Lunds Tekniska Högskola
Institutionen för datavetenskap

EDAF70 (a.k.a. EDA132)
VT 2019

Tillämpad Artificiell Intelligens
Applied Artificial Intelligence
Tentamen 2019–03–20, 08.00–13.00, MA10

You can give your answers in English or Swedish.
You are welcome to use a combination of figures and text in your answers.
96 points, 50% needed for pass.

1 Probabilistic reasoning, BNs (EAT) 3+3+3+3+3=15p

You have a set of 5 random variables. You know the following about them:

- Semantically speaking, the phenomenon represented by C is known to cause an effect on what is represented by E, which can then have an effect on what is represented by B and D respectively.
- $\mathbb{P}(B|A,C,D,E) = \mathbb{P}(B|E)$ and $\mathbb{P}(D|A,B,C,E) = \mathbb{P}(B|E)$
- $\mathbb{P}(A|B,C,D,E) = \mathbb{P}(A)$
- $\mathbb{P}(B,D) \neq \mathbb{P}(B)\mathbb{P}(D)$

Answer the following questions (motivate your answers!):

a) When are two random variables independent of each other? When are they conditionally independent?

b) Which of the network(s) i), ii), and iii) (see Fig. 1) is / are correct wrt the set of variables described above?

c) Which network is optimal (if any)?

d) In which order should you consider the variables to construct the optimal network?

e) Network iv) (see Fig. 2) represents a specific type of network. Give the name of this type and *explain* it!

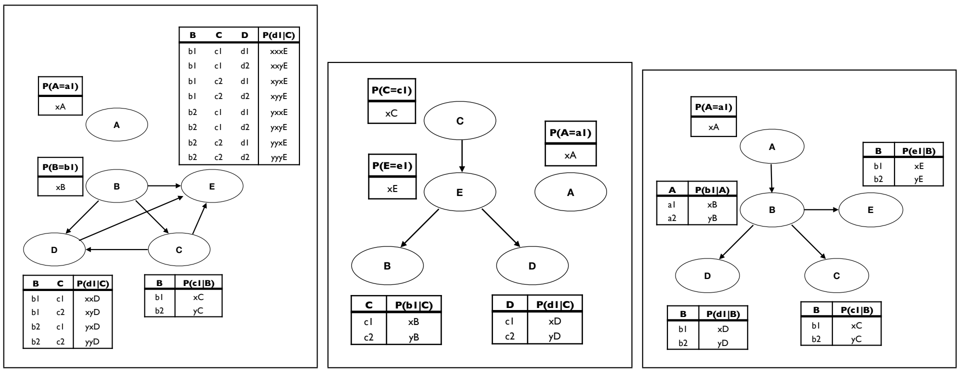


Figure 1: Networks i) (left), ii) (center), iii) (right)

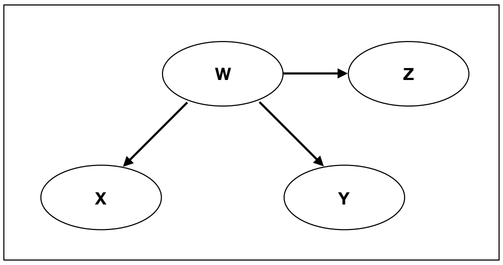


Figure 2: Network iv)

2 Probabilistic reasoning over time (EAT) 10+5=15p

a) A friend of yours has implemented an HMM-based forward filtering approach to localise and track a robot in a grid world, and consults you since you had to solve a similar task before. Your friend has the following setup to work with: the grid is of dimensions 4x8 (rows x columns, see below) and empty, the robot can only move straight and has a high probability to keep its heading (except when facing a wall), and the available sensor gives somewhat noisy information about the distance to some obstacle (wall) in each cardinal direction. Sensor readings are in this case more likely to be correct than you should have experienced yourself (somewhere between 0.125 and 0.25, as the probability for the sensor to be correct is much higher the closer it is to the wall), however, your friend's sensor is failing slightly more often entirely, namely with probability 0.15, independently from where the robot is. The probability for faulty readings resulting in a "one step off" reading are reported with the respective remaining probability for each field. Your friend is not really convinced that the own solution produces something useful, and tells you about it to get your opinion:

"I start with a known position and simply *track* the robot. Somehow, I

only get the really correct position reported in roughly 50% of the steps I do, and the average Manhattan distance in cases the estimate is *not* correct is 1.3. Something is wrong, I think, I might simply skip the calculations altogether!"
What do you answer? (Please argue based on the numbers and estimates as given, they are fictional!)

b) In the course, two other filtering techniques were mentioned as much more common for robot tracking than this HMM-based forward-filtering. Name two of them and explain briefly how they work!

