Lecture 10: Curve Fitting to Data

We will investigate algebraic systems that have more equations than unknowns. In general, these will not have a solution in the sense that one gets equality for each equation. Here we hope to be able to choose the unknowns so that the equations are as close to being satisfied as possible. For example, consider the following where there are two variables and three equations

$1x_1 + 1x_2 = 1$,

$2x_1 + 1x_2 = 3$ and

$3x_1 + 1x_2 = 7$.

Here there is no solution because there are three distinctly different equations and only two unknowns. Or, in matrix form, with the residual vector, r, inserted is

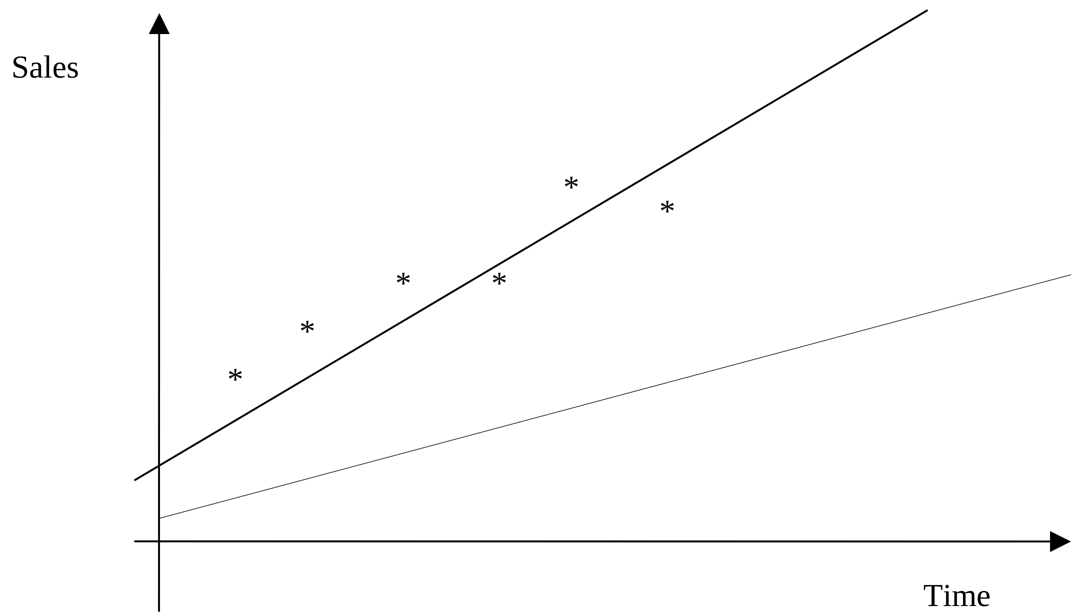
$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix} - \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}.$$

We will want to choose the solution x so that the "size" of the residual vector is a minimum. A very appropriate measure of "size" will be the sum of the squares of residual's components $r^T r = r_1^2 + r_2^2 + r_3^2$. Such x are called the *least squares* solution to the overdetermined system.

Application to Business Forecasting. Consider a computer company, which has recorded sales of 78, 85, 90, 96, 104 and 113 computers over the last six months:

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

They wish to make a prediction of the sales over the next six months. This will help them plan their production needs during this period. The data is increasing more or less in a linear fashion. Consequently, we are looking for a straight line that is "closest" to the data. An analytical way of saying this is that we are looking for y = mx + c, that is, the slope, m, and the y intercept, c, so that this line is "closest" to the graphed data. Once the m and the c are known, then the future sales can be predicted by putting the appropriate month into the x variable and computing the forecasted sales in the y variable.



In order to form the corresponding over determined system, note the components of the residual vector are just the vertical distances between the data points and the desired straight line. So, for each data point there is an equation, which has the two unknowns m and c. For the above data the six equations are

$m1 + c = 78$,

$m2 + c = 85$,

$m3 + c = 90$,

$m4 + c = 96$,

$m5 + c = 104$ and

$m6 + c = 113$.

Or, the matrix form with the residual vector is

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} 78 \\ 85 \\ 90 \\ 96 \\ 104 \\ 113 \end{bmatrix} - \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_6 \end{bmatrix}.$$

The *least squares solution* is the m and c so that the sum of the squares of the residual is a minimum, that is,

$f(m,c) = r_1^2 + \ldots + r_6^2$ is a minimum.

In order to find m and c, either one could use calculus or matrix algebra methods. In either method an algebraic system must be solved. This system is called the normal equations, and it will require the transpose operation, which we shall shortly describe.

