DSC 40A - Group Work Session 5

Due Wednesday, September 4 at 11:59pm

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. **One person** from each group should submit your solutions to Gradescope and **tag all group members** so everyone gets credit.

This worksheet won't be graded on correctness, but rather on good-faith effort. Even if you don't solve any of the problems, you should include some explanation of what you thought about and discussed, so that you can get credit for spending time on the assignment.

In order to receive full credit, you must work in a group of two to four students for at least 50 minutes in your assigned discussion section. You can also self-organize a group and meet outside of discussion section for 80 percent credit. You may not do the groupwork alone.

1 Bayes' Theorem and the Law of Total Probability

Bayes' Theorem describes how to update the probability of one event given that another has occurred.

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B) \times \mathbb{P}(B)}{\mathbb{P}(A)}$$

You can think of Bayes' Theorem as a restatement of the multiplication rule. If we multiply both sides of the expression above by $\mathbb{P}(A)$, we get two equivalent expressions for $\mathbb{P}(A \cap B)$ using the multiplication rule.

Another useful rule that is helpful in many Bayes' Theorem problems is the **Law of Total Probability** which says that if we have events $E_1, E_2, \ldots, E_k$ that partition our sample space,

$$\mathbb{P}(A) = \mathbb{P}(A \cap E_1) + \mathbb{P}(A \cap E_2) + \cdots + \mathbb{P}(A \cap E_k).$$

Problem 1.

There are two boxes. Box 1 contains three red and five white balls and box 2 contains two red and five white balls. A box is chosen at random, with each box equally likely to be chosen. Then, a ball is chosen at random from this box, with each ball equally likely to be chosen. The ball turns out to be red. What is the probability that it came from box 1?

Solution: Write your solution here.

2 Independence and Conditional Independence

Recall that two events A, B are **independent** if knowledge of one event occurring does not affect the probability of the other event occurring. There are three equivalent definitions of independence:

$$\mathbb{P}(A|B) = \mathbb{P}(A) \tag{1}$$
$$\mathbb{P}(B|A) = \mathbb{P}(B) \tag{2}$$
$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B) \tag{3}$$

Two events that are not independent are also called **dependent**.

Two events A and B are **conditionally independent given** C if

$$\mathbb{P}((A \cap B)|C) = \mathbb{P}(A|C) \times \mathbb{P}(B|C).$$

Notice the similarity between this definition and the third definition of independence given above. Conditional independence given C means that *when C occurs*, A and B are independent in that case. But they may or may not be independent in general.

Problem 2.

Let A and B be events in a sample space with $0 < \mathbb{P}(A) < 1$ and $0 < \mathbb{P}(B) < 1$. If A is a subset of B, can A and B be independent? If yes, give an example, otherwise prove why not.

Solution: Write your solution here.

Problem 3.

Consider two flips of a fair coin. The sample space is S = all outcomes of 2 flips of a coin = $\{HH, HT, TH, TT\}$, where each has equal probability $\frac{1}{4}$. We define the event A as A = first flip is heads = $\{HH, HT\}$, and the event B as B = second flip is heads = $\{HH, TH\}$. You can verify that A and B are independent by showing $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$.

Now, suppose that the coin is not fair, and instead flips heads the first time with probability p and flips heads the second time with probability q.

Are A and B still independent?

Solution: Write your solution here.

Problem 4.

A box contains two coins: a regular coin and one fake two-headed coin ($\mathbb{P}(H) = 1$). Choose a coin at random and flip it twice. Define the following events.

- A: First flip is heads (H).
- B: Second flip is heads (H).
- C: Coin 1 (regular) has been selected.

Are A and B independent? Are A and B conditionally independent given C?

Prove your answers using the definitions of independence and conditional independence. Also explain your answers intuitively.

Solution: Write your solution here.

3 Naive Bayes

Naive Bayes is a classification algorithm that uses Bayes' Theorem to predict what category an object belongs to, based on certain features. In this section, we'll work out an example of Naive Bayes by hand.

Problem 5.

In parts of the world other than San Diego, the weather changes from day to day. In these places, people try to guess tomorrow's weather using the current conditions.

Weather data for 20 random days in Columbus, Ohio are recorded below, along with the next day's weather (rainy, cloudy, or sunny).

Next Day's Weather	**Humidity**	**Temperature**	**Air Pressure**
Rainy	$> 50\%$	Cool	Low
Rainy	$> 50\%$	Hot	Low
Rainy	$> 50\%$	Cool	Low
Rainy	25%-50%	Hot	High
Rainy	25%-50%	Hot	Low
Rainy	25%-50%	Cool	Low
Rainy	25%-50%	Cool	Low
Rainy	$< 25\%$	Cool	Low
Rainy	$< 25\%$	Hot	Low
Rainy	$< 25\%$	Hot	High
Cloudy	$> 50\%$	Cool	Low
Cloudy	$> 50\%$	Cool	Low
Cloudy	25%-50%	Hot	High
Cloudy	$< 25\%$	Cool	High
Cloudy	$< 25\%$	Cool	Low
Sunny	$> 50\%$	Cool	Low
Sunny	$> 50\%$	Hot	High
Sunny	$> 50\%$	Cool	High
Sunny	25%-50%	Hot	High
Sunny	$< 25\%$	Hot	High

a) Suppose that today's humidity is $> 50\%$, the temperature is hot, and the air pressure is low. Use Naive Bayes without smoothing to predict whether tomorrow will be rainy, cloudy, or sunny. Show your work.

Solution: Write your solution here.

b) Repeat the algorithm with smoothing to predict whether tomorrow will be rainy, cloudy, or sunny given that today's humidity is $> 50\%$, the temperature is hot, and the air pressure is low. Show your work.

Solution: Write your solution here.