Name: Matriculation Number:

Retake Exam
Künstliche Intelligenz 2

Feb. 19., 2018

You have 90 min(sharp) for the test;
Write the solutions to the sheet.

The estimated time for solving this exam is 67 minutes, leaving you 23 minutes for revising your exam.

You can reach 134 points if you solve all problems. You will only need 125 points for a perfect score, i.e. 9 points are bonus points.

Different problems test different skills and knowledge, so do not get stuck on one problem.

	To be used for grading, do not write here													
prob.	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	4.1	4.2	4.3	4.4	Sum	grade
total	14	16	8	8	20	8	6	12	8	6	8	20	134	
reached														

The "solutions" to the exam/assignment problems in this document are supplied to give students a starting point for answering questions. While we are striving for helpful "solutions", they can be incomplete and can even contain errors.

If you find "solutions" you do not understand or you find incorrect, discuss this on the course forum and/or with your TA and/notify the instructors.

In any case, grading student's answers is not a process of simply "comparing with the reference solution".

In the course Artificial Intelligence I/II we award 5 bonus points for the first student who reports a factual error (please report spelling/formatting errors as well) in an assignment or old exam and 10 bonus points for an alternative solution (formatted in LaTeX) that is usefully different from the existing ones.

1 Bayesian Reasoning

Problem 1.1 (AFT Tests)

Trisomy 21 (*Down syndrome*) is a genetic anomaly that can be diagnosed during pregnancy using an amniotic fluid test. 14 pt

7 min

The probability of a fetus having Down syndrome is strongly correllated with the age of the mother during pregnancy. For 25 year old mothers the probability is 0.08%, for 43 year old mothers it increases to 2% (we only consider those two age groups).

However, diagnostic tests are never perfect. We distinguish two kinds of errors:

- **Type I Error (False Positive)**: The test result is positive even though the child is healthy.
- **Type II Error (False Negative)**: The test result is negative even though the child has trisomy 21.

The probabilities of Type I and Type II Errors are both merely 1% for amniotic fluid tests for Down syndrome.

1. Express all the numbers given above as conditional probabilities. Use the random variable F with Domain $\{Age_{25}, Age_{43}\}$ for the age of a mother and the Boolean random variables Pos and $Down$ for the propositions "*The amniotic fluid test is positive*" and "*The child has Down syndrome*" respectively.
2. Assume now we have a fixed mother of age 25 (i.e. for any event X you may assume $P(X \mid F = Age_{25}) = P(X)$). How can we compute the probability that a child has Down syndrome, given that the amniotic fluid test is positive? Give an equation that only uses the probabilities for which we have actual numbers.

Solution:

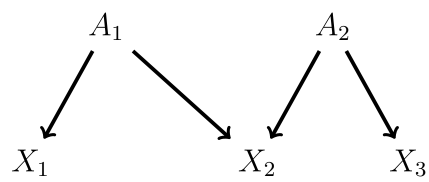
1. $P(Down \mid F = Age_{25}) = 0.0008$, $P(Down \mid F = Age_{43}) = 0.02$, $P(Pos \mid \neg Down) = 0.01$, $P(\neg Pos \mid Down) = 0.01$.
2. We normalize to $F = Age_{25}$ and compute:

$$P(Down \mid Pos) = \frac{P(Pos \mid Down) \cdot P(Down)}{P(Pos)} = \frac{P(Pos \mid Down) \cdot P(Down)}{P(Pos \land Down) + P(Pos \land \neg Down)}$$
$$= \frac{P(Pos \mid Down) \cdot P(Down)}{P(Pos \mid Down) \cdot P(Down) + P(Pos \mid \neg Down \cdot P(\neg Down))}$$
$$= \frac{(1 - P(\neg Pos \mid Down)) \cdot P(Down)}{(1 - P(\neg Pos \mid Down) \cdot P(Down)) + P(Pos \mid \neg Down) \cdot (1 - P(Down))}$$

Problem 1.2 (Bayesian Networks)

Consider the following Bayesian network with boolean variables: 16 pt

8 min



1. Which nodes in the network are
   - Stochastically independent
   - Conditionally independent and under which conditions?
2. What exactly (formal criterion) does an arrow between two nodes in a bayesian network mean for the associated events?