Not all data has the form of a straight line. For example, the following data has a graph that will look more like a parabola.

x_i	y_i
1	1
2	3
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 this means for $r_i = y_i - y(x_i) = y_i - (a\, x_i^2 + b\, x_i + c)$

$f(a,b,c) = r_1^2 + \ldots + r_5^2$ is a minimum.

Definition. Let x be a nx1 column vector with components $x_1, \ldots, x_n$.

x^T , called the *transpose of x*, is 1xn row vector with components $x_1, \ldots, x_n$.

$x^T = [\, x_1\ x_2\ \cdots\ x_n]$. Let x and y be two nx1 column vectors.

$x^T y$, called the *dotproduct of x and y*, is a single number equal to the sum of the products of the components of x and y

$$x^T y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Let A be an mxn matrix A = [a_1 a_2 …. a_n] where a_j are the nx1 column vectors. A^T, called the *transpose of A*, is an nxm matrix formed from A by interchanging the rows and columns.

$$A^T = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix}$$

Examples.

$$Let\ x = \begin{bmatrix} 1 \\ 7 \\ 8 \end{bmatrix}, y = \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix} and\ A = \begin{bmatrix} 1 & 3 & 2 \\ -1 & 7 & 9 \end{bmatrix}.$$

$$x^T x = 1 \cdot 1 + 7 \cdot 7 + 8 \cdot 8 = 114$$

$$x^T y = 1 \cdot (-1) + 7 \cdot 2 + 8 \cdot 3 = 37$$

The columns of A are

$$a_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, a_2 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, and\ a_3 = \begin{bmatrix} 2 \\ 9 \end{bmatrix}.$$

$$A^T = \begin{bmatrix} a_1^T \\ a_2^T \\ a_3^T \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 3 & 7 \\ 2 & 9 \end{bmatrix}.$$

Proposition 5.

1. Let x and y be nx1 column vectors. $x^Ty = y^Tx$.
2. Let A and B be mxn matrices. $(A + B)^T = A^T + B^T$.
3. Let A be a mxn matrix and c a single number. $(cA)^T = c(A^T)$.
4. Let A be a mxn matrix and x a nx1 column vector. $(Ax)^T = x^TA^T$.
5. Let A be mxn and B be nxp matrices. $(AB)^T = B^T A^T$.

Proof of 4.

$$\begin{aligned}(Ax)^T &= (\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} x)^T \\ &= (a_1 x_1 + a_2 x_2 + \cdots + a_n x_n)^T \\ &= a_1^T x_1 + a_2^T x_2 + \cdots + a_n^T x_n \\ &= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} \\ &= x^T A^T.\end{aligned}$$

Examples. There are additional examples in the Matlab demo oper_transpose.m

$$Let\ A = \begin{bmatrix} 2 & 7 \\ 8 & -1 \end{bmatrix} and\ B = \begin{bmatrix} 1 & 3 & 4 \\ 7 & 8 & 9 \end{bmatrix}.$$

$$A^T = \begin{bmatrix} 2 & 8 \\ 7 & -1 \end{bmatrix} and\ B^T = \begin{bmatrix} 1 & 7 \\ 3 & 8 \\ 4 & 9 \end{bmatrix}$$

$$AB = \begin{bmatrix} 2 & 7 \\ 8 & -1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 4 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 51 & 62 & 71 \\ 1 & 16 & 23 \end{bmatrix}$$

$$(AB)^T = \begin{bmatrix} 51 & 1 \\ 62 & 16 \\ 71 & 23 \end{bmatrix}$$

$$B^T A^T = \begin{bmatrix} 1 & 7 \\ 3 & 8 \\ 4 & 9 \end{bmatrix} \begin{bmatrix} 2 & 8 \\ 7 & -1 \end{bmatrix} = \begin{bmatrix} 51 & 1 \\ 62 & 16 \\ 71 & 23 \end{bmatrix}$$

Homework.

1. Verify the properties of transpose in Proposition 5 for the following:

$$Let\ A = \begin{bmatrix} 4 & 2 \\ 3 & -1 \end{bmatrix}, x = \begin{bmatrix} 1 \\ 4 \end{bmatrix} and\ B = \begin{bmatrix} 11 & 0 & -1 \\ 2 & 3 & 3 \end{bmatrix}.$$

2. Prove properties 1-3 in Proposition 5.
3. Prove property 5 in Proposition 5.
4. Find the coefficient matrix for the parabolic curve fit to the data x_i and y_i.