Mth 351 Least Squares

Mth 351 Summer 2002

July 28 2002 Maple 5 and 6

Bent E. Petersen

Filename: 351u2002_least_squares.mws

```
> restart;
```

Maple has support for least squares fitting in the stats[fit] package and also in the linalg package. The stats[fit] package is the more convenient one to use for fitting data to a given equation.

```
> with(stats[fit]):
```

It will also be useful to have some nice plot commands available:

```
> with(plots):
Warning, the name changecoords has been redefined
```

We will need some random data to illustrate the commands.

```
> Xdata:=[seq(k/2,k=0..20)];
```

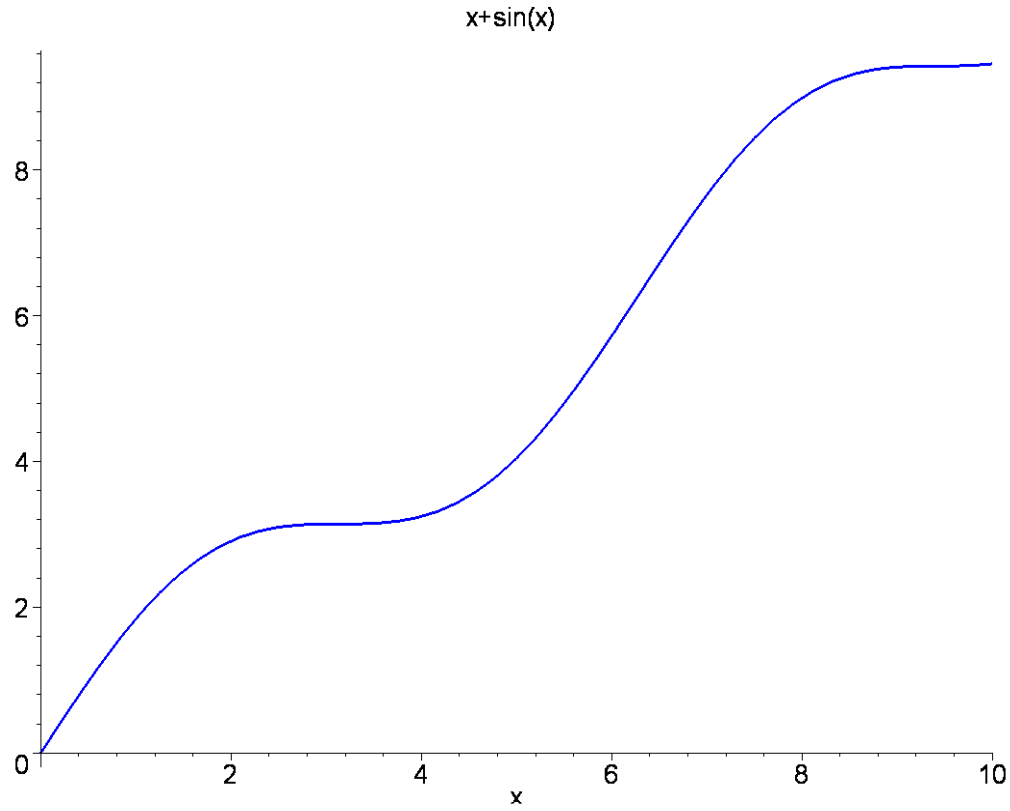
$$Xdata := \left[0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, 3, \frac{7}{2}, 4, \frac{9}{2}, 5, \frac{11}{2}, 6, \frac{13}{2}, 7, \frac{15}{2}, 8, \frac{17}{2}, 9, \frac{19}{2}, 10\right]$$

```
> Ydata:=map(x->x+sin(x),evalf(Xdata));
```

Ydata := [0., .9794255386, 1.841470985, 2.497494987, 2.909297427, 3.098472144, 3.141120008, 3.149216772, 3.243197505, 3.522469882, 4.041075725, 4.794459674, 5.720584502, 6.715119988, 7.656986599, 8.437999977, 8.989358247, 9.298487113, 9.412118485, 9.424848880, 9.455978889]

