Lab 10: Intermediate Curve Fitting

10.1 Introduction

This lab focuses on generating polynomial, general linear, linearized, and nonlinear fits to different data sets as well as how to identify likely nonlinear models for data. For some of the problems you will be presenting graphical representations of the models as well as calculating statistical information to quantify the goodness of fit. For some of the problems, you will also be making predictions based on the models. Note that in most cases, you should figure out how to apply extant code in the book or on Pundit to novel problems.

10.2 Resources

The additional resources required for this assignment include:

- Books: Chapra and Palm
- Pratt Pundit Pages: `MATLAB:Fitting`, `MATLAB:Plotting` (specifically the `General Plotting Tips`)

10.3 Getting Started

1. Log into one of the PCs in the lab using your NET ID. Be sure it is set to log on to acpub.
2. Start a browser and point it to `http://pundit.pratt.duke.edu/wiki/Lab:B209`. Check out the `Using the PCs to Run MATLAB Remotely` section for how to get connected and make sure the connection is working.
3. Once connected to a machine you believe will also display graphics, switch into your `EGR53` directory and create a `lab10` directory inside it:

   ```
   cd EGR53
   mkdir lab10
   ```

4. Switch to your ~/`EGR53/lab10` directory:

   ```
   cd lab10
   ```

5. Copy all relevant files from Dr. G's public `lab10` directory:

   ```
   cp -i ~mrg/public/EGR53/lab10/* .
   ```

   Do not forget the space and the "." at the end.
6. Open MATLAB by typing `matlab &` at the prompt that appears in your terminal window. It will take MATLAB a few seconds to start up.

10.4 Assignment

10.4.1 Based on Palm 5.38, p. 350

For this problem, instead of just using a quadratic polynomial, you will solve and plot linear, quadratic, and cubic models. In addition to finding the model coefficients and the statistical information, you will determine the value of the stopping distance predicted by each model for a car going 55 mph and for a car going 95 mph. Fill out the following table:

Order	Equation ($d = f(v)$)	S_t	S_r	r^2	$d(55)$ (ft)	$d(95)$ (ft)
1						
2						
3						

For the plot, make sure to use different line styles for the models and asterisks for the original data. Include a legend to indicate which is which. Note that you do not need to plot the estimates for any of the models. Plot the models for speeds between 20 and 70 miles per hour.

In the lab report, state which model you believe works best and how much confidence you have in each of the predictions of stopping distance for 55 mph and for 95 mph. Be sure to indicate the reason for your confidence or lack thereof in each model.

Coding hints: Take the example code for a polynomial fitting and figure out how to process three different models simultaneously. Think about which variables are applicable to all three models and which ones need to be differentiated from each other.

10.4.2 Based on Chapra 14.10, pp. 332-333

For this problem, you will use general linear least-squares to determine the initial concentrations of three different organisms in a sample of seaweed. The model is made up of three exponentials with different decay rates. To solve, you should:

- Determine the coefficients of the equation above using general linear regression.
- Create a `linspace` with 100 times evenly spaced between the values given in the table, then generate a `plot` with the original data as black asterisks and the model as a solid black line. Be sure to include proper labels, a legend, and a title. Note that you do not need to plot the estimates.
- Create another plot that shows the models of each individual population. That is make a plot that has three lines on it - one showing the population of type "A" organisms, one showing type "B," and one showing type "C." Be sure to include proper labels, a legend, and a title.
- Typeset the model you found for $p(t)$ in your report; you may assume that p, A, B, and C all have units of organisms per cubic centimeter, so that you do not need to put units in the equation. You should put the units on the axis labels, however. Be sure to put the terms in the same order (A, then B, then C).
- Determine the S_t, S_r, and r^2 values, then state whether this is a good fit and why you believe that.

Coding hint: Take the example code for a general linear model and adjust it to this particular model.

