Homework 1
Due in class September 30, 2015

A. Mathematics:

Compute the real value(s) of τ at which each of these functions reach a slope of 0. For each value **in question 2 and 3 (if there are any values)**, does it identify a local maximum or a local minimum?

1. $f(\tau) = \frac{3\tau^3\alpha^4 - 36\tau\alpha^2}{4\alpha^2 - 5}$ Assume α is a real number

2. $f(\tau) = 48\log(\tau) - 2\tau^3$

3. $f(\tau) = \frac{10\tau^2 - 6\tau\rho^3}{5\tau}$

Consider a Gaussian random variable A, described by the Gaussian function P(A) with mean $\mu = 4$ and variance $\sigma = 2$.

4. What is P(A=5.0)?

5. What is P(A=3.5)?

6. A neurology patient who has a headache has a 70% chance of having a tumor. But any given patient has only a 0.5% chance of having a **headache**. What is the probability that a randomly selected patient both has both a headache and a tumor?

7. We have developed a program that uses your shopping patterns to determine if you are a vegetarian. We know 25% of the population is vegetarian. If you are a vegetarian, the program will say so with probability 70%. If you do are not a vegetarian, the program will say that you are not a vegetarian with probability 90%.

Assume that the program has said you are a vegetarian. What is the probability you actually are a vegetarian?

B. Programming

In class September 16, we discussed classification using Likelihoods (MLE) and using Posterior probabilities (MAP). For this assignment, you will pretend you are a very intelligent squirrel. While you are eating nuts on the sidewalk, you feel the ground rumble. You will write some classifiers to determine whether the ground is rumbling due to a passing pigeon, a human walking by, a truck driving by, or an earthquake.

Assume a Gaussian likelihood P(rumbleFelt|rumbleMaker).
rumbleFelt is a real value corresponding to the amount of rumbling felt by you, the squirrel.
rumbleMaker is a member of the set of classes {Pigeon, Human, Truck, Earthquake}.

Our first job is to 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 variance of $\sigma = 1$ (also known as "unit variance"). We can determine the mean by finding the average amount of rumble for each class, i.e., for each "rumbleMaker."

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 Matlab 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 rumbleMaker class – 1 for Pigeon, 2 for Human, 3 for Truck, 4 for Earthquake – and the second column represents the corresponding rumbleFelt for the example data point.

Programming assignments:
1. Write a function called **learnMean** that takes in the training set (the full trainData matrix) and a class number, and returns the learned mean rumbleFelt for that class.

2. Write a function called **labelRumbleMLE** that takes in a rumbleFelt measurement and a vector containing the mean rumbleFelt values for the four rumbleMaker classes. labelRumbleMLE then will return the Maximum Likelihood Estimate for the input rumbleFelt measurement. **For both label questions, the function is to return the class number (1, 2, 3, or 4) with highest probability.**

Let us assume the prior probabilities for the rumbleMaker classes are P(Pigeon)=0.35, P(Human)=0.5, P(Truck)=0.3, P(Earthquake)=0.05.

3. Write a function called **labelRumbleMAP** that takes in a rumbleFelt measurement, a vector containing the mean rumbleFelt values for the four rumbleMaker classes, and a vector containing the rumbleMaker prior probabilities. labelRumbleMAP then will return the Maximum A Posteriori label for the input rumbleFelt measurement. **For both label questions, the function is to return the class number (1, 2, 3, or 4) with highest probability.**

4. Write a function called **evaluateMLE** that takes in all the test data (as a matrix) and the class rumbleFelt means (as a vector), and outputs the fraction of correctly-labeled data points in the test set.

5. Write a function called **evaluateMAP** that takes in all the test data (as a matrix), the class rumbleFelt means (as a vector), and the class priors (as a vector), and outputs the fraction of correctly-labeled data points in the test set.

6. Report the percent of correctly labeled test data for MAP and MLE 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.

Note: a "vector" typically means a matrix with 1 row and multiple columns. Sometimes it means a matrix with multiple rows but only 1 column.

Quick Matlab programming advice:
There is a function in Matlab called `find` that returns the indices in the input vector/matrix fitting a certain condition. For example, if we have a vector:

```
Vect1=[1 2 3 2 1];
```

and we use find with the == equality test:

```
index2 = find(Vect1==2);
```

`index2` now will hold the values [2 4] … that locations in `Vect1` that contain the number 2.

One last note: For your code, you may not use any built-in functions for calculating MLE, MAP, etc.. You should restrict yourself to the commands covered in our class slides, in the Matlab practice assignment, and the find function introduced above.