3 Reasoning (JM) (12p)

1. *"Heads I win; tails you lose"* can be represented in propositional logic in the following way:

   $Heads \rightarrow WinMe$
   $Tails \rightarrow LoseYou$
   $\neg Heads \rightarrow Tails$
   $WinMe \rightarrow LoseYou$

   - (2p) Is it possible to use the inference method *Forward Chaining* to derive $LoseYou$? Motivate your answer.
   - (2p) Is it possible to use the inference method *Backward Chaining* to derive $LoseYou$? Motivate your answer.
   - (2p) Is it possible to use resolution to derive $LoseYou$? Motivate your answer.

2. Mark the following formulae as one of:

   - satisfiable but not valid
   - valid
   - not satisfiable

   assuming interpretations belong to the specified set:

   (a) (2p) $(\neg q \wedge \neg p) \wedge \neg(\neg p \vee q)$
   Interpretations: all possible (assignments for p and q).

   (b) (2p) $(p \wedge \neg q) \vee (\neg p \wedge q)$
   Interpretations: all possible (assignments for p and q).

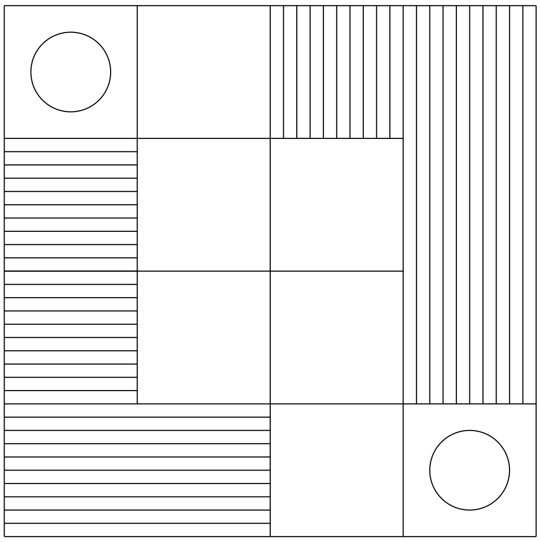
   (c) (2p) $(\neg p \wedge q) \vee (p \wedge \neg q)$
   Interpretations: $\{< p = \text{False}, q = \text{True} >, < p = \text{True}, q = \text{False} >\}$.

4 Games (JM) (12p)

L-game is played by two players on a 4×4 board. Each player has an L-shaped figure which can be turned or rotated in all directions. A move consists of two parts:

1. The player lifts her L-figure and puts it down on the board in a different position than before.
2. If she wants, she can also move one of the two neutral pieces[1] to a new position.

The player who cannot move her L to a new position, looses the game. The following picture illustrates the board in the beginning of the game.



(6p) Assume now that you want to write a program that could play the L-game. Your task consists of representing the problem as an *adversarial search* problem. Describe how state and operators (possible moves) could be represented, how the goal state will be recognized (a *goal-test* function), how possible moves will be generated (a *successor* function), how one of the possible moves will be chosen (a *choose-move* function), etc.

In order to avoid misunderstanding, some precision is necessary in your answer. Therefore it would be a benefit if you could use e.g. list structures (or whatever you deem appropriate) to define the necessary data types you choose to represent state and operators. You can write your functions using a pseudocode.

Remember that your program is going to play against an opponent. Therefore during the search for the next best move you should take into account the possible moves of the opponent.

[1] A neutral piece, marked by a circle on top of it, covers just one square of the board.

