MA 690 Project 1

Project Statement: This project will focus on the **Kumaraswamy Distribution** with parameters $a > 0$ and $b > 0$, which is the distribution defined by the probability density function

$$f(x \mid a, b) = \begin{cases} abx^{a-1}(1-x^a)^{b-1} & \text{if } x \in [0,1] \\ 0 & \text{if } x \notin [0,1]. \end{cases}$$

Provide solutions for the following problems:

1. (**By Hand, MATLAB**) Write MATLAB code which generates samples from a Kumaraswamy distribution by
   (a) Finding the cdf of the Kumaraswamy distribution
   (b) Finding the inverse of that cdf
   (c) Transforming uniform samples appropriately (as was done in a previous homework).
   (d) Checking your work by plotting the pdf of the Kumaraswamy distribution over a histogram of the generated samples.
2. (**By Hand**) Suppose that independent samples $x_1, x_2, ..., x_N$ are chosen from a Kumaraswamy distribution with *known* parameter a and unknown parameter b. Find the MLE for b.
3. (**MATLAB**) In the case that both a and b are unknown, then the maximum likelihood estimation problem gets much more difficult, because, one must simultaneously find the maximum likelihood estimate for both a and b by solving a system of equations. In the previous part, you derived a formula for b given a. Using this formula, find the MLE estimates for both a and b by proceeding as follows:
   (a) Generating a large number of samples from a Kumaraswamy distribution with some parameters a and b using your result from the first part.
   (b) Generate a vector of potential a values using the `linspace` command
   (c) For each of those potential a values, compute the associated estimate for b using the formula from the previous part.
   (d) Then calculate the likelihood for a set of samples and store the likelihood into a vector.
   (e) Use the `find` and `max` commands to find the maximum likelihood and the index at which it occurs.
   (f) Use that index to pick the value of a which corresponds to the maximum likelihood of a and use your work form Part 2) to find the MLE estimate for b.

   You will be able to tell if your script is working if you can estimate the parameters a and b well.
4. (**MATLAB**) Download the files *project1_training.mat* and *project1_testing.mat* from the CANVAS website. The file *project1_training.mat* contains samples drawn from three independent Kumaraswamy distributions. Implement a Bayesian classifier on the data by
   (a) Using your code from Part 4), find the maximum likelihood estimates for a and b for each group of samples in *project1_training.mat*. **Note:** You may assume that $0 \leq a, b \leq 15$.
   (b) Writing code that computes the likelihood of each data point in *project1_testing.mat* for each of the three groups (based upon your answers from part 1 of this problem).
   (c) Classifies the data by choosing the group which corresponds to the largest maximum likelihood computed.

   Your output should be the number of points that your classifier put the testing data into group 1, group 2, and group 3. **Note:** The file *project1_help.m* on the Canvas site will parse the input files for you.