Solution:

1. - A_1 and A_2 are stochastically independent.
   - X_1 and X_3 are stoachastically independent.
   - A_1 and X_3 (respectively A_2 and X_1) are stochastically independent.
   - X_1 and X_2 are conditionally independent given A_1.
   - X_2 and X_3 are conditionally independent given A_2.
2. We draw an arrow from X_j to X_i if X_j is in the smallest set $\texttt{Parents}(X_i)$ with the property $P(X_i|X_{i-1},...,X_1) = P(X_i|\texttt{Parents}(X_i))$

---

2 Decision Theory

Problem 2.1 (Decision Preferences)

8 pt

4 min

1. Name and state three of the axioms for preferences (i.e. $\succ$).
2. How are preferences related to value functions?

Solution:

1. Pick any three of the following:

   Orderability: $A \succ B \vee B \succ A \vee A \sim B$

   Transitivity: $A \succ B \wedge B \succ C \Rightarrow A \succ C$

   Continuity: $A \succ B \succ C \Rightarrow (\exists p.[p, A; (1-p), C] \sim B)$

   Substitutability: $A \sim B \Rightarrow [p, A; (1-p), C] \sim [p, B; (1-p), C]$

   Monotonicity: $A \succ B \Rightarrow (p>q) \Leftrightarrow [p, A; (1-p), B] \succeq [q, A; (1-q), B]$

   Decomposability: $[p, A; (1-p), [q, B; (1-q), C]] \sim [p, A; (1-p)q, B; (1-p)(1-q), C]$

2. Ramsey's theorem states that given a set of preferences that obey the constraints above, there is a value function U with

$$(U(A) \geq U(B)) \Leftrightarrow A \succeq B \quad \text{and} \quad U([p_1, S_1; \ldots; p_n, S_n]) = \sum_i p_i U(S_i)$$

Problem 2.2 (Expected Utility)

What is the formal(!) definition of *expected utility*? Explain every variable in the defining equation. 8 pt

4 min

Solution:The expected utility EU is defined as

$$EU(a|e) = \sum_{s'} P(R(a) = s'|a, e) \cdot U(s')$$

where

1. a is the action for which we want to find out the expected utility, given the evidene e.
2. $U(s')$ is the utility of a state s'.
3. $R(a)$ is the result of the action a.

Problem 2.3 (Decision Network)

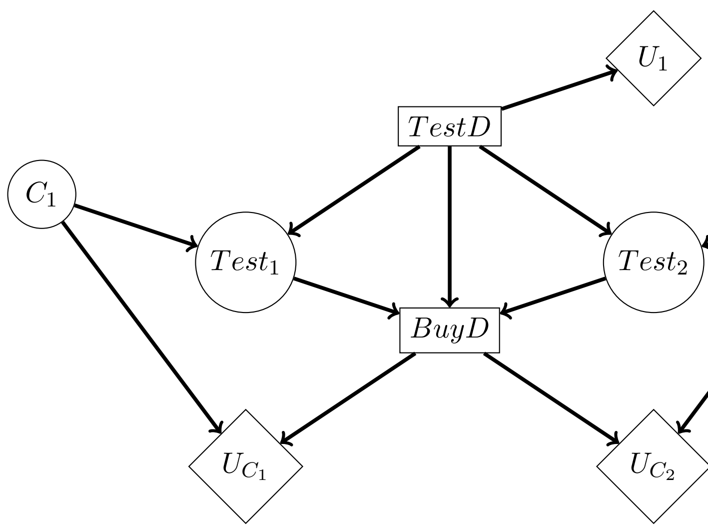
You need a new car. Your local dealership has two models on offer – C_1 for 1500\$ and C_2 for 1250\$. Either car can be of good quality or bad quality. If C_1 is of bad quality, repairing it will cost 200\$, if C_2 is of bad quality repairing it will cost 500\$. You have the choice between two tests: 20 pt 10 min

1. You can either take C_1 on a short test drive for 30 minutes, which will confirm that it is of good quality with certainty 75% probability, and that it is of bad quality (if it is) with certainty 65%.
2. Or you can borrow C_2 for a whole weekend, which will confirm that it is of good quality with 80% certainty, and that it is of bad quality (if it is) with certainty 70%.

The a priori probability that C_1 is of good quality is 70% and for C_2 it is 65%.

1. Draw the decision network for which test to apply and which car to buy in either case. (assume a meaningful utility function given).
2. Explain *formally* how to compute which test to apply.
3. Assume you apply test 1 and it seems like the car is of bad quality. Explain how to compute which car you should actually buy.

