Vera Koponen
Department of Mathematics
Uppsala University

2023-12-14

Module 3

Logic, probability and statistical relational learning

Examination of Module 3

Deadline: Your solutions to the assignments below need to be uploaded on Canvas, in **one pdf-file**, at the latest on **21 January 2024**, before midnight. You may submit hand written solutions but they must be converted into **one pdf-file**.

You may work together to solve the assignments, but each person must submit individually written solutions. If I find two or more solutions that are identical then the corresponding persons will have to rewrite the solutions. One can get maximally 40 points on this exam and at least 28 points are necessary for passing this exam.

All references below refer to my *Notes on logic and probability* that are available on Canvas.

1. (Inductive logic programming in the context of propositional logic) Le p, q, r be propositional variables. (6p)

(a) Suppose that all the formulas $\neg p \vee r$, $\neg q \vee \neg r$ and p are true. Determine the truth values of q and r.

(b) Construct a CNF φ which avoids $\mathbf{N}$ and covers $\mathbf{E}$ where

$$\mathbf{N} = \{(0,1,0), (1,0,0), (1,1,0)\} \text{ and}$$
$$\mathbf{E} = \{(0,0,0), (0,0,1), (0,1,1), (1,0,1), (1,1,1)\}.$$

Try to simplify φ as much as possible by removing unnecessary [1] literals (if such exist).

In the rest of the assignments L and $\mathcal{M}$ refer to the language L and L-structure $\mathcal{M}$ that we now define. Let $L = \{P, Q, R, S, T\}$ where the relation symbols P, Q, R, S, T have arities $1, 1, 2, 2, 2$, respectively.

Let $\mathcal{M}$ be the L-structure with domain $M = \{1, 2, 3, \ldots, 20\}$ and the following interpretations of the relation symbols:

$$P^{\mathcal{M}} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\},$$
$$Q^{\mathcal{M}} = \{1, 2, 3, 4, 5, 11, 12, 13, 14, 15\},$$
$$R^{\mathcal{M}} = \{(i, j) : i \neq j, 1 \leq i \leq 5 \text{ and } 1 \leq j \leq 5\},$$
$$S^{\mathcal{M}} = \{(2,5), (3,1), (4,3), (1,2), (5,4), (2,3), (4,2), (1,3), (3,5), (4,5)\},$$
$$T^{\mathcal{M}} = \{(2,5), (3,1), (4,3), (1,2), (5,4), (3,2), (2,1), (1,4), (2,4), (5,2)\}.$$

[1] By "unnecessary literal" I mean that such a literal can be removed and the thus obtained formula still covers $\mathbf{E}$ and avoids $\mathbf{N}$.

2. (Inductive logic programming and concept learning) Learning the concept $R(x, y)$: Find literals $\theta_1(x, y), \ldots, \theta_n(x, y)$ which do not contain R and such that $\mathcal{M} \models \forall x \forall y\big((\theta_1(x, y) \wedge \ldots \wedge \theta_n(x, y)) \rightarrow R(x, y)\big)$ and $\mathcal{M} \models \exists x \exists y\big(\theta_1(x, y) \wedge \ldots \wedge \theta_n(x, y)\big)$ Observe that a literal denoted $\theta_k(x, y)$ above need not necessarily contain both variables x and y and it may be of the form $x = y$ or $\neg(x = y)$.

There may be several possibilities, but to get maximal credit on this assignment you need to find such $\theta_1(x, y), \ldots, \theta_n(x, y)$ so that

$$\mathcal{M} \models \forall x \forall y\big((\theta_1(x, y) \wedge \ldots \wedge \theta_n(x, y)) \leftrightarrow R(x, y)\big)$$

and note the symbol '$\leftrightarrow$' in the last formula. (8p)

