Homework 1

Part A: due in class February 13, 2018

Part B: due online by evening (11:59pm) of Saturday February 9, 2018

A. Mathematics:

Compute the real value(s) of τ at which each of these functions reach a slope of 0.

1. $f(\tau) = \log(3 + (\tau + 5)^2)$

2. $f(\tau) = \frac{4\tau^k e^{-3\tau}}{k!}$ For some arbitrary integer k, assuming $\tau > 0$

3. $f(\tau) = \sum_{i=1}^{20} \frac{6\tau^3}{12i}$

4. I have a somewhat faulty alarm system. If the system is going if, there is a 30% chance there is a robber breaking into my house. At any given moment, there is a 10% chance my alarm system is going off. What is the probability that right now my alarm system is going off **and** a robber is breaking into my house?

5. You have developed a program that determines from a user's movie-watching history whether s/he is an adult or a teenager. We know 20% of users are teenagers. If user X is a teenager, the program will say so with probability 90%. If user X is an adult, the program will say s/he is an adult with probability 70%.

Assume that the program says user X is a teenager. What is the probability s/he is actually an adult?

6. Consider a Bayesian classification problem where we wish to determine if a recently-registered user of our Fordham network is a student, an administrator, or a dangerous hacker. We use three features - x_1 (VisitFrequency), x_2 (LoginLocation), x_3 (LoginDuration). Each feature takes on one of five values, shown below:

VisitFrequency	Never	Monthly	Daily		
LoginLocation	OnCampus	USA	OutsideUSA	InCity	In State
LoginDuration	MultipleWeeks	UnderAMinute	FewMinutes	FewDays	FewHours

We classify each user as either y^i=Hacker or y^i=TrustedUser. Based on a large training set, we wish to estimate all joint probability **likelihoods**, e.g.,
P(x_1=Monthly,x_2=InCity,x_3=FewDays | y=Hacker), P(x_1=Daily,x_2=USA,x_3=FewHours | y=Hacker).

a) Assuming the features **are not** independent, how many total parameters need to be estimated, accounting for classifying students, administrators, and hackers?

b) Assuming the features **are** independent, how many total parameters need to be estimated, accounting for classifying students, administrators, and hackers?

11. I have written a classifier to determine if my dog is sick or healthy. I record the sounds my dog makes once a minute for six minutes, obtaining six sound measurements s_1 , s_2 , … , s_6 . At each time, there is a likelihood my dog will make the sounds: bark, growl, no-sound, pant, whimper:

s	P(s|y=sick)	P(s|y=healthy)
Bark	0.1	0.2
Growl	0.05	0.1
No-sound	0.6	0.5
Pant	0.15	0.15

We compute P(y | s_1, …, s_6) = P(y) $\prod_j P(s_j|y)$
Note, P(y=healthy)=0.7

Provide the y=sick and y=healthy likelihoods and the max-posterior classification for each of the following sound recordings

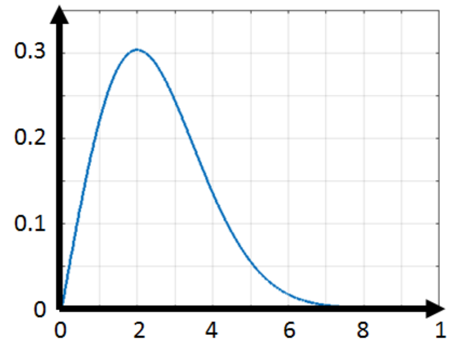
a) s_1 = Growl , s_2 = Bark , s_3 = Growl , s_4 = Whimper, s_5 = Bark, s_6 = No-sound

b) s_1 = No-sound , s_2 = Bark , s_3 = No-sound, s_4 = Bark, s_5 = Bark, s_6 = Pant

c) s_1 = Whimper, s_2 = Whimper, s_3 = Bark, s_4 = No-sound, s_5 = Bark, s_6 = No-sound

12. In lecture, we used the Gaussian probability function to express the likelihood of light-intensity conditioned on the weather being Cloudy, Eclipse, or Non-cloudy (clear skies). However, the Gaussian function allows for both positive **and negative** light intensities. Let us instead consider the Rayleigh probability as the likelihood for P(light|weather):

$$P(light|w) = \frac{light}{{\sigma_w}^2} e^{-light^2/(2{\sigma_w}^2)}$$



This function **only** is defined for light ≥ 0 . It has one parameter, σ_w, determined by the weather.

a) Assuming $\sigma_w = 3$, compute:
   - $P(light = 1|\sigma_w = 3)$
   - $P(light = 5|\sigma_w = 3)$

b) We have 500 measurements of light-intensities $light^1, light^2, \cdots, light^{500}$ during an eclipse and we wish to estimate the corresponding parameter $\sigma_{eclipse}$.