10.4.3 Chapra 14.10 Alternate

Assume one of your colleagues suggests that the "Type B" organism actually has a slower decay rate, so that the the general population model is actually:

$$p(t) = Ae^{-1.5t} + Be^{-0.1t} + Ce^{-0.05t}$$

Run the same analysis as the problem above, including making the two graphs and performing the statistical analysis. Is this a better *mathematical* fit? Is this a better *scientific* fit? Defend each answer using the data and graphs available to you. Note that the code for this problem should be nearly exactly the same as that above so it isn't truly a whole new problem.

Coding hint: You will probably just do a `Save As...` on the program from the previous problem. Just be sure to change the names of the files into which the graphs go *before* running the new version. *Note:* You may either choose to do this and the previous problem in two separate scripts or in one script keeping track of the two different models.

10.4.4 Palm 5.40, p. 350

Now you can actually do Palm 5.40 the way it was meant to be done. Use multiple linear regression to determine the coefficients of the plane and typeset the equation in your lab report. In addition to finding the maximum percent error in the predictions, you should report the y, x_1, x_2, and $\hat{y}$ value at which that maximum percent error occurs. Finally, calculate the S_t, S_r, and r^2 value and state whether the fit is mathematically good. You do *not* need to make a plot for this problem.

Coding hints: Start with the example code for the general linear model, but be sure to note that your equation now has two variables in it. Turns out, this will not have an impact on the `A` matrix. Also - remember that for linear fits, the data *must* be in columns. So for this one, turns out the form in which the values were given in the problem is useful.

10.4.5 Based on Chapra 14.14, p. 333

Using the data given in Chapra's Problem 14.14:

(1) Make a plot similar to *bottom half* of the one on p. 301 of the Chapra book - that is, make three subplots, where each uses some transformation on the original data. From these plots determine if it is more likely that the data will fit an exponential model, a power model, or a saturation growth model. Support your argument. You should title your plots - be sure to indicate the linearization model used as well as which data set is being used. Also be sure to label the axes. For these graphs, you should connect the data points with lines to help see the pattern - use black asterisks for the data points and a solid black line to connect them.

(2) Solve for the coefficients of a linear fit, a quadratic fit, an exponential fit, a power-law fit, and a saturation growth model. Note that you do *not* need to make plots of these models.

(3) Determine the S_t for the data set and the S_r and r^2 values for each of the models. Which model is the best mathematical fit?

In your lab report, include the following table:

Fit	Equation	S_t	S_r	r^2
Linear				
Quadratic				
Exponential				
Power Law				
Saturation Growth				

Coding hints: Pick one of the example scripts to serve as a foundation for this problem - probably the linearized example since that is the more complicated code - then think about how to keep track of *five* different fits simultaneously. A further complication is that these five models come from two different categories - two are polynomials and three are linearized. Plan out your program before you start typing!

10.4.6 Chapra 14.11, p. 333

Use *nonlinear regression* to solve for the maximum photosynthesis rate P_m and the optimal solar radiation I_{sat}. Note that you will need to be sure to give the `fminsearch` function good starting guesses for these values to obtain realistic results. Once you have the coefficients,

(1) Typeset the equation in your lab report. You can either substitute the numbers into the equation or report the symbolic version of the equation followed by numerical values for the variables you found. For example, if you determine that P_m is 8 mg m^{-3} d^{-1} and I_{sat} is 12 μE m^{-2} s^{-1} you can either report:

$$P(I) = 8\frac{I}{12}e^{-\frac{I}{12}+1}$$

or you could state

$$P(I) = P_m\frac{I}{I_{sat}}e^{-\frac{I}{I_{sat}}+1}$$

with P_m= 8 mg m^{-3} d^{-1} and I_{sat}= 12 μE m^{-2} s^{-1}.

(2) Describe how you decided what to use for an initial guess.

(3) Make a plot with the original data as black asterisks and the model equation as a black line. Make sure it is a proper graph.

(4) Calculate the statistical measures to determine goodness of fit and state your assessment of the model.

Coding hints: Start with the nonlinear regression example. Carefully note the *names* of the variables you are trying to find and figure out how you might make a reasonable first-guess as to what each might be. *Note:* `[0 0]` is a particularly poor guess.