CS 6316 – *Machine Learning*
Homework 0: Probability Warmup – ***Solutions***

OBJECTIVE:

- Reviewing background on general math/probability including Bayes Theorem

INTRODUCTION:
These 10 warmup exercises test your knowledge on general math/probability. You are also tested on topics related to Bayes Theorem which was introduced in class. Carrying out this assignment should serve as a review.

EXERCISE SOLUTIONS:

Question 1:

		Reason:
(a) 0.7		
(b) 1.101	**<<<**	*the result (value) of a probability experiment cannot be > 1*
(c) (3*5) / (2*8)		
(d) 0		
(e) 1		
(f) -0.3007	**<<<**	*the result (value) of a probability experiment cannot be < 0*
(g) 9/10		
(h) 10/9	**<<<**	*the result (value) of a probability experiment cannot be >1*
(i) (2*8) / (3*5)	**<<<**	*the result (value) of a probability experiment cannot be >1*
(j) 20%	**<<<**	*the result (value) of a probability experiment should be between 0 and 1, and not a percentage. Therefore it should be 0.20 not 20%*

(Please see extended note about this on the next page – thanks!)

Extended note about Question 1:

One or two students have asked me about this, so I thought I would provide a bit of an explanation here:

The normal measure of probability is **a number between 0 and 1 inclusive**, where 0 means "impossible", 0.5 means "as likely as not" and 1 means "certain to happen". Sometimes the result of a probability experiment is expressed as a **fraction**, for example **1/4**. This is accepted because the value associated with this fraction (**0.25** in this case) is still a number between 0 and 1. *In everyday language* (non-mathematical, non-scientific), the number is more likely to be expressed as a percentage ("25% chance"), however this is not strictly the correct format of the result of a probability experiment. For instance, look at the solution to Question 6. The actual result is expressed as "0.04" however the statement in parenthesis, "*There is only a 4% chance...*" expresses this result in more everyday language (not part of the solution.)

However, for questions on the exam or results related to your project, I will accept percentages if you do something like, "... = 0.45 = 45%" or have results in a tabular format with a column headed by "Percentage."

Note: For this question this is why "20%" is incorrect (mathematically/scientifically speaking it should be a number between 0 and 1).

Question 2:
Solution:
There are a total of 6*6 combinations when rolling two dice = 36
List the outcomes that add up to 7: (1,6), (2,5), (3,4), (4,3), (5,2), (6,1)
This gives you the probability of 6/36 = **1/6**
Note: (1,6) is a separate case from (6,1). So, the total outcomes that add up to 7 will be 6 (not 3.)

Question 3:
(a)
Solution:
Let R, B, G represent red ball, blue ball, and green ball, respectively.
Sample space S :{1R1B1G, 1R2B, 1R2G, 1B2G, 2B1G, 3G}

(b)
Solution:
For example: Event E: One of the three balls is green
P(E) = **5/6**

Question 4:
(a)
Solution:
There are 27 combinations that at least one of the dice shows a number larger than 3, and 5 of them have the sum equal to 8.
P(A|C) = **5/27**

(b)
Solution:
There are 27 combinations that at least one of the dice shows a number larger than 3, and 21 of them have the sum larger than 6.
P(B|C) = **21/27**

(c)
Solution:
No. $P(A \cap C) \neq P(A) * P(C)$

Question 5:
(a)
Solution:
$P(W) = P(B \cap W) + P(L \cap W) + P(U \cap W) + P(D \cap W)$

(b)
Solution:
$P(W) = P(B|W)*P(B) + P(L|W)*P(L) + P(U|W)*P(U) + P(D|W)*P(D)$

Question 6:
Solution:
Event A: Watchman not on duty: P(A) = 2/7
Event B: Supervisor visits warehouse: P(B) = 1/7
Rule of multiplication for independent events:
P(A and B) = P(A) * P(B) = (2/7) * (1/7) = 2/49 = **0.04**
(There is only a 4% chance that the watchman will be caught not doing his job!)

Question 7:
Solution:
In a normal distribution, 68.26% of the readings lie between the mean ± 1 standard deviation (in this case, it would be 120 ± 30.) So, for this person 68.26% of the readings will fall between 90 and 150. The rest of the readings will be either above or below this range: 1 – 68.26 = 31.74% will be above or below.
If event A is the probability of blood sugar being above 150 or below 90, then
P(A) = 0.3174 (as an approximation, the answer of **0.32** will be accepted.)

Question 8:
(a)
Solution:
Find P(John | D)
Using the rule of conditional probability, $P(J \mid D) = P(J \cap D) / P(D)$
$P(J \cap D) = 20/300 = 0.0667$
$P(D) = (20+6+21) / 300 = 0.15667$
$P(J \mid D) = P(J \cap D) / P(D) = 0.0667 / 0.15667 =$ **0.4256**

(b)
Solution:
Find P(J|D)
P(J) = 100/300 = 0.333
P(A) = 60/300 = 0.2
P(E) = 140/300 = 0.467

Using rule of conditional probability,
$P(J \mid D) = P(J \cap D) / P(D)$
$= [P(D \mid J) * P(J)] / P(D)$
$= [P(D \mid J) * P(J)] / [P(D \cap J) + P(D \cap A) + P(D \cap E)]$
$= [P(D \mid J) * P(J)] / [P(D \mid J)*P(J) + P(D \mid A)*P(A) + P(D \mid E)*P(E)]$
$= (0.2 * 0.333) / (0.2*0.333 + 0.15*0.2 + 0.1*0.467)$
$= 0.066 / 0.1433 =$ **0.4648**

Question 9:
Solution:
P(AI and ML) = 0.45
P(AI) = 0.55

Conditional probability: P(A|B) = P (A ∩ B) / P (B)
Plugging in the figures: P (AI ∩ ML) / P (AI) = 0.45 / 0.55 = **0.8182**

Note: According to this question, please note that while in the whole class only 45% of the students are studying ML, but from the group that is studying AI, more than 81% of the students are studying ML

Question 10:
Solution:
Total number of pens = 10 = sample space = S = Denominator
Probability of drawing first black pen = 5/10 = 1/2
Probability of drawing second black pen = 5/10 = 1/2
Probability of drawing first white pen = 2/10 = 1/5
Probability of drawing the second white pen = 2/10 = 1/5

Probability of drawing 2 black pens and 2 white pens
= 1/2 * 1/2 * 1/5 * 1/5 = 1/100 = 0.01