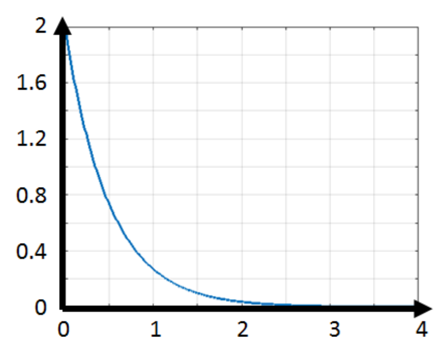
Derive the maximum likelihood estimate estimate for $\boldsymbol{\sigma_{eclipse}}$. Start with

$$P(D|\theta) = \prod_i P\left(light^i|\sigma_w\right)$$

Show at least three mathematical steps to get your estimate of $\boldsymbol{\sigma_{eclipse}}$.

c) Let us assume we have a prior probability on σ values defined by the exponential probability distribution P(σ):

$$P(\sigma) = 2e^{-2\sigma}$$



This function **only** is defined for $\sigma \geq 0$.

Derive the maximum a posteriori estimate for $\sigma_{eclipse}$. Show at least three mathematical steps to get your estimate of $\sigma_{eclipse}$

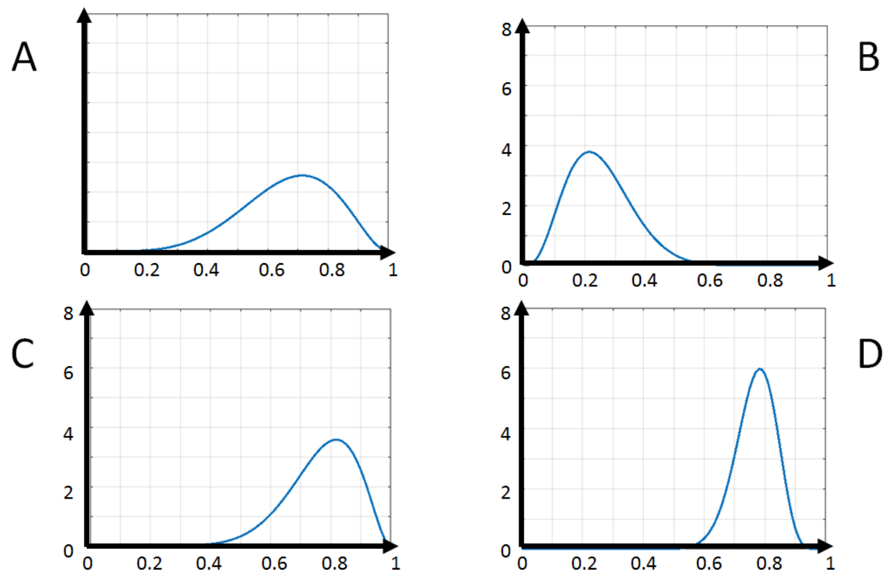
d) Actually, the Rayleigh function is not perfect for the likelihood of light intensity either, as there is probably a **maximum** possible light-intensity as well. Let us instead consider the Beta distribution as the likelihood for light intensity $P(light|\alpha_w, \beta_w)$:

$$P(light|\alpha_w, \beta_w) = \frac{light^{\alpha-1}(1-light)^{\alpha-1}}{B(\alpha,\beta)}$$

We convert the light intensity measure so that: $0 \leq light \leq 1$

What α, β combination will produce each of the likelhoods below. (All but one option fits a curve)

Option I : $\alpha = 10, \beta = 3$
Option II : $\alpha = 6, \beta = 3$
Option III : $\alpha = 4, \beta = 12$
Option IV : $\alpha = 30, \beta = 9$
Option V: $\alpha = 5, \beta = 5$



B. Programming

Detailed submission instructions: Code must be left in your `private/CIS5800/` directory. Include all function definitions and your answers to questions 1 and 7 (as comments) in the file `hw1.py` . For this homework, we will require several numpy array inputs.

