CS 3190: Foundations of Data Analysis **Due: Tuesday 9.13 at 11:50pm**

Homework 1: Probability and Bayes' Rule

Instructions: Your answers are due at 11:50pm submitted on canvas. You must turn in a pdf through canvas. I recommend using latex (`http://www.cs.utah.edu/~jeffp/teaching/latex/`, see also `http://overleaf.com`) for producing the assignment answers. If the answers are too hard to read you will lose points, entire questions may be given a 0 (e.g. **sloppy pictures with your phone's camera are not ok, but very careful ones are**)

Please make sure your name appears at the top of the page.

You may discuss the concepts with your classmates, but write up the answers entirely on your own. **Be sure to show all the work involved in deriving your answers! If you just give a final answer without explanation, you may not receive credit for that question.**

---

1. **[20 points]** Using the probability table below for the random variables X and Y, derive the following values

   (a) $\mathbf{Pr}(X \neq 0)$

   (b) $\mathbf{Pr}(X = 0 \cup Y = 0)$

   (c) $\mathbf{Pr}(Y = 1 \mid X = 1)$

   (d) Are X and Y independent? and explain why.

	$X = 0$	$X = 1$
$Y = 0$	1/10	2/10
$Y = 1$	4/10	3/10

2. **[25 points]** An "adventurous" track athlete has the following running routine every morning: He takes a bus to a random stop, then hitches a ride, and then runs all the way home. The bus, described by a random variable B, has four stops where the stops are at a distance of 5, 8, 11, and 12 miles from his house – the first three stops have probability 1/6 of occurring. The 12 mile stop has probability 1/2 of occurring. Then the random hitchhiking takes him further from his house a uniformily distributed number of miles on the distances -4 to 5; that is it is represented as a random variable H with pdf described

$$f(H = x) = \begin{cases} 1/9 & \text{if } x \in [-4, 5] \\ 0 & \text{if } x \notin [-4, 5] \end{cases}$$

   Note that a negative distance means that the runner is taken closer to his house. For example, if $H = -1$, then the runner is taken 1 mile *closer* to his home.

   What is the expected distance he jogs each morning?

3. **[30 points]** Consider a data set D with three data points $\{-1, 0, 1\}$. Assume the data has Laplacian noise defined with location M and scale 1, so from a model M a data point's

probability distribution is described by $f_M(x) = \frac{1}{2}\exp(-|M - x|)$. We want to choose M from the space $\Omega = \{-3, -1, 7\}$. Also assume we have a prior knowledge assumption on the model that $\mathbf{Pr}(M = -3) = 0.75$, $\mathbf{Pr}(M = -1) = 0.1$, and $\mathbf{Pr}(M = 7) = 0.15$. Assuming all data points in D are independent, which model is most likely?

4. **[25 points]** The Laplace Distribution, indexed by location parameter μ and scale parameter σ has probability density function given by $f(x) = \frac{1}{2\sigma}\exp(-\frac{|x-\mu|}{\sigma})$ for $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $\sigma > 0$. Plot the pdf and cdf of a Laplace random variable with $\mu = 3$ and $\sigma = 1$ for values of x is range $[-3, 9]$. The function `scipy.stats.laplace` may be useful.