Tutorial 6: Linear-Least-Squares Fitting

Joe Zalesky

October 17th, 2014

A linear least squares approximation takes a set of data and fits a function with linear terms to it. In this case, we'll be making a function that fits all of the data to a line. In this homework, you'll be making a bunch of small functions which all work together to do the approximation. Because of this, its convenient to put all of your functions on a single .pro file so you only have to compile once. Be sure that everything has proper comments and indentation for full credit. Your linear least squares function will take an arbitrary set of x values and corresponding y values and perform a linear regression to the dataset, producing a plot demonstrating the regression.

1 Linear Algebra and LLS

For a set of data points with x and y values a linear relationship can be expressed as,

$$y = mx + b = x' \cdot A \tag{1}$$

where x' is a column vector,

$$x' = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{pmatrix}$$

and A is the column vector,

$$A = \begin{pmatrix} m \\ b \end{pmatrix}$$

where m will be the slope of our line of best fit and b is the y-intercept of the same fit. Great so how do we calculate A from a bunch of data? We can begin by making x square via,

$$x'^T \cdot y = x'^T \cdot x' \cdot A \tag{2}$$

where x'^T is just the transpose of x'. From here we see A can be calculated by multiplying both sides of Equation 2 by the inverse of the square matrix we just made,

$$A = \left(x'^T \cdot x'\right)^{-1} \cdot x'^T \cdot y \tag{3}$$

Thus the fitted y values can be stated as,

$$y' = x' \cdot A \tag{4}$$

2 Preforming a Linear Fit in IDL

Hopefully the mathematical walk-through provided above will be helpful for those who have not had much practice with matrices / have not taken Math 54 yet. In order to practice our skills in modular programming we ask that you please complete this assignment by writing multiple functions, each of which will solve a small part of the overall problem. The functions/procedure you will need to write are as follows:

- **x_vector**: This function will take in your x values of the data and output the vector x'.
- **find_coeffs**: This function will take the x' vector you made and y values of the data and output the A vector
- **regress**: This function will take your x' and A vectors and output the fitted y-values y'.
- **main**: As stated before this is your procedure that calls the above helper-functions and uses them in sequence to make a pretty plot. First read in your data (what function does this for you?) into some variables and produce a plot showing both a scatter plot of the data and your line of best fit using the above three functions. Make sure your plot has proper axis labels and that the title is the line of best fit you found (I.E y=mx+b where m and b are the fitted values you found).

Once your code is running and making nice plots make sure to save a copy of your plot (recall how to save plots in IDL from last week) in your directory along with your code. Push everything up and you're done!