MTH 580/680 - Applied Mathematics **Project #1**

1. Write a program the draws a *composite* Bezier curve from a list of control points. Let the input of your program be a $n \times 8$ matrix, where each row consists of the coordinates of the four control points of each piece of the curve. Include the code with your project.

   (a) Use your code to recreate the letter 'R' pictured in Figure 1.13 on p.68.

   (b) Find appropriate control points to recreate the following heart shape using a composite Bezier curve. Include the list of control points and the curve created by your program.

2. Consider the points

x	0	1	2	3	4	5	6	7
y	4.0	3.1	2.8	2.5	2.3	2.2	2.1	2.0

   (a) (*Linear Regression*) Use the method of least squares to find the line $y = mx + b$ that best fits the data.

   (b) (*Cubic Regression*) Use the method of least squares to find the cubic polynomial $y = ax^3 + bx^2 + cx + d$ that best fits the data.

   (c) (*Exponential Regression*) Use the method of least squares to find the exponential function $y = Ce^{kx}$ that best fits the data.

   (d) Plot all three functions on the same graph along with the original points.

   (e) Which function do you think best models the given data?

3. Apply the SIR model to the spread of Covid-19 in Greene County, Missouri. (https://www.springfieldmo.gov/5068/Coronavirus)

$$\frac{dS}{dt} = -aSI$$
$$\frac{dI}{dt} = aSI - bI$$
$$\frac{dr}{dt} = bI$$

   (a) Use the data found in 'coviddata.csv' to set up an overdetermined system of equations involving the infection rate a and the recovery rate b, and then use the least squares method to determine the best estimate of a and b. (You will need to find a way to approximate the derivatives.)

   (b) Use your values of a and b to model the spread of the infection. Plot the functions $I(t)$ and $R(t)$ from your model on the same graph as the original data for comparison.

   (c) Make a plot showing the long term behavior of $I(t)$ and $R(t)$.

   (d) Estimate the time when $I(t)$ will reach its maximum value. What is that value?

   (e) Approximately what percentage of the total population does your model predict will have been infected in the long run?

4. The file 'wine.csv' contains data about alcohol content and color intensity for 178 different wines of 3 different types. Use k-means clustering with $k = 3$ to see how well you can classify these wines into their types based only on their alcohol content and color intensity. Make a scatter plot of the data that is color coded by cluster. What percentage of wines are clustered correctly? (Data from: http://archive.ics.uci.edu/ml/datasets/Wine)