Assignment 6

Due: Thursday November 19, 2020 at 11:59pm
MAE 107 Computational Methods for Engineers
Professor Michael Tolley

Homework Submission

- Homework submission is through Gradescope. Please leave extra time for submission, especially if you have not used Gradescope previously.
- Assignments are graded based on correctness, how well you organize your homework (i.e. it should be easy to understand your thinking and easy to find your responses), and how well you follow the submission instructions below.
- You will not receive credit if you just give an answer. Your solution must demonstrate how you got the answer.
- If you ever think a problem is stated incorrectly, not enough information is given, or it is impossible to solve, don't panic! Simply make a reasonable assumption that will allow you to solve the problem (but clearly state what this assumption is), or indicate why it is not possible to solve the problem.
- Please attend office hours (listed in the syllabus) if you have any questions regarding the assignment.
- You are welcome to discuss the assignment with other members of the class, but everything you submit should be your own (i.e. you wrote it, typed it, or generated the plot in Matlab).

Problem 1 (25 points): Lagrange Polynomials and Spline Interpolation

A Nielson study found the following average self-reported wellbeing scores on a scale of 1-10:

Age (years)	19.5	27.5	35.5	43.5	51.5	59.5	67.5	75.5	83.5
Wellbeing	6.742	6.430	6.490	6.371	6.261	6.388	6.653	6.801	6.922

Assuming this data is accurate and precise, Prof. Tolley wants you to tell him what wellbeing score he should expect to have when he turns 40. To find this:

a) By hand, find a third-order Lagrange interpolating polynomial (choosing the base points for the best accuracy).

b) By hand, write equations for quadratic splines connecting the four nearest points (set the second derivative to zero at the first point). Express these equations as a matrix equation $Ax = b$ where x contains the unknown coefficients, and solve for the unknown coefficients using Matlab.

c) Plot both interpolation functions on top of the given data points, and report the interpolated values at $Age = 40$.

Problem 2 (25 points): Linear Regression

Imagine that you have run an experiment and obtained the data found in the file `data_testing.txt`. Load the `data_testing.txt` dataset into Matlab. The first column of the data are the x-coordinates, and the second column are the y-coordinates.

a) Using the relationships discussed in lecture, find the coefficients a_0 and a_1 for this data. Plot the data and the corresponding linear regression that you found. Note: you can either do this using the equations derived by Prof. Tolley in lecture, or using the equivalent linear algebra relationship demonstrated by Mohammad, but do not use specialized Matlab functions.

b) Using your own code (not specialized Matlab functions), determine the Standard Error of the regression S.

c) Using your own code (not specialized Matlab functions), determine the coefficient of determination of the regression r^2.

Problem 3 (25 points): Polynomial Regression

Fitting an experimental data set to a least squares approximation can be useful to provide a physical understanding of the measured data. When testing higher order polynomial fits, we need to use caution to avoid “overfitting”.

a) For the same data from Problem 2, fine least squares approximation of the data for linear, quadratic, cubic, quartic, and 5th order polynomials. Plot all the fits in one figure. For this problem, you can use specialized Matlab functions like `polyfit` and `polyval`.

b) Determine the standard error S, and coefficient of determination r^2 of each regression.

c) Based on this data alone, which is the best fit line for this set of data? Why?

d) Now, imagine that you run a second (validation) set of experiments and obtain the data contained in `data_validation.txt`. Load the `data_validation.txt` dataset into Matlab. The first column of the data are the x-coordinates, and the second column are the y-coordinates. Plot this data and determine the standard error S, and coefficient of determination r^2 for the fits generated in part (a).

e) Based on this new information, which regression do you think best describes the experimental results? Explain why your answer is the same as, or different from, your answer in part (c).

Problem 4 (25 points): Regression with Linearized Data

You are designing a manufacturing process, and measure the cooling of a part after a key step in the process:

t (min)	0	1	2	3	8	13	18	23	28	33	38	43	48	53
T ($^\circ C$)	86	82	79	77	69	62	57	53	49	47	44	42	40	39

You decide to model this cooling process as an exponential decay, of the form:

$$T(t) = T_0 e^{-Ct}$$

a) Write the equation for the error e of the exponential decay approximation.

b) Write the two linear equations that can be solved for the coefficient values that minimize the error e. (Hint: You will need to make variable substitutions to obtain linear forms of these equations.)

c) Write the equations from part (b) as a matrix equation of the form $Ax = b$.

d) Solve in MATLAB to find the coefficient values for a least-squared error fit.

e) Plot the data points along with the approximation fitting curve using the calculated coefficient values from part (c) on two different plots with different scaling for the T axis:

   i) t and T

   ii) t and $\ln(T)$