3. (Learning a Bayesian network) Construct a Bayesian network from $\mathcal{M}$ in the way explained in Chapter 5.1 using the following order of the formulas, also viewed as random variables from M^2 to $\{0, 1\}$: $P(x)$, $P(y)$, $Q(x)$, $Q(y)$, $R(x, y)$, $S(x, y)$, $T(x, y)$. As explained in Chapter 4, we can identify $P(x)$ with a random variable $P_x : M^2 \rightarrow \{0, 1\}$ such that $P_x(a, b) = 1$ if $\mathcal{M} \models P(a)$ and $P_x(a, b) = 0$ otherwise. $P(y)$ can be identified with a random variable $P_y : M^2 \rightarrow \{0, 1\}$ such that $P_y(a, b) = 1$ if $\mathcal{M} \models P(b)$ and $P_y(a, b) = 0$ otherwise. If you need you can use conditions like '$x = y$' or '$x \neq y$' when specifying conditions for conditional probabilities (although we don't view such formulas as vertices of the network).

It is also possible, and acceptable, to construct a lifted Bayesian network with vertices P, Q, R, S, T (considered in this order) as defined in Definition 6.2.1 in my notes. The procedure will be essentially the same and you have to do the same calculations. The resulting lifted Bayesian network will be similar to the one obtained by the first approach, but not identical since the later has 5 vertices while the first has 7. (In the later approach $P(x)$ and $P(y)$ are identified, and $Q(x)$ and $Q(y)$ are identified.)

When determining (conditional) independence you can use the usual mathematical definition of independence, as there is no need for "approximation" in this assignment. Use the *counting measure* (i.e. uniform probability distribution) $\mu : M \rightarrow \mathbb{R}$ when determining (conditional) probabilities. (12p)

4. (Incomplete data) Now assume that for each of the pairs $(1, 4), (5, 10), (14, 6)$ and $(3, 5)$ we do not know if it belongs to $R^{\mathcal{M}}$ or not, and for each of the pairs $(4, 3), (1, 3)$ and $(6, 8)$ we do not know if it belongs to $S^{\mathcal{M}}$ or not, but otherwise all assumptions are as before.

Estimate $\mu(R(x, y))$ (also denoted $\mu(R = 1)$) and $\mu(S(x, y) \mid R(x, y))$ (also denoted $\mu(S = 1 \mid R = 1)$) by the methods described in Chapter 5.3. (6p)

5. (Inference/prediction) Let $\mathbb{G}$ denote the Bayesian network constructed in assignment 3. Let $D_n = \{1, \ldots, n\}$ be a domain. Let $\mathbf{W}_n$ be the set of all L-structures with domain D_n. For each positive integer n, the Bayesian network $\mathbb{G}$ determines a probability distribution $\mathbb{P}_n$ on $\mathbf{W}_n$ in the way explained by examples in sections 4.2.9, 5.2.1 and 5.2.2. (If one uses the approach in Chapter 6 then Definition 6.2.5 tells how $\mathbb{G}$ determines a probability distribution on $\mathbf{W}_n$. However, it is important to understand what that definition means in the case of the concrete Bayesian network obtained in assignment 3.) Material on the slides from the lectures may also be of help to understand how one can reason about the distribution $\mathbb{P}_n$ on $\mathbf{W}_n$.

Since we are interested in results for large n you can choose some lower bound to n if you need it.

(a) Suppose that $a, b \in D_n$ and $a \neq b$. Let $\mathbb{P}_n(R(a,b))$ be an abbreviation of $\mathbb{P}_n(\{\mathcal{A} \in \mathbf{W}_n : \mathcal{A} \models R(a,b)\})$, and similarly for S in place of R. Determine $\mathbb{P}_n(R(a,b))$ and $\mathbb{P}_n(S(a,b))$ and show how you get to your conclusions.

(b) For any first-order sentence φ let $\mathbb{P}_n(\varphi)$ be an abbreviation of $\mathbb{P}_n\big(\{\mathcal{A} \in \mathbf{W}_n : \mathcal{A} \models \varphi\}\big)$. Determine $\lim_{n\to\infty} \mathbb{P}_n\big(\forall x \exists y \exists z (S(x,y) \wedge \neg S(x,z))\big)$ and show how you get your conclusion.

(8p)