Now, on to the programming:
In class, we discussed classification using Max Likelihood and using Max Posterior. For this assignment, you will create both classifiers to label online shoppers based on their purchase history. Specifically, you will determine if each shopper is a "minor" (y='M'), a "young adult" (y='Y'), a "middle-aged adult" (y='A'), or a senior citizen (y='S'). You will make this determination based on a single feature x – the number of time the shopper has ordered alcohol in the past month.

Assume a Gaussian likelihood P(amountAlcohol|shopperAge).
amountAlcohol is a real value corresponding to the number of times the shopper has ordered alcohol in the past month.
shopperAge is a member of the set of classes {Minor, Young-adult, middle-aged Adult, Senior-citizen }.

We first check if the likelihood is Gaussian. Then, we learn the parameters to best describe our likelihood function for each class. Recall the Gaussian is described by mean μ and variance σ. For convenience, we will assume a standard deviation of $\sigma = 2$. We can determine the mean by finding the average amount of alcohol for each class, i.e., for each "shopper."

Accessing our data
The file hw1data.mat is available on our website (and on erdos using `cp ~dleeds/MLpublic/hw1data.mat .`) Load this file into your Python session to get access to the trainData and testData matrices. For each matrix, each row is one example data point. The first column represents the shopper class – 1 for Minor, 2 for Teen, and 3 Adult – and the second column represents the corresponding amountAlcohol for the example data point.

Programming assignments:
1. Inspect the distribution of the amountAlcohol feature for each class and determine if it follows a Gaussian or Exponential distribution. (Note, exponential was shown earlier in this homework and Gaussian was shown in class.) Record this result as a comment in hw1.py .

You can inspect the distribution of value in a list/vector of numbers `vector` through a histogram. In Matlab, use `hist(vector)`. In Python,

```
import matplotlib.pyplot as plt
plt.hist(vector)
plt.show()
```

Regardless of our results from question 1, we will assume all distributions really are Gaussian for the rest of this assignment.

2. Write a function called **learnMean** that takes in the training set (the full trainData matrix) and a class number, and returns the learned mean amountAlcohol for that class. Specifically, the function will be called as:

```
learnMean(Data, classNum)
```

where `Data` is a numpy array with shape/size [N,2] (where N is the number of data points) and `classNum` is a single number, the function will return a single number.

```
learnMean(np.array([[1,3],[2,5],[1,9],[2,10]]) , 2)
```

would return `7.5` ((10+5)/2)

3. Write a function called **labelML** that takes in an amountAlcohol measurement and a vector containing the mean amountAlcohol values for the four shopper classes. labelML then will return the Maximum Likelihood class for the input amountAlcohol measurement. Specifically, it will be called as:

```
labelML(amountAlc, meanVector)
```

where `amountAlc` is a single number and `meanVector` is the numpy array with shape/size [1,4] and with contents $[\mu_{Minor}, \mu_{YoungAdult}, \mu_{Adult}, \mu_{Senior}]$. The function will return a single character, M, Y, A, or S, for the most likely class.

For both label questions, the function is to return the class letter (M, Y, A, or S) with highest probability.

```
labelML(6, np.array([1, 3, 8, 5]))
```

would return `S` (Senior citizen gives highest probability)

Let us assume the prior probabilities for the shopper classes are P(y=Minor)=0.3, P(y=Young-adult)=0.4, P(y=middle-aged-Adult)=0.2, P(y=Senior-citizen)=0.1 .

4. Write a function called **labelMP** that takes in an amountAlcohol measurement, a vector containing the mean amountAlcohol values for the four shopper classes, and a vector containing the shopper prior probabilities. labelMP then will return the Maximum Posterior probability (Bayes) label for the input amountAlcohol measurement. Specifically, it will be called as:

```
labelMP(amountAlc, meansVector, priorVector)
```

where `amountAlc` is a single number, `meanVector` is the numpy array with shape/size [1,4] and with contents $[\mu_{Minor}, \mu_{YoungAdult}, \mu_{Adult}, \mu_{Senior}]$, and `priorVector` is the numpy array with shape/size [1,3] and with contents $[P(y = Minor), P(Y = YoungAdult), P(Y = Adult)]$. The function will return a single character, M, Y, A, or S, for the most probable class. **For both label questions, the function is to return the class letter (M, Y, A, or S) with highest probability.**

For questions 5 and 6 below, recall the following class labels in hw1data.mat: 1 for Minor, 2 for Young-adult, 3 for middle-aged Adult, 4 for Senior citizen . The outputs of labelML and labelMP are letters instead: M, Y, A, and S.