(3p) Assume that the search space for the L-game is huge (which is not true in reality). In such case you would probably like to use some form of search-tree cutoff in order to minimize the search time. Given your original solution, how would you introduce α-β pruning in it? I would appreciate an illustrative example using the L-game (NOT the one from the book!).
(3p) Finally, let us look at the α-β pruning. In both α-cutoff and β-cutoff we cut some part of the search tree because we conclude that there is no use in expanding this particular branch of the tree. We perform same reasoning in both α and β cases. *Therefore there is no difference between them!* Please convince me that α-cutoff and β-cutoff differ from each other (that is, that the preceding sentence is false). The best way of doing this is to describe HOW do they differ, I suppose.

5 Search (JM) (12p)

Bilbo and his companions have reached the mountain where Smaug, the dragon, holds its treasure. They enter the mountain's internal labyrinth via a large cave (marked S in the figure 3) and discover a system of caves connected by corridors. They start exploring the caves, looking for ones holding gold (G1 and G2, it happens that there are two of them, but Bilbo and his companions need not to be aware of this fact).

Each corridor takes some hours to pass (denoted by numbers by the arrows describing direction in which a corridor might be taken). Each cave holds a sign saying " you still need at least x hours to reach treasure" (denoted by the numbers inside the circles).

Luckily, Bilbo has once read an old manuscript describing strategies of exploring labyrinths and plans to use his knowledge now. For each of the following search strategies, indicate which gold cave is reached (if any) and list, in order, all the caves visited in between (i.e., states popped off of the OPEN/FRINGE list. When all else is equal, caves should be explored in alphabetical order.

You may use the form below or copy its structure on an answer sheet. Remember that some nodes may occur on this list more than once!

Breadth-First (2p) Gold cave reached: ______

States popped off OPEN:____________________________

Best-First (Greedy) using $f = h$

(2p) Gold cave reached: ______

States popped off OPEN:____________________________

Iterative Deepening (2p) Gold cave reached: ______

States popped off OPEN:____________________________

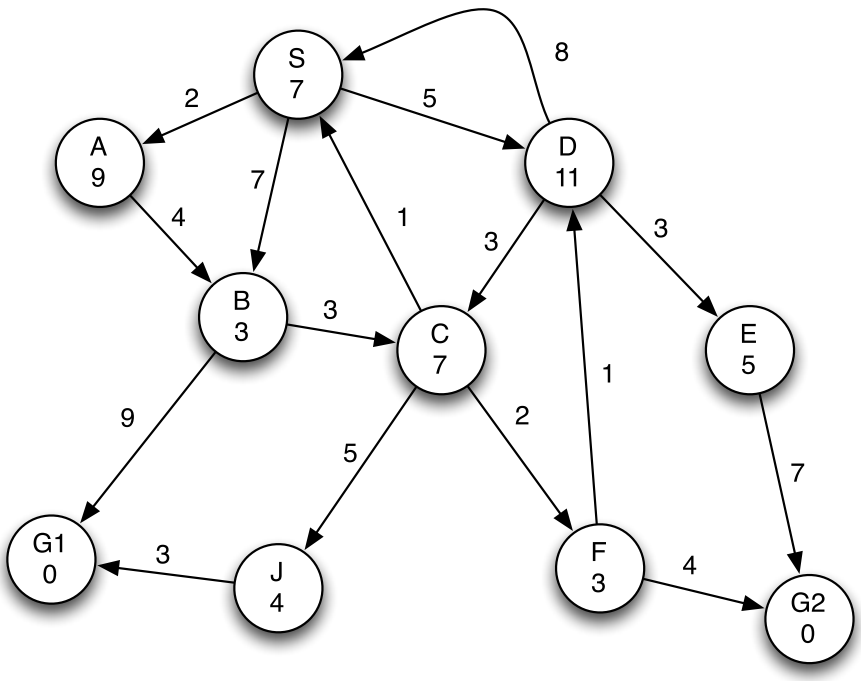


Figure 3: A search space, with costs (over arcs) and heuristic estimates (in nodes).

A* (3p) Gold cave reached: ______

States popped off OPEN:____________________________

— (1p) Is the heuristic estimate illustrated in Fig. 3 admissible? Why or why not?

— (1p) Is the heuristic estimate illustrated in Fig. 4 admissible? Why or why not?

