CS 441: Discrete Structures for Computer Science

Spring 2020

Practice Worksheet for sections 7.3 and 8.5

Username (abc123): _____

1. Suppose that a test for opium use has a 2% false positive rate and a 5% false negative rate. That is, 2% of people who do not use opium test positive for opium, and 5% of opium users test negative. Furthermore, suppose that 1% of people actually use opium.

   (a) Find the probability that someone who tests negative for opium use does not use opium.

   Let T = "The person tests positive for opium" and U = "The person uses opium"

   $P(T|\bar{U}) = 0.02$, $P(\bar{T}|U) = 0.05$, and $P(U) = 0.01$

   $$\begin{aligned} P(\bar{U}|\bar{T}) &= \frac{P(\bar{T}|\bar{U})P(\bar{U})}{P(\bar{T}|\bar{U})P(\bar{U}) + P(\bar{T}|U)P(U)} \\ &= \frac{0.98 * 0.99}{0.98 * 0.99 + 0.05 * 0.01} \\ &= 0.9995 \end{aligned}$$

   (b) Find the probability that someone who tests positive for opium use actually uses opium.

   $$\begin{aligned} P(U|T) &= \frac{P(T|U)P(U)}{P(T|U)P(U) + P(T|\bar{U})P(\bar{U})} \\ &= \frac{0.95 * 0.01}{0.95 * 0.01 + 0.02 * 0.99} \\ &= 0.324 \end{aligned}$$

2. Suppose that 8% of all bicycle racers use steroids, that a bicyclist who uses steroids tests positive for steroids 96% of the time, and that a bicyclist who does not use steroids tests positive 9% of the time. What is the probability that a randomly selected bicycle user who tests positive actually uses steroids?

Let T = "The racer tests positive for steroids" and U = "The racer uses steroids".

$P(T|U) = 0.96$, $P(T|\bar{U}) = 0.09$, and $P(U) = 0.08$

$$P(U|T) = \frac{P(T|U)P(U)}{P(T|U)P(U) + P(T|\bar{U})P(\bar{U})}$$
$$= \frac{0.96 * 0.08}{0.96 * 0.08 + 0.09 * 0.92}$$
$$= 0.481$$

3. Suppose that a Bayesian spam filter is trained on a set of 10,000 spam messages and 5000 messages that are not spam.

   - The word "enhancement" appears in 1500 spam messages and 20 non-spam messages.
   - The word "herbal" appears in 800 spam messages and 200 messages that are not spam

   Estimate the probability that a received message containing both the words "enhancement" and "herbal" is spam. Will the message be rejected if the threshold for rejecting spam is 0.9?

Let S = "The message is spam", E = "The message contains 'Enhancement' ", and H = "The message contains 'Herbal' "

$$P(S|H \cap E) = \frac{P(E|S)P(H|S)}{P(E|S)P(H|S) + P(E|\bar{S})P(H|\bar{S})}$$

$P(E|S) = \frac{1500}{10000} = 0.15$

$P(E|\bar{S}) = \frac{20}{5000} = 0.004$

$P(H|S) = \frac{800}{10000} = 0.08$

$P(H|\bar{S}) = \frac{200}{5000} = 0.04$

$P(S|H \cap E) = \frac{0.15*0.08}{0.15*0.08+0.004*0.04} = 0.9868$

$0.9868 > 0.9$, so the message should be rejected as spam.

4. Find the number of elements in $A_1 \cup A_2 \cup A_3$ if there are 100 elements in each set and if

   (a) the sets are pairwise disjoint.

   $$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_2 \cap A_3| - |A_1 \cap A_3| + |A_1 \cap A_2 \cap A_3|$$
   $$|A_1 \cup A_2 \cup A_3| = 100 + 100 + 100 - 0 - 0 - 0 + 0 = 300$$

   (b) there are 50 common elements in each pair of sets and no elements in all three sets.

   $$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_2 \cap A_3| - |A_1 \cap A_3| + |A_1 \cap A_2 \cap A_3|$$
   $$|A_1 \cup A_2 \cup A_3| = 100 + 100 + 100 - 50 - 50 - 50 + 0 = 150$$

   (c) there are 50 common elements in each pair of sets and 25 elements in all three sets.

   $$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_2 \cap A_3| - |A_1 \cap A_3| + |A_1 \cap A_2 \cap A_3|$$
   $$|A_1 \cup A_2 \cup A_3| = 100 + 100 + 100 - 50 - 50 - 50 + 25 = 175$$

5. There are 2504 computer science students at a school. Of these, 1876 have taken a course in Java, 999 have taken a course in Linux, and 345 have taken a course in C. Further, 876 have taken courses in both Java and Linux, 231 have taken courses in both Linux and C, and 290 have taken courses in both Java and C. If 189 of these students have taken courses in Linux, Java, and C, how many of these 2504 students have not taken a course in any of these three programming languages?

Let:

$C =$ students who have taken C

$J =$ students who have taken Java

$L =$ students who have taken Linux

$C \cup J \cup L =$ students who have taken C or Java or Linux

$C \cap J \cap L =$ students who have taken C and Java and Linux

$$|C \cup J \cup L| = |C| + |J| + |L| - |C \cap J| - |J \cap L| - |C \cap L| + |C \cap J \cap L|$$

$$|C \cup J \cup L| = 345 + 1876 + 999 - 290 - 876 - 231 + 189 = 2012$$

$$|\bar{C} \cap \bar{J} \cap \bar{L}| = 2504 - |C \cup J \cup L| = 2504 - 2012 = 492$$

There are 492 students that have not taken a course in any of the three programming languages.