AEM 3103 Exam #1 - Spring 2024

Wednesday, February 28

Name:

A single 8.5 x 11 inch sheet of notes (both sides) and a non-communicating calculator is allowed. Please write answers on both sides of the exam page. If needed, additional pages can be used and attached to the submission. Please write your name on all sheets submitted.

1. (20 points) Two weeks into an aerodynamics project, a student has the file structure (shown below) on their computer.

   (a) Propose a new file structure that organizes all project-related files into the `AeroLab-2024` folder. This folder can have sub-directories as well. Draw out the proposed folder structure clearly listing all folders/files (i.e. similar to the format shown).

   (b) The `load_and_plot.m` has a hard-coded path to the data `lab-data-1.csv` file. Given your proposed folder structure, what should the new path be? Express it in two ways (either a Windows or MacOS/Linux file path is acceptable).

      i. An absolute path: (*)\data\lab-data-1.csv

      ii. A relative path (relative to the new home of `load_and_plot.m`): $\underbrace{..}_{\text{parent directory}}$\data\lab-data-1.csv

   (c) The project is expected to take 2 months and so it seems prudent to start using version control. On your proposed folder structure, clearly mark the directory where a Git repository, for this particular project, should be initialized (i.e. where should they run `git init`?).

c) run "git init" here.

a) One Example:

(*) C:\Users\Student\Documents\Coursework\AeroLab-202

```
C:\Users\Student\
 └─Documents
    └─Coursework
       └─AeroLab-2024
          ├─load_and_plot.m
          ├─curve_fit.m
          └─compute_aero_params.m
 └─Desktop
    ├─Lab-Writeup.docx
    └─Lab-Analysis-Readme.txt
 └─Downloads
    ├─lab-data-1.csv
    └─lab-data-2.csv
```

```
├─docs
│  └─Lab-Writeup.docx
├─data
│  ├─lab-data-1.csv
│  └─lab-data-2.csv
├─scripts
│  ├─load_and_plot.m
│  ├─curve_fit.m
│  └─compute_aero_params.m
└─Lab-Analysis-Readme.txt
```

Clear folder names separating data, scripts, and docs

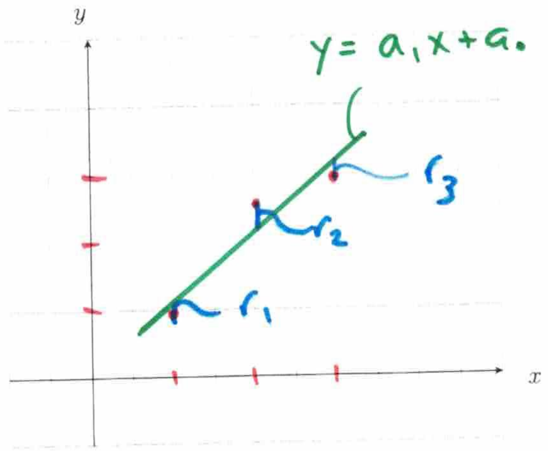
2. (25 points) Consider the following arrays of data: $\bar{x} = [1, 2, 3]$ and $\bar{y} = [1, 2.5, 3]$. Assume we want to find a linear model, $y(x) = a_1 \cdot x + a_0$, to curve-fit this data.

   (a) Write the linear least-squares cost-function for this scenario. This function is minimized to solve for the coefficients (a_1, a_0). Write the function by following these steps:

      i. Write out the residual expression for a generic data point (e.g. $r_i(x_i, y_i, a_1, a_0)$).

      ii. Plug in the data then square and sum the residuals.

   (b) Use a diagram and simple terms to explain the concept of *least-squares* and why the name is appropriate.



b) The goal is to pick the coeffts for the model (e.g a_1, a_0) that minimize the sum of the residuals squared.

Hence the name "least-squares"

   (c) Given the expression in 2(a)ii (e.g. $E(\bar{x}, \bar{y}, a_1, a_0)$), explain in simple terms how to solve for the choice of (a_1, a_0) which minimizes this function.

   (d) Given the array definitions `xdata` and `ydata`, write out the MATLAB code to:

      i. Curve-fit this data to a 1st-order polynomial (i.e. linear model).

      ii. Apply the model to `xdata` and store the result as `yfit`.

   If unsure, rough syntax or pseudocode is acceptable.

a) i.) $r_i = y_i - (a_1 x_i + a_0)$

a) ii) $E = \sum_{i=1}^{3} r_i^2$

$$= (1 - (a_1 \cdot (1) + a_0))^2 + (2.5 - (a_1(2) + a_0))^2 + (3 - (a_1(3) + a_0))^2$$

c) Take partial derivatives w.r.t each coeff, set to zero, & solve the system of equations.

$\frac{\partial E}{\partial a_1} = 0$ $\quad \frac{\partial E}{\partial a_0} = 0$ $\Rightarrow$ solve for a_1, a_0

This choice minimizes E.

d)
```
>> p = polyfit(xdata, ydata, 1);
>> yfit = polyval(p, xdata);
```

3. (25 points) Interpolation is a procedure which returns the exact value where data is available, and estimated values between data points. Consider the following arrays of data: $\bar{x} = [0, 1, 2]$ and $\bar{y} = [1, 2, 3]$.

The general equation for a Lagrange polynomial with n data points is:

$$f(x) = \sum_{i=1}^{n} y_i \prod_{\substack{j=1 \\ j \neq i}} \frac{(x - x_j)}{(x_i - x_j)}$$

(a) Write out the Lagrange polynomial $f(x)$ for this data set.

(b) Verify that $f(x_i) = y_i$ for all three data points.

(c) Interpolate the data by finding the value of at $x = 0.5$ (i.e $f(0.5)$).

(d) A future experiment will have 51 pairs of data points. A friend tells you to interpolate the data using a single 50^{th} order polynomial (it is true that a m^{th} order polynomial has $m + 1$ unknowns). Thinking back to AEM 3103, is this a good idea? Why or why not?

(e) Other factors make it desirable to use some combination of standard polynomials for the future experiment and not Lagrange or Netwon's polynomials. What method of interpolation should you recommend? Describe how it works in simple words.

a) $f(x) = 1 \frac{(x-1)}{(0-1)} \frac{(x-2)}{(0-2)} + 2 \frac{(x-0)}{(1-0)} \frac{(x-2)}{(1-2)} + 3 \frac{(x-0)}{(2-0)} \frac{(x-1)}{(2-1)}$

$i=1 \quad j=2 \quad j=3 \qquad i=2 \quad j=1 \quad j=3 \qquad i=3 \quad j=1 \quad j=2$

b) $f(0) = \frac{2}{2} = 1$ ✓

$f(1) = (-2)(1)(1-2) = 2$ ✓

$f(2) = \frac{3}{2}(2)(2-1) = 3$ ✓

c) $f(0.5) = \frac{3}{2} = 1.5$

↑ plug int (a)

d) No. Higher-order polynomials have numerical issues & are unreliable between data points.

e) Use a collection of piece-wise low-order polynomials, also known as a spline.

4. (20 points) The arrays of data $\bar{x} = [x_1, x_2..., x_N]$ and $\bar{y} = [y_1, y_2..., y_N]$ have been collected as part of an experiment (of length N). Two data model candidates are being considered for a curve-fitting task.

(a) $y(x) = c_3 \cdot x^3 + c_2 \cdot x^2 + c_1 \cdot x + c_0$

(b) $y(x) = c_1 \cdot x^{c_0}$

(typo)

Model (a) has four unknown coefficients (c_3, c_2, c_1, c_0) and model (b) has two unknown coefficients $(\cancel{c_2}\ c_1, c_0)$. Assume $N >> 4$ (i.e. much greater than 4).

(a) Can linear curve-fitting techniques like linear least-squares be applied to model (a)? Why or why not?

(b) Model (b) has a non-linearity in the coefficients. Apply a transformation to the equation so as to put it in a linear form. Define the new unknown coefficients in terms of c_1 and c_0.

a) Yes, the polynomial is linear in the unknowns.

$$y_i = \underbrace{[\, x_i^3 \quad x_i^2 \quad x_i \quad 1 \,]}_{\text{Known \#'s.}} \begin{bmatrix} c_3 \\ c_2 \\ c_1 \\ c_0 \end{bmatrix}$$

b) $y(x) = c_1 x^{c_0}$ is non-linear. Take $\ln()$ of both sides.

$$\ln(y) = \ln(c_1 x^{c_0})$$

$$\ln(y) = \ln(c_1) + \ln(x^{c_0})$$

$$\underbrace{\ln(y)}_{\tilde{Y}} = \underbrace{\ln(c_1)}_{\tilde{C}_0} + \underbrace{c_0 \ln(x)}_{\tilde{C}_1 \tilde{X}}$$

$$\tilde{Y} = \tilde{C}_0 + \tilde{C}_1 \tilde{X}$$

5. (10 points) For the following list of properties, circle the option that best matches the statement.

A. Array indices begin with 1.	MATLAB	Python	both	neither
B. Can be downloaded and installed without paying for a commercial license.	MATLAB	Python	both	neither
C. Is commonly used for CAD drawings of parts and assemblies.	MATLAB	Python	both	neither
D. Can be used for engineering and data analysis tasks.	MATLAB	Python	both	neither
E. Use the element-by-element operator (`x.^2`) to square every element of an array `x`.	MATLAB	Python	both	neither

The rest of this page, and the back, can be used to show work for other problems. If used, clearly label them as such.