ELE 364: Assignment #3

1. (10 pts) Consider the dataset shown below. It has two descriptive features: Chest Pain (CP) and Blood Pressure (BP) and the target feature is Heart Disease (HD).

Table 1: A pruning set

ID	Chest Pain (CP)	Blood Pressure (BP)	Heart Disease (HD)
1	false	high	false
2	true	low	true
3	false	low	false
4	true	high	true
5	false	high	false

Suppose the Shannon mutual information between random variables X and Y is given by:

$$I(X,Y) = \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}$$

   (a) (8 pts) What is the Shannon mutual information between descriptive feature BP and target feature HD?

   (b) (2 pts) If the two random variables X and Y are independent of each other, what is their Shannon mutual information?

2. (10 pts) The following dataset is used to predict whether someone is at high risk for diabetes (Sys. BP: systolic blood pressure; Resp.: respiration rate; Temp.: body temperature; BMI: body mass index; FBS: fasting blood sugar).

User ID	Sys. BP	Heart Rate	Resp.	Temp.	BMI	FBS	Result
1	120	72	16	99	26	80	NO
2	180	40	12	96	30	120	YES
3	150	110	22	100	35	150	YES
4	84	80	18	102	22	70	NO
5	112	60	17	98	38	142	YES
6	70	70	16	97	19	75	NO
7	100	100	15	95	21	83	NO
8	190	50	14	96	32	126	YES

   (a) (5 pts) Create a Naive Bayes model that uses probability density functions to model the descriptive features in the above dataset (assuming that the features are normally distributed).

   (b) (5 pts) What prediction will the naive Bayes model return for the following query?

User ID	Sys. BP	Heart Rate	Resp.	Temp.	BMI	FBS	Result
9	110	77	17	97	23	92	

3. (10 pts)

   (a) (5 pts) For Bayesian network I given below, find $P(\neg a \mid b)$.

   (b) (5 pts) For Bayesian network II given below, find $P(\neg a, b, c, \neg d)$.

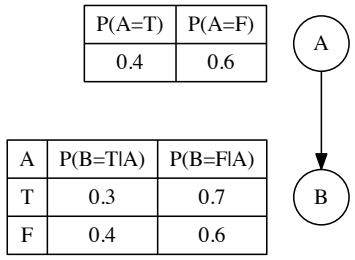

Figure 1: Bayesian network I.

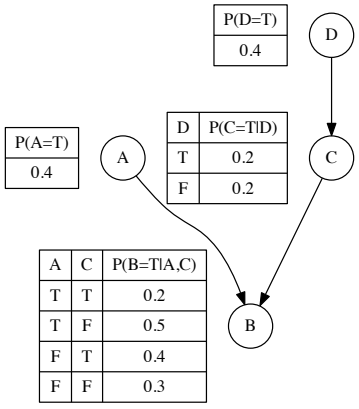

Figure 2: Bayesian network II.

4. (10 pts) The following is a description of the causal relationship between storms, behavior of burglars and dogs, and house alarms: "Stormy nights are rare. Burglary is also rare, and if it is a stormy night, burglars are likely to stay at home. Dogs don't like storms either, and if there is a storm, they like to go inside. Your house alarm is designed to be triggered if a burglar breaks into your house, but sometimes it can be set off by your dog coming into the house, and sometimes it might not be triggered even if a burglar breaks in (the alarm could be faulty or the burglar may be very good)."

   (a) (2 pts) Define the topology of a Bayesian network that encodes these causal relationships.

   (b) (2 pts) The table below lists a set of instances from the house alarm domain. Using the data in this table, create the conditional probability tables (CPTs) for the Bayesian network you created.

   (c) (6 pts) What value will the Bayesian network predict for ALARM, given that there is a storm but we don't know if a burglar has broken in or where the dog is?

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:

ID	STORM	BURGLAR	DOG	ALARM
1	false	false	false	false
2	false	false	false	true
3	false	false	true	false
4	false	true	false	false
5	false	true	false	true
6	false	true	true	true
7	true	false	true	true
8	true	false	true	false
9	true	false	true	false
10	true	true	false	true

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.

See the Jupyter notebook for more details.