AI ASSIGNMENT 3
Date of Submission : 2023 November 18

Total Marks : 25 Marks
Theory section : 11 Marks
Computational section : 14 marks

Instructions:

1. Assignments are to be attempted individually.
2. Submit the assignment in a zip folder with name AI3_Rollnumber
3. Submit a file named AI3_Rollnumber_Result.pdf and combine the answers/results for the theory and computational question.
4. Submit a file named AI3_Rollnumber_Codes.py file for the python codes.
5. If any of the file is missing a penalty of -5 marks.
6. **Programming Language :** Python
7. **Extension and Penalty clause:**
   - Even a 1 minute late submission on google classroom will be considered as late. Please turn-in your submissions atleast 5 minutes before the deadline.
   - Not explaining the answers properly will lead to zero marks.
   - Please make sure to turn in the Assignment. Missing files will receive zero marks.

Theory

1. (6 marks) Given the following statements below, answer the questions. The precision is up to 3 decimal places. So round off the values computed to upto 3 decimal places. A statement may have more than one proposition.
   (i) (2.5 marks) Identify the random variables in the statements below , and write each of the following statements using symbols for random variables, logical connectives where necessary, and conditional probability notation.
   (ii) (1 mark) Verify that these propositions create a valid probability distribution. List the set of axioms that they satisfy.
   (iii) (1 mark) Populate the full joint probability distribution table.
   (iv) (1.5 marks) Use the joint distribution table and check for conditional independence between all the random variables that you have identified.

(a) About 82.5 % people have travelled and have caught either corona or other diseases.

(b) Of the people who had travelled 15 % have mild and 22 percentage have severe cases of corona, respectively.

(c) Given that a person travelled the chance they caught a disease other than corona is 0.485 rounded to 3 decimal places.

(d) About 24 % of people died of diseases other than corona after travelling.

(e) There is 0.025 probability that a person has not travelled and has severe case of corona.

(f) Given a person has not travelled the probability that the person is severely sick is about 0.457 rounded to 3 decimal places.

(g) The probability that a person has died and did have corona is 0.059.

(h) About 70 % people had mild or severe cases of any disease.

(i) There is 80 % chance that a person has travelled given that he is severely sick.

(j) There is 50 % chance a person had corona whether they travelled or not.

2. (5 marks) You are playing a game where there are rooms, and you need to escape from one room to go to the next level. Each room shows you three doors. Behind one door is the key to the next room, and if you choose the other two doors, you lose a life. You pick one door. A person who knows what's behind the door waits for you to choose your door. After you decide, he opens another door, revealing that you lost your life.

   (a) (1 mark) Should you switch your choice to the other unopened door to maximize your chance of winning the key?

   (b) (2 marks) However, if he occasionally makes a mistake, he reveals the loss of life with a probability of 1/3 and correctly reveals the key with a probability of 2/3. In this scenario, should you switch your choice to maximize your chances of winning the key?

   (c) (1 mark) If you choose to switch, what is the conditional probability that you win the key if the man has mistakenly revealed the door that shows life lost?

   (d) (1 mark) Additionally, what is the conditional expectation of your prize (key/life lost) based on your choice to switch or stick, considering both possible scenarios? Would you choose to switch or stick based on the conditional expectation?

Computational

1. Create a Bayesian Network (14) :

   (a) (2 mark) Dataset : Load the wine quality dataset https://archive.ics.uci.edu/dataset/109/wine. The dataset is continuous and therefore discretization would be required to build the network. You can explore continuous/hybrid models also. Build classification model based on class variable in the data for performance evaluation (accuracy) https://scikit-learn.org/stable/modules/generated/sklearn.metrics.

accuracy_score.html of the network. You can use any available open-source packages to build the network, example `bnlearn` package {https://pypi.org/project/bnlearn/}.

(b) Instructions to Construct the network: You will need to construct and evaluate a total of three networks A, B and C as described below:

- (2 mark) Construct a Bayesian network (A) for the data. Visualise the network and the probability distribution. Describe a few examples of parent and child nodes.
- (1 mark) Prune the network (A) for better performance on the class variable. Let the new network be (B). Explain your method of pruning.
- (1.5 marks) Use methods other than pruning to construct the best model on the dataset. Let (A) be the new improved network.
- (2 marks) Use feature selection method {https://scikit-learn.org/stable/modules/feature_selection.html} on the data and construct a Bayesian network on new data (C). Plot the probability distribution $P(class|F_1, F_2)$ to F_1 and F_2 where F_1 and F_2 are features that are significant for the class prediction. Describe the change in probability distribution ? (Whether it is the same ?(why/why not)) (Example fig:14.6 page 521 in the book Artificial Intelligence: A Modern Approach by Russsell & Norvig, 3^{rd} edition)
- (1 marks) Compute for at least 4 cases the posterior probabilities of variable(s) of interest (queries) given evidence (observations). Do the results match your intuition and/or domain knowledge? Explain your answer. (You can use API)
- (1 mark) Compare the prediction performance based on the three models (A), (B) and (C).
- (3.5 marks) Presentation, Viva and demonstration of the Code.
- There will be relative grading based on your explanation and innovative implementations.

(c) Instruction for submission : Please follow the guidelines below to submit the findings. There are penalties for not following the format outlined below.

- (-1 mark) The data should be imported correctly, and any pre-processing should be explained. Otherwise explain why there was no pre-processing.
- (-1 mark) Make a proper pdf file with all the results.
- (-3 mark) Make a table with the performance measures of the three Bayesian networks.
- (-1 mark) The explanations and descriptions should be proper.
- (-1 mark) The code is incorrect.

END