CMP9794M Assessment Item 1 Briefing 2022-2023

Module Code & Title: CMP9794M Advanced Artificial Intelligence

Contribution to Final Module Mark: 50%

Description of Assessment Task and Purpose:

This individual assessment **item 1** is an assignment. Your task is to use materials covered during the lectures and workshops of this module to implement a software application to solve Medical Diagnosis problems by calculating probabilities of Stroke and Heart Disease from data. The sources of data and random variables of each of these datasets are:

- **Heart Disease**: age, gender, chest pain type, BP, cholesterol, FBS over 120, EKG results, max HR, exercise angina, ST depression, slope of ST, number of vessels fluro, thallium, heart disease
- **Stroke**: gender, age, hypertension, heart disease, ever married, work type, residence type, average glucose level, BMI, smoking status, stroke

Since data will be pre-processed, your work should use the data provided via Blackboard – do not use the original sources of data.

Task 1a: Discrete Bayesian Networks (30%)

The first task consists of implementing *discrete Bayesian Networks* to answer probabilistic queries such as:

```
P(heart disease=true|gender=male, chest pain type=4)
P(stroke=true|gender=female, age=37, smoking status=smokes)
```

or

```
P(heart disease|gender= female, chest pain type=3)
P(stroke|gender=male, age=47, smoking status=formerly smoked)
```

The queries above in the discretised dataset correspond to the following:

```
P(target=1|gender=1, cp=4)
P(stroke=1|gender=Female, age=2, smoking_status=smokes)
```

or

```
P(target|gender=0, cp=3)
P(stroke|gender=Male, age=2, smoking_status=formerly smoked)
```

Your software solution (programs) should be written in Python and it should do the following to be able to answer probabilistic queries:

1. Read the datasets (in CSV format) available via Blackboard.
2. Read from a configuration file a predefined structure of each of your Bayesian networks or use randomly generated structures.
3. Learn the parameters of the Bayesian networks (one network per dataset) using Maximum Likelihood Estimation for example.
4. Answer probabilistic queries using one of the algorithms provided in the module.

Your implementation above should use at least one dataset (of your choice) but preferably both of them.

Task 1b: Comparison of Inference and Structure Learning Algorithms (40%)

This task consists of comparing the performance of different inference algorithms such as

a. Rejection sampling
b. Likelihood weighting
c. Gibbs sampling

Whilst the algorithms can be those discussed/provided during the lectures and workshops, at least one of them should be an implementation not provided as part of the module. You should compare the performance of these algorithms in terms of predictive power (disease classification accuracy), AUC (area under curve) score, Brier score, and inference time (in seconds), among others.

In addition to the above, this task consists of comparing the performance of structure learning algorithms such as

d. Conditional independent tests
e. Network scores
f. Hybrid algorithm

The structure learning algorithms can be those discussed/provided during the lectures and workshops – but at least one of them should be an implementation not provided as part of the module. Your comparison of algorithms can also be in terms of predictive power (e.g., disease classification accuracy), statistical distance (e.g., Kullback-Liebler Divergence), and inference time (e.g., runtime seconds) -- others are optional.

Task 1c: Gaussian Discrete Bayesian Networks (30%)

This final task asks you will extend task 1a (and task 1b if you opt to address it) to support *conditional Gaussian Bayesian Nets* – combining discrete and continuous random variables as opposed to only discrete variables. As part of this task, you should compare their performance (using the original datasets without missing values) against *discrete Bayesian Networks* (using the discretised datasets) using metrics discussed during the lectures/workshops such as Log Likelihood (LL) and BIC (Bayesian Information Criterion), among others.

Optionally, your implementation could graphically display the structure of the targeted Bayesian network and the probability distributions required to answer a probabilistic query.

Please indicate and justify in your report the methods used for implementing the tasks above and read the Criterion Reference Grid for details on how your work will be graded.

Learning Outcomes Assessed:

- [LO1] Critically appraise a range of AI techniques for knowledge representation, reasoning and decision-making under uncertainty, identifying their strengths and weaknesses, and selecting appropriate methods to serve particular roles.

- [LO2] Design and develop a software algorithm for solving complex AI problems in an application domain of interest.

Knowledge & Skills Assessed:

- *Subject Specific Knowledge, Skills and Understanding*: e.g., literature searching, referencing, project planning, techniques and skills subject-specific knowledge.
- *Professional Graduate Skills*: e.g., independence and personal responsibility, adaptability, verbal communication, written communication, creativity, critical thinking, IT skills, problem solving, effective time management, working under pressure to meet deadlines.
- *Emotional intelligence*: e.g., self-awareness, self-management, motivation, resilience, self-confidence.

Assessment Submission Instructions:

You must make an electronic submission of your work in PDF format by using the assessment link on Blackboard for this component. You must attend the lectures and workshops for further details, guidance and clarifications regarding these instructions.

DO NOT include this briefing document with your submission.

The deadline for submission of this work is included in the School Submission dates on Blackboard.

Date for Return of Feedback:

Please see the School's assessment dates spreadsheet.

Format for Assessment:

Your submission describing the solutions to the medical problems above should include a concise report of 3 pages (excluding references) using the IEEE template provided via Blackboard, which must be submitted as a PDF file on Blackboard. Whenever possible, you should cite previous works from the related literature to justify your arguments or choices.

The software implemented (source code) to solve the targeted problems should be submitted as a ZIP file in the assignment support documentation on Blackboard.

Feedback Format:

Written and numerical feedback will be provided via Blackboard, and additional feedback can be provided upon request in a meeting or via email.

Additional Information for Completion of Assessment:

This assessment is an individually assessed component. Your work must be presented according to the Lincoln School of Computer Science guidelines for the presentation of assessed written work. Please make sure you have a clear understanding of the grading principles for this component as detailed in the accompanying Criterion Reference Grid. You are expected to take the following into account:

- Your submission – using the provided template (in MS Word or Latex) via Blackboard – should be a concise report of maximum **3 pages** as main body + **one additional page** only for references.

- Submissions failing to stick to the length requirement above and/or omitting source code **will not be marked**.
- Please make sure that you submit **your own work** (writing, results) and not somebody else's. Failure to do so will incur plagiarism or collusion, which will the reported to the School for investigation of potential academic misconduct – see more details below.

If you are unsure about any aspect of this assessment component, please seek the advice of a member of the delivery team.

Assessment Support Information:

Assignment support will be provided during workshop sessions and surgery hours.

Important Information on Dishonesty & Plagiarism:

University of Lincoln Regulations define plagiarism as 'the passing off of another person's thoughts, ideas, writings or images as one's own...Examples of plagiarism include the unacknowledged use of another person's material whether in original or summary form. Plagiarism also includes the copying of another student's work'.

Plagiarism is a serious offence and is treated by the University as a form of academic dishonesty. Students are directed to the University Regulations for details of the procedures and penalties involved.

For further information, see www.plagiarism.org