

**CSCI 7350/6600 Homework 2 (Due: midnight Tuesday, Sept 29<sup>th</sup>)**

Please type your solution. Save your file in **PDF format** before submitting to the D2L Dropbox.

1. Given two dice, one red and one green. Each die rolls to one of six values, 1 to 6. For variable  $A = \text{red die value} - \text{green die value}$ , and variable  $B = \text{green die value}$ :
  - a. Are variables A and B independent of each other? Explain your answer.
  - b. Are variables A and B independent of each other conditioned on B? Explain your answer.
2. We have collected a data set of 14 data objects representing 14 different mushrooms. Each mushroom is labeled by a domain expert whether it is edible or poisonous. We would like to learn Bayes classifiers that will help us determine for any mushroom we may find in the future whether it is edible or poisonous. The three attributes chosen for describing the mushrooms, together with the possible values for each attribute are shown below:
  - Cap Shape: bell, flat, or convex
  - Cap Color: brown, grey
  - Odor: almond, spicy, foul

**Data:**

Object	Cap Shape	Cap color	Odor	Class
X1	bell	brown	almond	edible
X2	flat	grey	almond	edible
X3	convex	grey	spicy	poisonous
X4	bell	brown	almond	edible
X5	flat	grey	almond	edible
X6	flat	grey	spicy	edible
X7	convex	grey	almond	edible
X8	bell	brown	almond	edible
X9	convex	brown	foul	poisonous
X10	bell	brown	spicy	edible
X11	bell	grey	almond	edible
X12	convex	grey	spicy	poisonous
X13	flat	brown	almond	edible
X14	flat	grey	foul	poisonous

- a. Show the naïve Bayes classifier learned from this data  
 Use Laplace smoothing for model parameter estimation:  

$$\theta_{yi} = \frac{N_{yi} + 1}{N_y + n},$$
 where  $N_{yi} = \sum_{x \in T} x_i$  is the number of times feature  $i$  appears in a sample of class  $y$  in the training set  $T$ ,  $N_y = \sum_{i=1}^n N_{yi}$  is the total count of all features for class  $y$ , and  $n$  is the total number of data in the entire data set.
- b. Show the Bayes classifier with full joint density estimator learned from this data. Use  $10^{-6}$  for probability of data not observed in the training set. Break ties consistently.
- c. Classify the following three mushrooms using the Bayes classifiers learned from problems a and b.

**Data:**

Object	Cap Shape	Cap color	Odor	class
q1	bell	grey	almond	??
q2	convex	grey	spicy	??
q2	flat	brown	foul	??