Solution:See also `http://mas.cs.umass.edu/classes/cs683/lectures-2010/Lec21_Uncertainty6-F2010-4up.pdf`

1. 

---

3 Markov Models

Problem 3.1 (Bellman Equation)

State the Bellman Equation and explain 1) every symbol in the equation and 2) what the equation is used for and how. 8 pt

4 min

Solution:

$$U(s) = R(s) + \gamma \cdot \max_a \left(\sum_{s'} U(s') \cdot T(s, a, s') \right)$$

Problem 3.2 (Stationary)

Define what it means for a Markov model to be *stationary*, and why we are interested in stationarity. 6 pt

3 min

Solution:A Markov process is called stationary if the transition model is independent of time, i.e. $\mathbf{P}(\mathbf{X}_t \mid \mathbf{X}_{t-1})$ is the same for all t.

We like stationary Markov processs, since they are finite.

Problem 3.3 (Stock Market Predictions)

You bought SpaceY stock recently and try to predict whether to buy more or sell. The stock market is in one of two possible states; bull state or bear state. In a bull state, it will (in the long term) be advantageous to buy stock; in a bear state it will be more advantageous to sell. 12 pt 6 min

If the market is in a bull state, the probability it will still be in a bull state tomorrow is 60%. If it is in a bear state, the probability it will remain so tomorrow is 80%.

If the market is in a bull state, the probability that your stock will rise that day is 90%. If it is in a bear state, your stock will more likely fall (with 60% probability).

1. Explain what kind of probabilities *prediction*, *filtering* and *smoothing* compute in this scenario.
2. Give the underlying equations for the first two of these algorithms and explain what each variable in the equation represents.

Solution:

1. **Prediction** Given the *behavior* of the stock market up to time t_0, compute the probability of the state the stock market will be in at time $t_1 > t_0$

   Filtering Given the behavior of the stock market up to now, compute the probability of the state the stock market is in right now

   Smoothing Given the behavior of the stock market up to t_0, compute the probability that sta stock market was in some state at an earlier point $t_1 < t_0$.

2. $$P(X_{t+1}|e_{1:t+1}) = \alpha P(e_{t+1}|X_{t+1}) \cdot \sum_{x_t} P(X_{t+1}|x_t, e_{1:t}) \cdot P(x_t|e_{1:t})$$

   Where X represents the state and e the behavior of the stock market.

---

4 Learning

Problem 4.1 (Gradient Descent)
Explain a gradient descent algorithm. 8 pt

4 min

Problem 4.2 (Information Entropy)

Explain and define *information entropy*. 6 pt

3 min

Solution:Information entropy of a (set of) random variable is the average level of "information", "surprise", or "uncertainty" inherent in the variable's possible outcomes.

$$I(\langle P_1, ..., P_n \rangle) = \sum_{i=1}^{n} -P_i \log_2(P_i)$$

Problem 4.3 (Overfitting)

Explain what *overfitting* means and why we want to avoid it. 8 pt

4 min

Solution:Overfitting is a modeling error that occurs when a function is too closely fit to a limited set of data points. Overfitting the model generally takes the form of making an overly complex model to explain idiosyncrasies in the data under study rather than the underlying mechanisms.

Problem 4.4 (Tennis Trees)

Consider the following decisions on whether or not to go play tennis. The target is "PlayTennis". 20 pt

10 min

Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Strong	No
Rain	Mild	Normal	Weak	Yes

Explain how to apply decision tree learning to this table. In particular, define the notion of *information entropy*.

Solution:http://courses.cs.tamu.edu/choe/17spring/633/lectures/slide05.pdf

$$E := I(\langle\frac{3}{5},\frac{2}{5}\rangle) = -\frac{3}{5}\log_2(\frac{3}{5}) - \frac{2}{5}\log_2(\frac{2}{5})$$

$$\texttt{Gain}(\texttt{Outlook}) = E - \frac{1}{5}I(\langle 0,1\rangle) - \frac{1}{5}I(\langle 1,0\rangle) - \frac{3}{3}I(\langle\frac{2}{3},\frac{1}{3}\rangle)$$

$$\texttt{Gain}(\texttt{Temperature}) = \ldots$$

Pick whichever attribute has the highest information gain, split there and build subtrees, iterate...
