{x_1,...,x_{t-1}} P(x_1, ..., x_{t-1}, o_1, ..., o_{t-1}, x_t = S, o_t)$$

where $o_1, ..., o_t$ is a partial observation sequence and x_t the state at time t.
The most likely state sequence is reconstructed working backwards:

$$\hat{X}_T = arg \max_S \delta_S(T)$$

$$\hat{X}_t = \psi_{\hat{X}_{t+1}}(t+1) \qquad (t = 1, ..., T-1)$$

Exercise 3 (Section Wolfram Burgard (15 credit points))

a) *(2 credits) What is the idea of reinforcement learning, and when does it make sense?*

b) *(2 credits) What is the difference between exploitation moves and exploration moves?*

c) *(3 credits) Which problem can arise if only exploitation moves are carried out? Describe briefly a situation in which it occurs.*

d) *(2 credits) There are different action selection methods: greedy, ϵ-greedy, and the Gibbs method given by the formula*

$$\frac{e^{Q(a)/\tau}}{\Sigma_a e^{Q(a)/\tau}}$$

Which method(s) would you implement to solve this problem and why?

e) *(2 credits) To achieve faster learning rates, how should the values ϵ and τ be adjusted?*

- *ϵ should be: ☐ lower ☐ higher*
- *τ should be: ☐ lower ☐ higher*

f) *(4 credits) In a ternary bandit problem, there are three possible choices for actions. Let the expected rewards be as given in the following table:*

action	*(deterministic) reward*
A	60
B	40
C	20

Assume you start with an initialization value of 0 for each action. Once a stable strategy is learned, what is the expected average reward per move

- *with greedy action-selection?*
- *with ϵ-greedy action-selection, $\epsilon = 0.1$?*
- *with ϵ-greedy action-selection, $\epsilon = 0.2$?*
- *with Gibbs method?*

g) (+2 credits) *The average sum evaluation formula can be improved by assigning an initial value s to each possible action (as if, as a kind of hypothetical prior knowledge, each action was chosen once with a reward of s).*
What is a sufficient condition for such an initialization so that even with the greedy action selection method, the best strategy is learned ? (No proof required, just motivate your answer.)

Exercise 4 (Section Bernhard Nebel (15 credit points))

a. Shortly define the following terms (3 credits):

- *action profile*
- *Nash equilibrium*
- *Iterated elimination of weakly dominated strategies.*

b. Which implications hold between the following statements? (3 credits)
(give your answers in the form, e.g.: (iv.) => (v.), (v.) <=> (vi.), (iv.) => (vi.), (vi.) and (vii.) independent)

(i.) Action profile π is a unique solution from iterated elimination of weakly dominated strategies

(ii.) Action profile π is a unique solution from iterated elimination of strongly dominated strategies

(iii.) Action profile π is a Nash equilibrium

c. Can you give an example for a Nash equilibrium that is not a unique solution from iterated elimination of strongly dominated strategies? (4 credits)

d. Let G be a strategic 2-player zero-sum game. Let -G be the game that has the same players and actions, and is obtained by changing each payoff from v to -v. Does hold: If G has a Nash equilibrium, then -G has a Nash equilibrium. (Find a proof or a counterexample.)
(5 credits)