These data points all lie on the graph of x + sin(x). Here's a plot of the graph for comparison with the least squares fits that we obtain below.

```
> plot(x+sin(x),x=0..10,title="x+sin(x)",thickness=3,color=blue);
```



Let's try a leastsquares line fit:

```
> eqn1:=y=m*x+b;
```

$$eqn1 := y = m\,x + b$$

```
> lsq1:=leastsquare[[x,y],eqn1,{m,b}]([Xdata,Ydata]);
```

$$lsq1 := y = .9789783571\,x + .2636407548$$

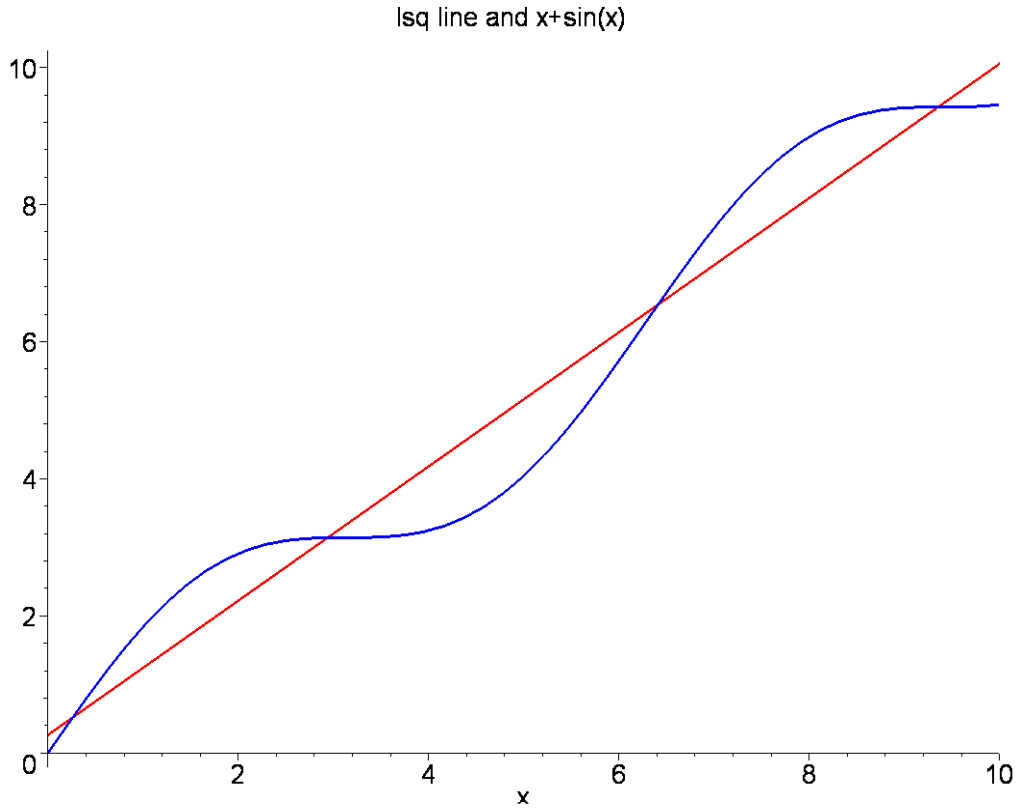
Note [x,y] tells Maple the names of the variables in eqn1 and {m,b} tells Maple which parameters to adjust. If you want to leave b as an undetermined parameter you could do it as foolws:

```
> lsq1b:=leastsquare[[x,y],eqn1,{m}]([Xdata,Ydata]);
```

$$lsq1b := y = (-.1463414634\,b + 1.017559931)\,x + b$$

To plot the solution lsq1 we note we just want to plot the right-hand side of the equation lsq1. Let's plot it together with x + sin(x) so we can compare them.

```
> plot([x+sin(x),rhs(lsq1)],x=0..10,color=[blue,red],thickness=3,tit
  le="lsq line and x+sin(x)");
```