5. Write a function called **evaluateML** that takes in all the test data (as a matrix) and the class amountAlcohol means (as a vector), and outputs the fraction of correctly-labeled data points in the test set. (The fraction will be a decimal number between 0.0 and 1.0, e.g., 0.65.) Specifically, it will be called as:

```
evaluateML(testData, meanVector)
```

where testData is a numpy array of size/shape [N,2] and `meanVector` is the numpy array with shape/size [1,4] and with contents $[\mu_{Minor}, \mu_{YoungAdult}, \mu_{Adult}, \mu_{Senior}]$. The function will return a single number between 0 and 1.

6. Write a function called **evaluateMP** that takes in all the test data (as a matrix), the class amountAlcohol means (as a vector), and the class priors (as a vector), and outputs the fraction of correctly-labeled data points in the test set. (The fraction will be a decimal number between 0.0 and 1.0, e.g., 0.65.) Specifically, it will be called as:

```
evaluateMP(testData, meanVector, priorVector)
```

where testData is a numpy array of size/shape [N,2] , `meanVector` is the numpy array with shape/size [1,4] and with contents $[\mu_{Minor}, \mu_{YoungAdult}, \mu_{Adult}, \mu_{Senior}]$, and `priorVector` is the numpy array with shape/size [1,3] and with contents $[P(y = Minor), P(Y = YoungAdult), P(Y = Adult)]$. The function will return a single number between 0 and 1.

7. Report the percent of correctly labeled test data for max likelihood and max posterior separately when means are learned:

- on the first 6 data points in the training set,
- on the first 18 data points,
- on the first 54 data points,
- on and the first 162 data points.

8. Consider shopper classification based on **both** how much alcohol they buy and how much soda they buy: $\cancel{\underset{shopperAge}{\operatorname{argmax}} \; P(amtAlcohol, amtSoda|shopperAge)}$

CORRECTION, FEB 4 11pm: MaxPosterior should compute the following argmax

$$\underset{\boldsymbol{shopperAge}}{\mathbf{argmax}} \;\; \boldsymbol{P(shopperAge|amtAlcohol, amtSoda)}$$

If you have already answered Q8 presuming argmax P(alc,soda|age) , we will accept that answer too, but that is technically max LIKELIHOOD.

`hw1dataQ8.mat` contains expanded data where the first column is subject class (1 through 4), the second column is amtAlcohol, and the third column is amtSoda.

Write a function **labelMP2** to perform Gaussian Naïve Bayes classification based on amtAlcohol and amtSoda. It will classify a single shopper based on her/his alcohol and soda purchases and based on the corresponding likelihoods and priors. Specifically, it will be called as:

```
labelMP(amountDrinks, meansMatrix, priorVector)
```

where `amountDrinks` is a numpy array with shape/size [1,2] contains [amtAlcohol, amtSoda] for a single shopper, `meansMatrix` is the numpy array with shape/size [2,4] and with contents $\begin{bmatrix} \mu_{Minor}^{alcohol} & \mu_{YoungAdult}^{alcohol} & \mu_{Adult}^{alcohol} & \mu_{Senior}^{alcohol} \\ \mu_{Minor}^{soda} & \mu_{YoungAdult}^{soda} & \mu_{Adult}^{soda} & \mu_{Senior}^{soda} \end{bmatrix}$, and `priorVector` is the numpy array with shape/size [1,3] and with contents $[P(y = Minor), P(Y = YoungAdult), P(Y = Adult)]$. **As in previous questions, we will assume all likelihood standard deviations are 2,** $\boldsymbol{\sigma = 2}$**.** The function will return a single character, M, Y, A, or S, for the most probable class. **Again, the function is to return the class letter (M, Y, A, or S) with highest probability.**

In Python, you can find dimensions of a numpy array `arr` with the statement `arr.shape()` . You can find the location of the value x in a numpy array using `where` :

```
import numpy as np
np.where(arr==N)
```

or

```
np.argwhere(arr==x)
```

If `arr=np.array([2,4,5,2,3])`, `np.where(arr==2)` would output `[0, 3]`

One last note: For your code, you may not use any built-in functions for calculating the labelML, labelMP, evaluateML, and evaluateMP functions.. You should restrict yourself to the commands covered in our class slides and the where function introduced above (in Python through numpy).