CSE 5523: Problem Set 1

Due via Carmen dropbox at 11:59 PM 9/13.

1. [6 pts.] Assume the following joint distribution for $\mathsf{P}(A, B)$:

$$\mathsf{P}(A{=}0, B{=}0) = 0.5$$
$$\mathsf{P}(A{=}0, B{=}1) = 0.2$$
$$\mathsf{P}(A{=}1, B{=}0) = 0.2$$
$$\mathsf{P}(A{=}1, B{=}1) = 0.1$$

   (a) What is the marginal probability of $\mathsf{P}(A{=}0)$?

   (b) What is $\mathsf{P}(B{=}1 \mid A{=}0)$?

   (c) What is $\mathsf{P}(A{=}B)$?

2. [10 pts.] Use the joint distribution from the above problem to show that the following cyclic conditional dependency structure would lead to a violation of the axioms of probability (in other words, show that this is true by plugging in the numbers and showing that one or more of the Kolmogorov probability axioms are violated):

$$\mathsf{P}(A, B) \stackrel{\text{def}}{=} \mathsf{P}(A \mid B) \cdot \mathsf{P}(B \mid A)$$

3. [10 pts.] Without making any independence assumptions, which of the following statements are always true?

   (a) $\mathsf{P}(X, Y, Z) = \mathsf{P}(X) + \mathsf{P}(Y) + \mathsf{P}(Z)$

   (b) $\mathsf{P}(X, Y \mid Z) = \mathsf{P}(X \mid Z) \cdot \mathsf{P}(Y \mid Z)$

   (c) $\mathsf{P}(X, Y) = \mathsf{P}(X) \cdot \mathsf{P}(Y)$

   (d) $\mathsf{P}(X, Y) = \sum_{z \in Z} \mathsf{P}(X, Y, Z{=}z)$

   (e) $\mathsf{P}(X, Y) = \mathsf{P}(X) + \mathsf{P}(Y) - \mathsf{P}(Z)$

4. [20 pts.] PROGRAMMING:

   (In general for programming problems you should hand in the following as separate files:

   - a copy of each program file you write,
   - a representative sample of each input file you use,
   - a representative sample of each output you produce.

   Your programs should be as short as possible.)

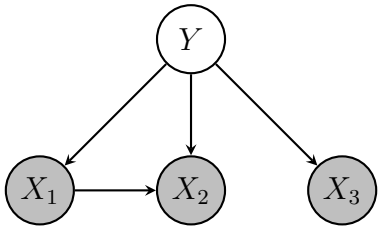
   Write a program called 'less_naive_bayes.py' that takes the following as command-line arguments in the following order:

(a) a filename of a training csv file, containing data in columns for one hidden variable followed by three observed variables, delimited by commas, and

(b) a filename of a test csv file, containing data in columns for three observed variables, delimited by commas, with headers matching the last three columns of the training file.

Your program should perform Naive Bayes training on the data in the training file and print out the full joint probability for each possible value of the hidden variable, similar to the Naive Bayes example program at the end of Lecture Notes 3, ***except*** that the first two observed variables of the data file should ***not*** be treated as independent of each other, either in training or testing.

In other words, the model should look like this when represented graphically:



You should not apply any kind of smoothing to the model.