— (1p) Modify any of the heuristic estimates shown in Figs 3 and 4, so that the A* search is optimal, i.e. it expands the least number of nodes searching for the optimal path.

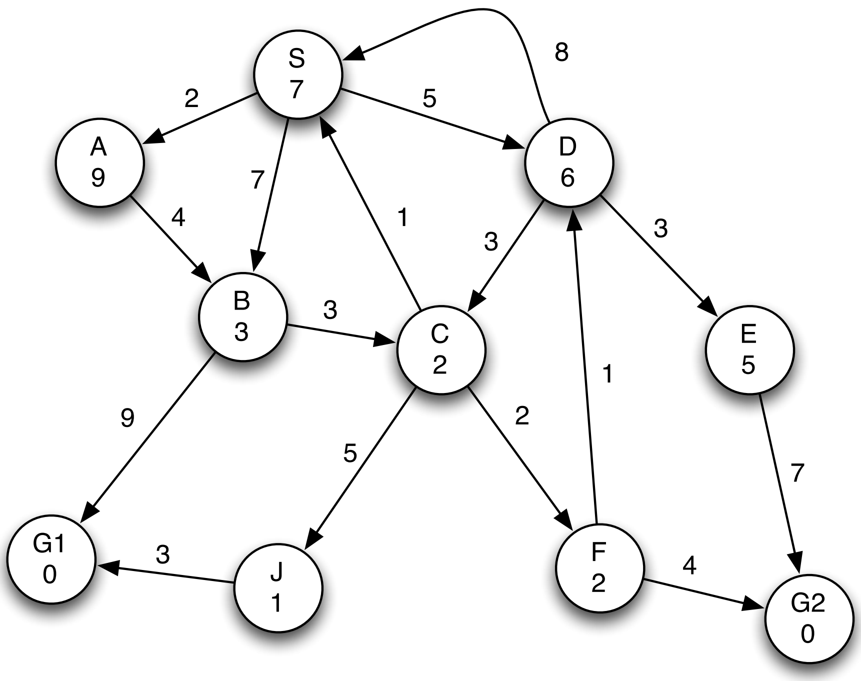


Figure 4: A search space, with costs (over arcs) and heuristic estimates (in nodes).

6 Logistic Regression (PN) (10 + 10 + 10 = 30p)

In 1944, Berkson[2] used logistic regression as a classifier and applied it to the dataset in Table 1. In this question, you will follow him in the exploration of this dataset. The goal of the analysis is to model and predict the outcome of drug ingestion.

6.1 Understanding Logistic Regression

1. (a) Write the logistic function: $y = Logistic(z)$;
   (b) Draw this function;
   (c) Rewrite the logistic function with $z = f(x)$, where $f(x) = mx+b$.
2. (a) In Table 1, each row describes the effect of the dosage of a potentially lethal drug when ingested by a population of individuals, who either survive or die. Describe the meaning of the first row with a dosage of 40. You will focus on the columns: number exposed, survived, died, and mortality rate;

[2] The original paper is provided as a matter of information. You should not need it for the examination.

(b) For dosages of 40 and 100, check that the mortality rates correspond to the ratios of individuals in the table;

3. (a) Given a dosage, logistic regression returns the probability $P(1|\text{dosage})$ of being a member of class 1 (die). Knowing $P(1|\text{dosage})$, explain how to derive the probability of being a member of class 0 (survive), $P(0|\text{dosage})$?

   (b) Using the logistic function, the parameters m and b, write the formulas giving the probabilities: $P_{(m,b)}(1|40)$, $P_{(m,b)}(1|100)$, $P_{(m,b)}(0|40)$, and $P_{(m,b)}(0|100)$; you will not try to compute them;

   (c) Berkson did not used the raw dosage, but applied a logarithmic transform to it, i.e. instead of x, he used $\log_{10} x$. Rewrite your probabilities with this transform;

   (d) Berkson, using a semi manual fitting process, found estimates for m and b of 5.659746 and -10.329884, respectively. Rewrite the logistic function with these values, $m = 5.659746, b = -10.329884$; do not forget the logarithm;

   (e) Compute the estimates of $P_{(m,b)}(1|40)$ and $P_{(m,b)}(1|100)$ with these values. (You will need a calculator). Check that they match the figures in the expected mortality column in Table 1?

