

Math 2270-2 Spring 2012  
Computer Lab 3  
Least Squares Approximation

This lab was written by Professor Chris Cashen in his Postdoc years at Utah. Some minor display, maple code display, and font size changes were made. It remains exactly the same as Chris Cashen wrote.

**Submit your project** on the due date in class as a worksheet print. **Submit study group** efforts as one worksheet print with multiple names. See Section 4.3 of the text for problem background.

## 1 Introduction

We have done several examples of using least squares approximations to find the best fit line through a collection of point in the plane.

Here is a quick refresher, along with some Maple code to visualize.

Consider the three points  $(0, 1)$ ,  $(1, 2.1)$ ,  $(2, 3.9)$ . If these points lie on a line  $y = mx + b$  then

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 2.1 \\ 3.9 \end{pmatrix}$$

These points, however, do not all lie on a common line. To find the best fit line using the least squares approximation method, multiply both sides of the equation by  $A^T$  and solve.

$$\begin{aligned} A^T A \begin{pmatrix} m \\ b \end{pmatrix} &= A^T \begin{pmatrix} 1 \\ 2.1 \\ 3.9 \end{pmatrix} \\ \begin{pmatrix} m \\ b \end{pmatrix} &= (A^T A)^{-1} A^T \begin{pmatrix} 1 \\ 2.1 \\ 3.9 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 8.7 \\ 5.3 \end{pmatrix} \end{aligned}$$

The best fit line is  $y = \frac{8.7}{6}x + \frac{5.3}{6}$ .

Plot this in Maple with the following commands:

```
with(LinearAlgebra): # It will be used in this session.
Points:= [ [0, 1], [1, 2.1], [2,3.9] ];
y:=(8.7/6)x+5.3/6;
plot([Points,y], x=-1..3,color=[red,blue],style=[point,line]);
```

## 2 Best fit line

Let's try this again with a larger data set. Get the three files

DataSet1.txt, DataSet2.txt, DataSet3.txt

from the course web site and save them in your home directory (or somewhere else that you can find them). In Maple enter

```
datadir := "path to the data files";  
myfile := cat(datadir, "DataSet1");  
DataSet1 := readdata(myfile, 2);
```

Take a look at this set of points by plotting them:

```
plot(DataSet1, x=0..10, color=red, style = point);
```

You should see a sequence of points that look approximately linear. Find the best fit line by least squares approximation.

At present, the data is in a list of lists. Make it a matrix by  $M := \text{Matrix}(\text{DataSet1})$ ;

However,  $M$  is not the matrix that we usually call  $A$ . Column 1 of  $M$  is column 1 of  $A$ , but column 2 of  $A$  should be all 1's, while column 2 of  $M$  is what we usually call  $\vec{b}$ .

We can pick out a column of  $M$  by multiplying by a vector.  $M \cdot \langle 1, 0 \rangle$  is the first column of  $M$ , and  $M \cdot \langle 0, 1 \rangle$  is the second column of  $M$ .

Do the following steps to get our usual setup (specify  $n$ , it should be the number of rows of  $M$ ):

```
b := M .<0, 1>;  
A := < M .<1, 0> | Vector(column, n, 1) >;
```

Now we can find the slope and  $y$ -intercept of the best fit line:  $(A^+ \cdot A)^{-1} \cdot A^+ \cdot b$ .

Plot the points of the data set along with the best fit line.

### 3 Parabola

Repeat these steps for DataSet2 (but use names other than  $M$ ,  $A$  and  $b$ , maybe  $M2$ ,  $A2$  and  $b2$ ). Notice when the points are first plotted they do not look approximately linear, they look parabolic. Find the best fit parabola using the least squares approximation.

This time the matrix  $A2$  should have all 1's in the third column, the  $x$ -coordinates of the points in the second column, and the squares of the  $x$ -coordinates in the first column.

Here is a Maple command that achieves this easily (you need to specify  $n$ ):

```
A2 := < map(x -> x^2, M2 .<1, 0>) | M2 .<1, 0> | Vector(column, n, 1) >;
```

This syntax will be useful again, so let's understand what it says. It defines  $A2$  to be a three column matrix. The third column is a column vector of dimension  $n$  whose entries are all 1. The second column is the column vector which results from multiplying the matrix  $M2$  by the column vector  $\langle 1, 0 \rangle$ , which picks out the first column of  $M2$ . The first column is the one which is new, it uses the *map* function, which applies an operation to all the entries of the input. In this case, the input is the column vector  $M2 \cdot \langle 1, 0 \rangle$ , the first column of  $M2$ . The operation is **square each entry** by applying the function  $x \rightarrow x^2$ .

### 4 Logarithm

Now try DataSet3, which is the height/weight data. This data should satisfy a power rule:  $\text{weight} = k \cdot \text{height}^n$  for some constants  $k$  and  $n$ . Take the logarithm of each side to make this a linear equation. Find the best fit line for the log data. Plot the data set along with the function  $k \cdot x^n$  for the appropriate  $k$  and  $n$ .

According to this model, what would be the weight of a person whose height is 5 feet 10 inches?