Delft University of Technology
Electrical Engineering, Mathematics and Computer Science

Exam
CS4375 – Artificial Intelligence Techniques

October 14, 2022, 9:00 – 12:00

- The exam consists of multiple choice questions for a total of 17 points and open questions for a total of 18 points.
- With respect to the multiple choice questions:
  - You can only select one answer.
  - In certain cases multiple options could be considered correct, in these cases you need to **select the *best* (i.e., most specific) answer**.
- With respect to the open questions:
  - Answer in correct Engels and write legible; first use scratch paper if needed.
  - Motivate your answers.
  - Be to-the-point. Don't provide irrelevant information, this can lead to deductions.
- It is not allowed to use the book.
- It is allowed to use a calculator.
- You are allowed to bring one page of A4 paper with handwritten notes in your own handwriting.
- Before handing in, make sure that you have written your name and study number on each page. Indicate the total number of pages on the first page.
- Total number of pages of this exam: 5.

Succes!

Multiple Choice Questions

Uncertainty

1. (1 point) What is the *qualification problem*?
   A. The problem of finding the structure of a probabilistic domain.
   B. Not having a complete theory to model a domain.
   C. Difficulty of weighting the relative importance of various goals in rational decision making.
   D. Having to specify all exceptions of desired action effects.

Bayesian networks and Inference

2. (1 point) What are the two main operations needed to implement variable elimination?
   A. Max-out, sum-out
   B. Max-out, pointwise product
   C. Sum-out, pointwise product
   D. Pointwise product, prior sample

3. (1 point) Which is not a method for approximate inference:
   A. Variable elimination
   B. Particle filter
   C. Rejection sampling
   D. Likelihood weighting

Making 'Simple' Decisions

4. (1 point) Tickets to a lottery cost $10. You buy one. There are three possible prizes: a $100 payoff with probability 1/50, a $10,000 payoff with probability 1/5000 and a $10,000,000 payoff with probability 1/2,000,000. What is the expected monetary value of a lottery ticket?
   A. $9
   B. $10
   C. $7
   D. $-1

Answer: $EMV = 100 \times 1/50 + 10000 \times 1/5000 + 10,000,000 \times 1/2,000,000 = 2 + 2 + 5 = 9$

5. (1 point) Economists often make use of an exponential utility function for money: $U(x) = -e^{-x/R}$, where R is a positive constant representing an individual's risk tolerance. Risk tolerance reflects how likely an individual is to accept a lottery with a particular expected monetary value (EMV) versus some certain payoff. As R (which is measured in the same units as x) becomes larger, the individual becomes less risk-averse. Consider the choice between receiving $100 with certainty (probability 1) or participating in a lottery which has a 50% probability of winning $500 and a 50% probability of winning nothing. Which of the following equations holds true in an exponential utility function that would cause an individual to be indifferent to these two alternatives.
   A. $100 = 0.5e^{-500/R} + 0.5$
   B. $e^{-100/R} = 0.5e^{-500/R}$
   C. $100 = 0.5e^{-500/R}$
   D. $e^{-100/R} = 0.5e^{-500/R} + 0.5$

6. (1 point) You are participating in a game show with 3 doors A,B and C. Two of them have goats behind them and 1 door hides a car. Initially you are allowed to choose one door, say door A. After you chose, the host Monty inspects the two doors you did not choose and opens one with a goat, say door B. (Note that the host would never open the door with the car, so in this example if the car were in door C, the host would be guaranteed to open B). Given the information that door B has a goat what is the probability that door C has the car?

   A. 1/3

   B. 1/2

   C. 2/3

   D. 1

Answer: The events $\mathcal{A}$ = Door A has a car, and similarly $\mathcal{B}$ and $\mathcal{C}$. $P(\mathcal{A}) = P(\mathcal{B}) = P(\mathcal{C}) = 1/3$. The events $\mathcal{H}$ =Host opens door B. To find $P(\mathcal{C}|\mathcal{H})$.

$$P(\mathcal{C}|\mathcal{H}) = \frac{P(\mathcal{H}|\mathcal{C})P(\mathcal{C})}{P(\mathcal{H}|\mathcal{A})P(\mathcal{A}) + P(\mathcal{H}|\mathcal{B})P(\mathcal{B}) + P(\mathcal{H}|\mathcal{C})P(\mathcal{C})}$$

The host always opens the door containing the goat. If the car were in door A, the host will arbitrarily chose between door B and C to open, so $P(\mathcal{H}|\mathcal{A}) = 1/2$. If the car were in door B, the host would not have opened B, so $P(\mathcal{H}|\mathcal{B}) = 0$. If the car were in door C, the host would be guaranteed to open B, so $P(\mathcal{H}|\mathcal{C}) = 1$. Finally,

$$P(\mathcal{C}|\mathcal{H}) = \frac{1 \times 1/3}{1/2 \times 1/3 + 0 \times 1/3 + 1 \times 1/3} = \frac{2}{3}$$

Time and Uncertainty

7. (1 point) What algorithm is best to compute the solution to $P(X_{t+k}|e_{1:t})$

   A. Forward without addition of new evidence

   B. Forward-backward

   C. Backward

   D. Variable Elimination

   Answer: B is wrong, C is nonsense. D is actually not far off: if applied "left to right" VE can be seen to correspond to 'forward', however, the more specific algorithm (A) is the best answer.

Making Sequential Decisions

8. (1 point) In infinite horizon MDPs under discounted rewards, what condition on discount factor γ needs to be satisfied for the Value Iteration algorithm to be able to guarantee convergence after an infinite number of iterations?

   A. $\gamma \geq 1$.

   B. $\gamma > 1$.

   C. $\gamma \leq 1$.

   D. $\gamma < 1$ **.**

Learning

9. (1 point) Which class of methods is expected to perform best in a supervised learning problem with a small data sample?
   - **A. Bayesian methods.**
   - B. Maximum likelihood methods.
   - C. Machine learning methods.
   - D. Clustering methods.

Quantification of objectives and Reward hacking

10. (1 point) Consider the scenario where an AI-based vacuum cleaner ejects collected dust and collects it again. What is most likely the reason for such behavior? The designer of the system did not consider environmental variables regarding the size and dimensions of the room that the vacuum cleaner needs to clean.
    - A. The algorithms that detect dust were designed as a classification problem, but they should have been a regression problem.
    - **B. The vacuum cleaner is "gaming" the designed objective function by ejecting the dust so it can collect it again and get more reward.**
    - C. The vacuum cleaner is learning from a biased dataset.
    - D. The vacuum cleaner is taking random actions, since no objective function was specified by the system's designer.

MDPs & POMDPs

11. (1 point) The "reward hypothesis" states that:
    - A. The cumulative reward of agents should be aligned with ideal task behavior.
    - **B. All tasks can be formulated as maximization of expected cumulative scalar reward signal.**
    - C. In case where both formulations are possible, rewards are more preferrable to penalties in reinforcement learning.
    - D. All of the above.

Model-Free Reinforcement Learning

12. (1 point) What statement about Monte Carlo estimation ("direct utility estimation") is wrong:
    - A. Monte Carlo estimation can be used to do policy evaluation.
    - B. Monte Carlo estimation is unbiased.
    - **C. Monte Carlo estimation needs only a small number of samples.**
    - D. Monte Carlo estimation does not exploit any knowledge of the Bellman equation

    Answer: Monte Carlo estimation is not biased, but has a high variance which means that one typically needs a large number of samples.

Model-based RL

13. (1 point) Which of the following statement most accurately describes the relation between model-based and model-free reinforcement learning techniques?
    - A. Model-free and model-based RL are mutually exclusive.
    - B. Model-free and model-based RL are complementary.
    - **C. Model-free RL techniques can be used within model-based RL.**
    - D. Model-based RL can be used to bootstrap model-free RL.

Answer: Model-free techniques can indeed be used (as planning techniques) within model-based RL.

Multiagent decision making

14. (1 point) A multiagent MDP (MMDP) is difficult to solve because
    A. It requires coordination.
    B. The puppeteer agent needs to communicate with all the agents constantly.
    C. The number of joint actions is exponential in the number of agents.
    D. None of the other options; an MMDP is easy to solve in polynomial time.

15. (1 point) In multi-agent problems it might not always be feasible to act on global information instead of only local observations. Which of the following is **not** a possible reason for this?
    A. Communication between agents might not be possible.
    B. It requires coordination.
    C. Scales poorly with the number of agents.
    D. Communication is not instantaneous or noise free.

Adversarial Search

16. (1 point) From the tree in Figure 1, what is the minimax value of the top node of this tree? (Terminal nodes are labeled with the utility of player MAX.)
    A. 18
    B. 3
    C. 4
    D. 14

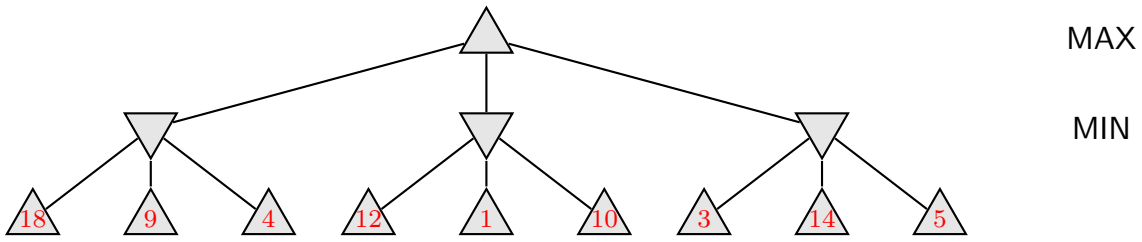

Figure 1: Two-player zero-sum game.

Game Theory

17. (1 point) What is the name of the auction in which the seller lowers the price until the item is sold (descending auction)?
    A. First-price open cry.
    B. First-price sealed bid.
    C. Dutch auction.
    D. Second-price sealed bid.

18. (1 point) Which of the following statements is *false*?
    A. A game can have more than one Nash equilibrium.
    B. Every game has at least one Pareto-optimal solution.
    C. A game can have more than one Pareto-optimal solution.
    D. Every game has at least one pure-strategy Nash equilibrium.

Advanced RL Topics

19. (1 point) In the Maximum Entropy Inverse Reinforcement Learning (IRL) algorithm, it is **FALSE** that:
    A. The algorithm employs the principle of maximum entropy to resolve the ambiguity problem in a principled manner.
    B. The algorithm requires that the human expert provides an optimal policy.
    C. If the human expert demonstrated erratic behavior in a certain condition, the learned reward that models this behavior should be low or zero.
    D. If the human expert demonstrated consistent behavior in a certain condition, the learned reward that models this behavior should be large.

Open Questions

20. (5 points) Solving an MDP.

    Consider the following MDP:

    - States: set S of states has 9 states in a 3 x 3 grid: <1, 1>, ... <3, 3>, where
      - <1,1> is start state, and
      - <2, 2> and <3, 3> are terminal states.
      - lower coordinates correspond to top-left corner of the grid.
    - Actions A = {up, right, left, down}
    - Reward model:
      - R(s = <2, 2>) = + 1
      - R(s = <3, 3>) = -1
      - R(s) = -0.2 for any non-terminal state s
      - Note that these rewards are given at the beginning of each time step (so for being in a state, not for transitioning to a state). In other words, these are $R(s)$ not $R(s')$.
    - Transition model:
      - where every action with probability 0.9 ends in the tile that is two tiles from the starting tile, and with probability 0.1 ends in the next tile. When bumping 'into the wall' (side of the grid) the agent will remain in its place.
      - That is, with 90% probability the agent will move 2 squares, but only if this is possible.
    - the discount factor $\gamma = 0.5$

    Compute the value of state $\langle 1, 1 \rangle$ after 2 iterations of value iteration.

    You can assume that the value of each state is updated in parallel (using the values of the previous iteration), such that the order of doing the sweeps over states does not matter.

End of exam.
Please make sure you answerred all 22 questions.

Answer:

- We start with initializing: $V(1,1) = \cdots = V(3,3) = 0$.
- Iteration 1, we apply

$$F(s,a) \leftarrow \sum_{s'} \Pr(s'|s,a)V(s')$$

$$V(s) \leftarrow R(s) + \gamma \max_a F(s,a)$$

for all states, and their possible successors

$$F(\langle 1,1\rangle, Up) = 1 \cdot V(1,1) = 0$$

$$F(\langle 1,1\rangle, Left) = 1 \cdot V(1,1) = 0$$

$$F(\langle 1,1\rangle, Down) = 0.9 \cdot V(1,3) + 0.1 \cdot V(1,2) = 0$$

$$F(\langle 1,1\rangle, Right) = 0.9 \cdot V(3,1) + 0.1 \cdot V(2,1) = 0$$

$$V(1,1) \leftarrow -0.2 + \gamma \max\{0,0,0,0\} = -0.2$$

idem for all other non-terminal states. For the terminal states, we get:

$$V(2,2) = +1$$

$$V(3,3) = -1$$

- Iteration 2, we only compute for $\langle 1,1\rangle$ since that suffices to get the answer:

$$F(\langle 1,1\rangle, Up) = 1 \cdot V(1,1) = -0.2$$

$$F(\langle 1,1\rangle, Left) = 1 \cdot V(1,1) = -0.2$$

$$F(\langle 1,1\rangle, Down) = 0.9 \cdot V(1,3) + 0.1 \cdot V(1,2) = 0.9 \cdot -0.2 + 0.1 \cdot -0.2 = -0.2$$

$$F(\langle 1,1\rangle, Right) = 0.9 \cdot V(3,1) + 0.1 \cdot V(2,1) = -0.2$$

$$V(1,1) \leftarrow -0.2 + \gamma \max\{-0.2, -0.2, -0.2, -0.2\} = -0.3$$

21. Modeling an MDP.
Your trainer says you are a talent in the sport Artificial Bowling that will be an Olympic sports in the next Olympic Games. Your trainer tells you that you should train a lot and eat regularly. As the trainer sees it, you preferably should limit your actions to the above. However, you don't have a sponsor and therefore, you have to work to pay for your food.
The action of working, brings a reward of -1, eating a reward of +5. Both the working and the eating action have a probability of 0.5 of making you drop a level, and 0.5 of staying at your current level of expertise.
A training session with a probability of 0.25 brings you to the next level of expertise in Artificial Bowling, with probability 0.5 you stay at this level, and with probability 0.25 you drop a level. There are four increasing levels of expertise: beginner, regional top, national top, world top. You cannot drop below the level of beginner. Once you are at the world top, training keeps you at that level with a probability of 0.75. Only at that level can you try for the Olympic Gold Medal, this action has a 0.01 probability of succeeding, which brings to you a terminal state with a reward of +100. With a probability of 0.99 you drop one level with a reward of -10. For any level of expertise, training gives a negative reward of -5.

(a) (6 points) Formalize this as an MDP, specify the states, actions and the reward and transition functions.

Answer:

```
States: S = { b, r, n, w, gold}. [1 point]

Actions A = {train, work, eat, olymp} [1 point]

Reward model R : S x A x S -> Integer, specified by the following,
for any s and s' from S\{gold}:
R(s, train, s') = -5
R(w, olymp, gold) = 100
R(w, olymp, n) = -10
R(s, work, s') = -1
R(s, eat, s') = 5
[2 points]

Transition model T: S x A x S -> [0, 1], specified by the following,
for any s from S\{gold}, and any action a from {eat, work}:
T(r, train, r) = T(n, train, n) = 0.5
T(b, train, b) = T(w, train, w) = 0.75
T(s, train, s-1) = 0.25, where  s - 1 refers to one level below s if that exists
T(s, train, s+1) = 0.25, where  s + 1 refers to one level above s if that exists
T(w, olymp, gold) = 0.01
T(w, olymp, n) = 0.99
T(r, a, r) = T(n, a, n) = T(w, a, w) = 0.5
T(b, a, b) = 1
T(s, a, s-1) = 0.5, where  s - 1 refers to one level below s if that exists
[2 points]
```

22. The following table represents a game of "rock, paper, scissors" played between 2 players Alice and Bob:

		Bob		
		R	P	S
$Alice$	R	$0,0$	$-1,1$	$1,-1$
	P	$1,-1$	$0,0$	$-1,1$
	S	$-1,1$	$1,-1$	$0,0$

A win gives 1 point and a lose results in -1 point. A draw gives 0 points to both players. Both players follow the mixed strategy of taking actions uniform randomly.

(a) (2 points) Is both players following this strategy an equilibrium? Explain your answer.

Answer: Yes, it's an equilibrium. We will write μ_i to denote a mixed strategy: a probability distribution over strategies s_i. The critical point is that at a mixed Nash equilibrium $\langle \mu^*_{Alice}, \mu^*_{Bob} \rangle$, both players need to be indifferent between the strategies that are played with positive probability. So, at equilibrium Alice to make Bob indifferent between R,P, and S. This means that μ^*_{Alice} needs to specify $\frac{1}{3}$ for each of her actions, leading to the following expected utilities $u_{Bob}(s_{Bob}) = \frac{1}{3}u_{Bob}(R, s_{bob}) + \frac{1}{3}u_{Bob}(P, s_{bob}) + \frac{1}{3}u_{Bob}(S, s_{bob})$ for Bob's strategies s_{Bob}:

	Bob		
	R	P	S
μ^*_{Alice}	0	0	0

By analogue reasoning, Bob also needs to play uniformly random. [Note, to be 100% complete, one would need to show that this is the only way to make Bob indifferent. For instance, in general in games with 3

strategies, there could be ways that we would make Bob indifferent between just 2 actions, which he would both prefer over the 3rd action. This is not the case in rock, paper, scissors, and mentioning this is not required for getting full points for the question.]

(b) (1 point) Is there a pure equilibrium strategy in this game? Explain your answer.

Answer: No, for each of the pure strategies $\langle s_{Alice}, s_{Bob} \rangle$ (i.e., for all entries in the payoff matrix) at least one of the players will want to move to a better strategy.

(c) (1 point) What is the expected value of this game for Alice?

Answer: Given that both players play uniformly random, there are 9 equally probable outcomes. 3 of them yield 1 point, another 3 yield -1 and remaining 3 yield 0 points. Therefore, expected value is

$$u(\langle \mu^*_{Alice}, \mu^*_{Bob} \rangle) = \frac{1}{9} \times 3 \times 1 \; + \; \frac{1}{9} \times 3 \times -1 \; + \; \frac{1}{9} \times 3 \times 0 = 0.$$

(d) (1 point) Does this game favor any particular player? Explain.

Answer: No. The expected value of the game for both players is 0, so neither Alice or Bob has an advantage.