ECE 364: Assignment #3

1. (10 pts)

   (a) (5 pts) Suppose we wish to calculate $P(F|E_1, E_2)$ and we have no conditional independence information. Which of the following sets of probabilities are sufficient for the calculation and why?

   i. $P(E_1, E_2)$, $P(F)$, $P(E_1|F)$, $P(E_2|F)$

   ii. $P(E_1, E_2)$, $P(F)$, $P(E_1, E_2|F)$

   iii. $P(F)$, $P(E_1|F)$, $P(E_2|F)$

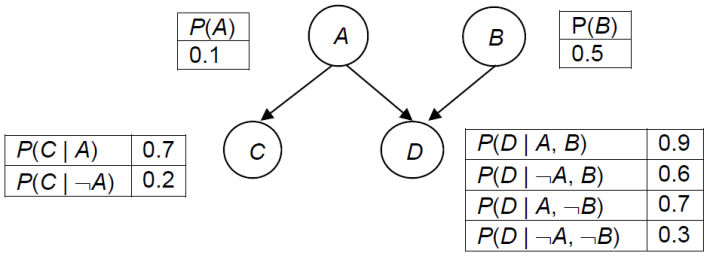
   (b) (5 pts) Suppose we know that $P(E_1|F, E_2) = P(E_1|F)$ for all values of F, E_1, E_2. Now which of the above three sets are sufficient and why?

2. (10 pts) Consider a naive Bayes model with three binary descriptive features, X_1, X_2, X_3, and one binary target feature, Y.

   (a) (5 pts) How many probabilities must be estimated to train such a naive Bayes model?

   (b) (5 pts) How many probabilities would have to be estimated to learn the above model if we did not make the naive Bayes conditional independence assumption?

3. (10 pts) Consider the following Bayesian Network containing four binary random variables.



   (a) (3 pts) Compute $P(\neg A, B, \neg C, D)$.

   (b) (7 pts) Compute $P(A|B, C, D)$.

4. (10 pts) You are designing a troubleshooting advisor for PCs. Let CF be a binary random variable representing whether the computer fails ($CF = true$) or not. Assume there are two possible causes of failure: Electricity-Failure and Malfunction-of-the-Computer represented using the binary random variables EF and MC, respectively. Assume the following probabilities:

   $P(EF) = 0.1$, $P(MC) = 0.2$, $P(CF|\neg EF, \neg MC) = 0.0$, $P(CF|\neg EF, MC) = 0.5$, $P(CF|EF, \neg MC) = 1.0$, $P(CF|EF, MC) = 1.0$.

   (a) (2 pts) Draw the Bayesian Network for this problem.

   (b) (2 pts) Compute $P(CF, \neg EF, MC)$.

(c) (3 pts) Compute $P(MC|EF)$.

(d) (3 pts) Compute $P(EF|CF)$.

5. (20 pts) **Coding project**

   In this project, you will train naive Bayes models and Bayesian networks to determine whether or not a patient has heart disease.

   The dataset consists of descriptive numerical features, including: age in years, trestbps: resting blood pressure, chol: serum cholesterol in mg/dl, thalach: maximum heart rate achieved, oldpeak: ST depression induced by exercise relative to rest, and ca: number of major vessels colored by fluoroscopy.

   The target feature is a binary variable, where 1 indicates the presence of heart disease.

   You will fit naive Bayes models and investigate the effect of different hyperparameters including different feature distributions and strategies for deriving categorical features. You will also fit Bayesian networks and do experiments with networks capturing other feature relationships.

   Please see the Jupyter notebook for more details.

   GitHub repository for ECE364 coding projects: https://github.com/JHA-Lab/ece364_2024