Drug concentration	Number exposed	Survive Class 0	Die Class 1	Mortality rate	Expected mortality
40	462	353	109	.2359	.2206
60	500	301	199	.3980	.4339
80	467	169	298	.6380	.6085
100	515	145	370	.7184	.7291
120	561	102	459	.8182	.8081
140	469	69	400	.8529	.8601
160	550	55	495	.9000	.8952
180	542	43	499	.9207	.9195
200	479	29	450	.9395	.9366
250	497	21	476	.9577	.9624
300	453	11	442	.9757	.9756

Table 1: Berkson's dataset. Adapted and simplified from the original article that described how to apply logistic regression to classification by Joseph Berkson, Application of the Logistic Function to Bio-Assay. *Journal of the American Statistical Association* (1944).

6.2 Determining the Error

One way to find m and b is to minimize the sum of squared errors between the observed classes and the predicted probabilities. In this section, you will determine the error resulting from estimates of m and b: The mean of the sum of squared errors is given by:

$$SSE = \frac{1}{\#\text{individuals}} \sum_{\text{individuals}} (y - P(1|\text{dosage}))^2.$$

1. (a) Write the formula to compute the sum of squared errors for all the individuals exposed to a dosage of 40, $(y - P(1|40))^2$, where the y value can be either 0 (survived) or 1 (died). This formula will involve the number of individual who survived this dosage, the number who died of it, as well as the estimated probability, $P(1|40)$. You will write the estimate $P(1|40)$ with the logistic function and the m and b parameters; do not forget the logarithm;

   (b) Compute this value with a calculator; do not forget to normalize it by the total number of individuals exposed to a dosage of 40 (i.e. divide it by 462);

2. You will now show that this value, the normalized sum of the squared errors, is minimal when $P(1|40)$ is exactly equal to the proportion of individuals who died (i.e. $\frac{109}{462}$).

   (a) Rewrite the mean of the sum of squared errors in Question 1a, Sect. 6.2 with p, the proportion of individuals who died and $\hat{p}$, the estimate of this proportion; i.e. you will replace the observed frequencies of death, $\frac{109}{462}$,with p and the estimated ones, $P(1|40)$, with $\hat{p}$ and you will obtain a polynomial of degree 2 with p and $\hat{p}$; in this exercise, you should replace any occurrence of the logistic function with $\hat{p}$ or a simple expression involving it, and any observed frequencies of death or survival with p or a simple expression involving it;

   (b) Compute the derivative of the polynomial with respect to $\hat{p}$ and show that the minimum is reached when $p = \hat{p}$.

   (c) Give the minimal value of the sum of squared errors; it should be a very simple function of p (a polynomial);

   (d) Compute the value for a dosage of 40 and check that it matches well the number you found in Question 1b, Sect. 6.2.

6.3 Fitting the Parameters

In this section, you will determine how to estimate m and b and describe the fitting process knowing that

$$\hat{p} = P(1|x) = \frac{1}{1 + e^{-(mx+b)}}.$$

1. Rewrite the mean of the sum of squared errors:

$$\frac{1}{\#\text{individuals}} \sum_{\text{individuals}} (y - P(1|\text{dosage}))^2$$

   with the logistic function, where we will use x to denote the dosage, m, and b, the two parameters to find. You will denote this equation: $SSE(b, m, x)$ and you will ignore the logarithm.

2. Compute the partial derivatives of this equation with respect to m and b (the gradient), $\frac{\partial SSE(b,m,x)}{\partial b}$ and $\frac{\partial SSE(b,m,x)}{\partial m}$, knowing that the derivative of the logistic function is $\frac{e^{-x}}{(1+e^{-x})^2}$;

3. Knowing the gradient descent is given by this iteration:

$$\begin{aligned} b &\leftarrow b - \alpha \frac{\partial SSE(b,m,x)}{\partial b}, \\ m &\leftarrow m - \alpha \frac{\partial SSE(b,m,x)}{\partial m}, \end{aligned}$$

   give the update rule for b and m.

4. When will you stop the iteration?

References

Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American Statistical Association*, 39(227):357–365.

Good Luck!