Homework Assignment #1

Due date – Feb. 20, 2018 (Tue), in class.

Bayesian Network

Consider a Bayesian network whose graph structure is shown in Figure 1. In the given BN, all variables take binary values (0 or 1), and the conditional probability distributions are defined as follows:

$$P(A1) = 0.9$$

$$P(B1) = 0.15$$

$$P(C1|A0, B0) = 0.1, P(C1|A1, B0) = 0.9, P(C1|A0, B1) = 0.8, P(C1|A1, B1) = 0.2$$

$$P(D1|C0) = 0.1, P(D1|C1) = 0.85$$

$$P(E1|C0) = 0.75, P(E1|C1) = 0.25$$

$$P(F1|D0, G0) = 0.1, P(F1|D1, G0) = 0.9, P(F1|D0, G1) = 0.8, P(F1|D1, G1) = 0.1$$

$$P(G1|D0, E0) = 0.25, P(G1|D1, E0) = 0.85, P(G1|D0, E1) = 0.75, P(G1|D1, E1) = 0.9$$

where $X1$ denotes $X = 1$ and $X0$ denotes $X = 0$.

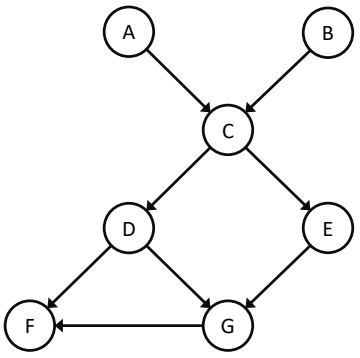

Figure 1: The directed acyclic graph of a Bayesian network.

Problem 1. Predicting the graph structure (100 points)

(a) Generate 50 observations based on the given BN. Use the following measures to predict the edges:

(i) correlation

(ii) partial correlation

(iii) mutual information.

For each approach, repeat the experiments and plot the overall ROC curve. Also plot a bar chart showing the average AUC for the different approaches (along with the error bar showing the standard deviation). Discuss your results.

(b) Repeat the experiment in part-(a) for 100 observations and 500 observations. Discuss your results.

(c) Generate 500 observations. Write a code for predicting the P-map based on the method that was discussed in the class. Use the code to predict the P-map from your data.

(i) According to your prediction, how many P-maps belong to the same I-equivalence class?

(ii) Draw the class-PDAG for the predicted P-map.

(iii) Evaluate the edge predictions. How many true/false positives/negatives did you get?

(iv) Evaluate the immorality predictions. How many true/false positives/negatives did you get?

(d) Repeat the experiment in part-(c) for 2,500 observations. Compare and discuss the results.

(e) Can you improve the prediction accuracy further? Please propose alternative methods or heuristics that may enhance the predictions. Verify the proposed method through simulations and compare it with the method you implemented in part-(c).