MATLAB for MAPH 3071 Lab 5

Modelling of Data – Regression

The purpose of linear regression fitting is to fit a polynomial to a set of data that minimises the squared distance from the curve to the data. Linear regression is probably the most widely used analysis in statistics.

Regression differs from interpolation in general, in the sense that the curve (NB. A curve can be a straight line !) does not necessarily pass through every data point. In statistical analysis indeed, the law of parsimony is generally followed, with a polynomial of the lowest possible degree which can be usefully used to model the data being fitted.

Example:

```
» x = [ 0 1 2 3 4 ];
» y = [ 1 3 8 9 14 ];
```

If we doing interpolation a cubic spline could be used to interpolate these data. However in linear regression we may want to model the increasing trend in the data (i.e. a x increase so does y) with a simple straight line model – i.e. a polynomial of degree 1.

The polyfit command in MATLAB will return the coefficients of the polynomial that you require for least squares.

```
» polyfit(x,y,1)

ans =

   3.2000    0.6000
```

Therefore the equation for the least squares regression line is 0.6 + 3.2x.

Run the program **regression.m** which is on the class library. This will plot the least squares line and will also plot the interpolating natural cubic spline polynomial for contrast.

It is relatively easy to write an m-file to find the least squares equation. Remember that the least squares equation in matrix form for a set of paired values in X and Y is,

(X` X) B = X` Y

So we can use gaussian elimination to solve for the solution B to this set of linear equations.

Here X is a matrix with the first column all 1’s (which fits the intercept) and the other columns the actual X values, and Y is a columnar vector.

i.e use the A\B command in MATLAB to solve for B

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \qquad y = \begin{bmatrix} 1 \\ 3 \\ 8 \\ 9 \\ 14 \end{bmatrix}$$

» A= X'*X

» Xty= X'*y

$$\overset{A=}{\begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix}} * \overset{B=}{\begin{bmatrix} B_0 \\ B_1 \end{bmatrix}} = \overset{Xty=}{\begin{bmatrix} 35 \\ 102 \end{bmatrix}}$$

Solution for B using gaussian elimination :

```
» A\Xty

ans =

    0.6000
    3.2000
```

To fit a quadratic to this set of data, we find the coefficients of

y = b_0 (constant) + b_1(x) + b_2(x^2)

which will be the same as above except that the x matrix will take the form;

x =

$$\begin{array}{c} \begin{array}{ccc} & x & x^2 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \end{array}$$

See L:\regression2.m for an example of fitting a quadratic to the same data

END