Lecture 11: Normal Equation and Polynomial Approximation

The model has the form of Ax = d where A is an mxn matrix with m larger than n, that is, there are more rows or equations than the unknowns in the x vector. Since there are no exact solutions, we try to find x so that the residual vector in r = d - Ax is as "small" as possible.

Definition. The vector x is called a *least squares* solution of the over determined system if and only if x is such that

$$r^T r = (d - Ax)^T (d - Ax) \text{ is a minimum of all } (d - Ay)^T (d - Ay)$$

Fortunately, this solution in most cases has a very nice answer. If x is the least squares solution, then for all y

$$(d - Ax)^T (d - Ax) \le (d - Ay)^T (d - Ay). \quad (1)$$

Let y = x + (y - x) so that Ay = Ax + A(y-x), and use the properties of transpose to get

$$\begin{aligned}(d - Ay)^T (d - Ay) &= ((d - Ax) - A(y - x))^T ((d - Ax) - A(y - x)) \\ &= ((d - Ax)^T - (A(y - x))^T)((d - Ax) - A(y - x)) \\ &= (d - Ax)^T (d - Ax) - (A(y - x))^T (d - Ax) \\ &\quad - (d - Ax)^T A(y - x) + (A(y - x))^T (A(y - x)) \\ &\ge (d - Ax)^T (d - Ax) - 2((y - x)^T A^T (d - Ax). \quad (2)\end{aligned}$$

If the last term on the right side of (2) is zero, then the inequality in (1) must hold. This prompts the following definition and proposition, which we have just established.

Definition. $A^T Ax = A^T d$ is called the *normal equation* associated with the least squares problem.

Proposition 6. If $A^T A$ has an inverse, then the solution of the normal equation is also a solution of the least squares problem.

Example 1. Consider the problem in the introduction of lecture 9 with three equations and two unknowns

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix} \text{ with}$$

$$A^T A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix} \text{ and}$$

$$A^T d = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 28 \\ 11 \end{bmatrix}.$$

The solution of the normal equation is $x_1 = 3$ and $x_2 = -7/3$. An interpretation of this is with x_1 = slope = m and x_2 = y intercept = c, is that the straight line y = 3x + (-7/3) is closest to the three data points (x_i , y_i) = (1,1), (2,3) and (3,7).

Example 2. Consider the sales data that was discussed the previous lecture.

Month	**Computers Sold**
1	78
2	85
3	90
4	96
5	104
6	113

The matrix equation for this data has the form

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} 78 \\ 85 \\ 90 \\ 96 \\ 104 \\ 113 \end{bmatrix}.$$

$$A^T A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \end{bmatrix} = \begin{bmatrix} 91 & 21 \\ 21 & 1 \end{bmatrix}$$

$$A^T d = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 78 \\ 85 \\ 90 \\ 96 \\ 104 \\ 113 \end{bmatrix} = \begin{bmatrix} 2100 \\ 566 \end{bmatrix}$$

The solution of the normal equation A^TA x = A^T d is x^T =[34/5 1058/15]. This means the slope m = 34/5 and the y intercept c = 1058/15. The predicted sales at 9 months is m9 + c = 132.

Example 3. Consider the following data that looks like data whose graphs looks like a parabola

$\mathbf{x_i}$	$\mathbf{y_i}$
1	1
2	5
3	8
4	17
5	16

So, a good fit to the data is to find a, b, and c such that y(x) = ax^2 + bx + c is "closest" to the data. In the least squares sense the means for $r_i = y_i - y(x_i) = y_i - (a\, x_i^2 + b\, x_i + c\,)$.
The matrix form of this is

$$Ax = d$$

$$\begin{bmatrix} 1^2 & 1 & 1 \\ 2^2 & 2 & 1 \\ 3^2 & 3 & 1 \\ 4^2 & 4 & 1 \\ 5^2 & 5 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 8 \\ 17 \\ 26 \end{bmatrix}.$$

The solution of the normal equations can be found by any of the following three Matlab commands:

(i). (A'*A)\(A'*d) (Matlab uses A' to denote A^T)

(ii). inv(A'*A)*(A'*d)

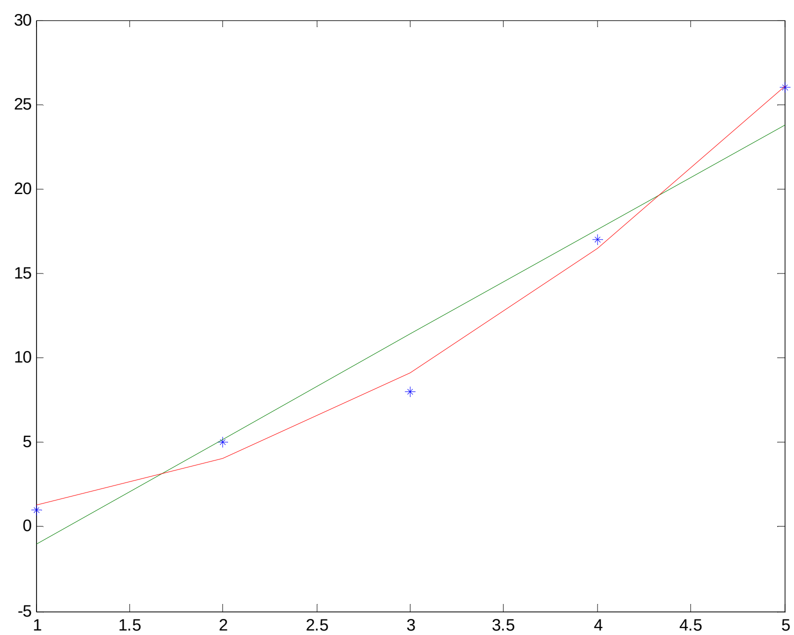
(iii). A\d (The preferred method uses the QR factorization.)

This is illustrated in the Matlab demo ls_quad.

The solution for the parabolic curve fit is a = 1.1429, b = -0.6571 and c = 0.8000. The solution for the linear curve fit is m = 6.2000 and c = -7.2000. One can compare the accuracy of the curve fit by either looking at r^Tr or by looking at the graphs:

r^Tr for the linear approximation is res_lin = 20.8000

r^Tr for the quadratic approximation is res_quad = 2.5143.



Homework.

1. Consider the following sales data, which looks like a straight line. Find the straight line that is a least squares fit to this data

Time	**Sales**
1	10
2	13
3	17
5	19

   Find the matrix representation, the normal equations, the m and c in $y = mx+c$.

2. Consider the following data, which looks like a parabola. Find the parabola that is a least squares fit to this data where x_i are in the left and y_i are in the right.

0	9
1	10
2	16
3	27
5	41

   Find the matrix representation, the normal equations, and the a, b and c in $y = ax^2 + bx + c$.

3. Consider the following data, which looks like a cubic. Find the cubic polynomial that is a least squares fit to this data where x_i are in the left and y_i are in the right.

1	-1
2	3
3	5
4	2
5	0
6	1
6	7

   Find the matrix representation, the normal equations, and the a, b, c and e in $y = ax^3 + bx^2 + cx + e$.