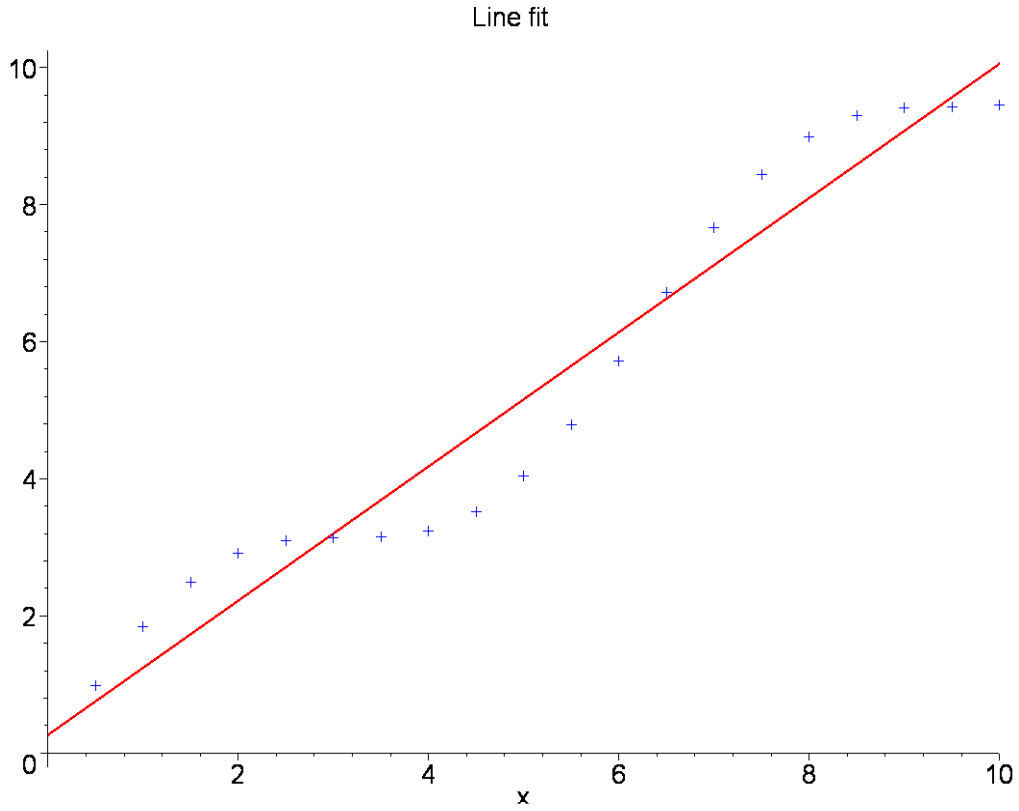


Of course it makes more sense to compare our least squares fit with the data points. First we need the Cartesian coordinates corresponding to our data:

```
> L:=[]: for k from 1 to nops(Xdata) do
  L:=[op(L),[Xdata[k],Ydata[k]]]: od:
> evalf(L,4);
```

[[0., 0.], [.5000, .9794], [1., 1.841], [1.500, 2.497], [2., 2.909], [2.500, 3.098], [3., 3.141], [3.500, 3.149], [4., 3.243], [4.500, 3.522], [5., 4.041], [5.500, 4.794], [6., 5.721], [6.500, 6.715], [7., 7.657], [7.500, 8.438], [8., 8.989], [8.500, 9.298], [9., 9.412], [9.500, 9.425], [10., 9.456]]

```
> img0:=PLOT(POINTS(op(L)),SYMBOL(CROSS,20),COLOR(RGB,0,0,1)):
> img1:=plot(rhs(lsq1),x=0..10,color=red,thickness=3,title="Line
  fit"):
> display([img0,img1]);
```



Looking at the data you may feel a polynomial of degree 4 ought to fit better. Let's try it:

```
> eqn2:=y=a+b*x+c*x^2+d*x^3+e*x^4;
```

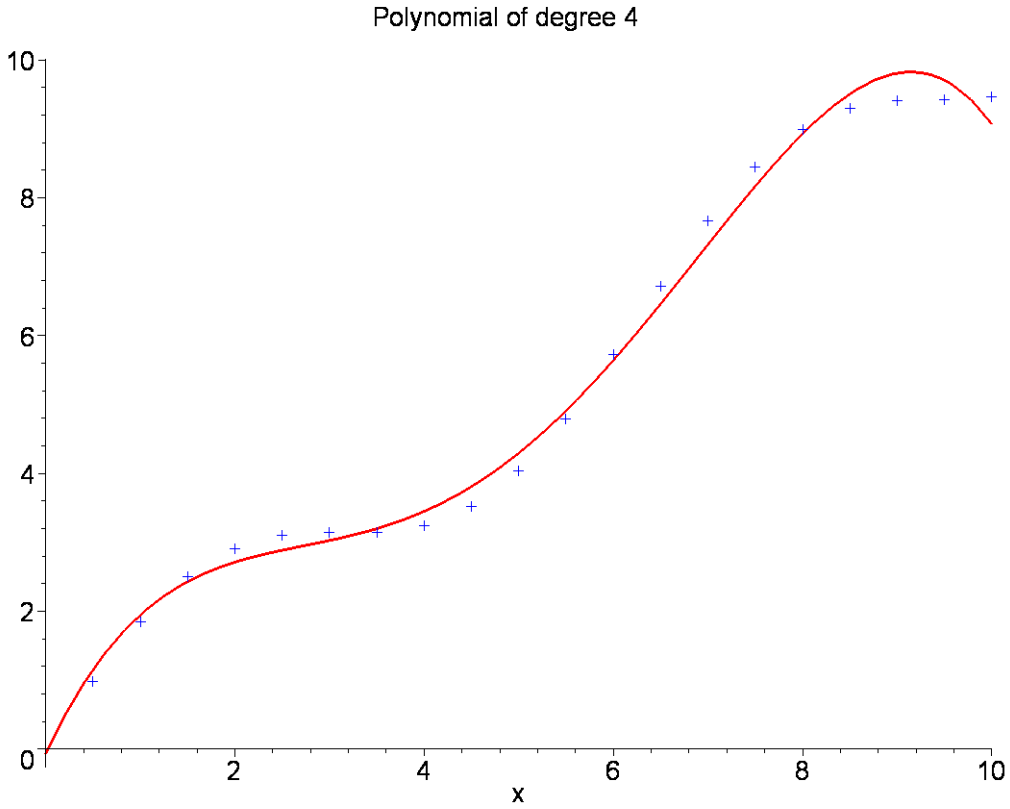
$$eqn2 := y = a + b\,x + c\,x^2 + d\,x^3 + e\,x^4$$

```
> lsq2:=leastsquare[[x,y],eqn2,{a,b,c,d,e}]([Xdata,Ydata]);
```

$$lsq2 :=$$

$$y = -.07174541678 + 2.973013639\,x - 1.138367815\,x^2 + .1940921577\,x^3 - .01008454155\,x^4$$

```
> img2:=plot(rhs(lsq2),x=0..10,color=red,thickness=3,title="Polynomi
  al of degree 4"):
> display([img0,img2]);
```



Don’t confuse correlation and causation!

The degree 4 fit certainly looks better, but you should be careful not to draw any unwarranted conclusions. Extrapolation is definitely out. For example the degree 4 polynomial here decreases after x = 10 whereas x + sin(x) continues to increase! If you were dealing with experimental data the good fit might mislead you! You really need to have an underlying model which implies a certain functional relation before a least squares fit with that function allows you to reasonably extrapolate or draw any conclusions other than statistical ones concerning properties of the actual data.

>