ESSAY PROBLEM, HOMEWORK 6

POLYNOMIAL FIT FOR A BIG DATA SET

For this problem, your task is to find some data (any data) and use a least squares fit to find the best fit line. You should use at least 15 data points, and your write up should include the following:

(1) The equation of the line.
(2) A graph of the data points with the axes labeled and the best fit line drawn.
(3) A paragraph explaining what data you used, what steps you used to determine the answer (including what program you used to do the linear algebra), and what the equation of the line tells you about your data.

You can work with others on this problem, but you should each use your own data set. You can get your data from any source (for example, x=state population and y=number of representatives in the House) or create your own (ask a bunch of friends how much time they spend each week doing homework and let x=age y=number of hours spent doing homework). Of course, try to pick two variables that have a chance at being correlated. You do not have to show your work or turn in your actual computer code.

Doing the problem in MATLAB. I will use last quarter's grades to illustrate how the process would go in MATLAB. I am going to put students' grades to the first midterm on the x-axis, and their grades on the second midterm on the y-axis.

Define the data:

```
>>x=[78; 84; 100; 90; 74; 92; 82; 88; 98; 90; 82; 78; 82; 72; ...
>>y=[76; 84; 98; 82; 50; 98; 50; 98; 94; 84; 100; 98; 84; 96; ...
```

Form the matrix A with ones in the first column and the x values in the second:

```
>>A=[ones(size(x)) x]
```

Reduce the matrix $[A^TA \quad A^T\vec{\mathbf{y}}]$ (Use an apostrophe to transpose in MATLAB)

```
>>rref([A'*A A'*y])
```

This gives you the answer. In our case, the equation for the line is $y = -6.4913 + 1.0505x$. Now we will plot the data and the line. MATLAB only works with matrices, so we need to set up vectors xx and yy corresponding to the line. This command makes xx go from 50 to 100 by steps of 1. Depending on your data, you will want to use different values. The semicolon is not necessary, but supresses the output.

```
>>xx=50:1:100;
>>yy=-6.4913+1.0505*xx;
```

Now we plot it with the following command:

```
>>plot(x,y,'o',xx,yy,'-')
```

The first three arguments draw the data that I stored in the variables x and y, marked by circles. The second three arguments draw the line, which is stored in xx and yy, and makes it look like a line graph. Here is my result: