6.4.2 Matrix formulation of least squares regression: normal system

We can generalize the model $M(x)$ defined in the previous section to include m model functions λ_j and model parameters p_j, where $j = 1, 2, ..., m$. So the model now takes the form $M(x) = p_1\lambda_1(x) + p_2\lambda_2(x) + \cdots p_m\lambda_m(x)$ with $\lambda_1(x) \equiv 1$, $\forall x$. Ideally, $M(x)$ would fit the data perfectly, i.e. $y_i \equiv M(x_i)$ whence we would arrive at a simple linear system written in matrix form as follows.

$$\Lambda \mathbf{p} = \mathbf{q} \tag{6.15}$$

$$\begin{pmatrix} 1 & \lambda_{12} & \lambda_{13} & \cdots & \lambda_{1m} \\ 1 & \lambda_{22} & \lambda_{23} & \cdots & \lambda_{2m} \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ 1 & \lambda_{n2} & \lambda_{n3} & \cdots & \lambda_{nm} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ . \\ . \\ p_m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_m \end{pmatrix} \tag{6.16}$$

Here we have used the notation used earlier: $\lambda_{ik} = \lambda_k(x_i)$. These λ_{ij}s populate the matrix Λ. There are n rows in Λ corresponding to n equations for n data points (x_i, y_i), $i = 1, 2, ..., n$ and m columns corresponding to m model parameters.

However, like we observed in Figure 6.8, it is unlikely for a finite dimensional model to exactly fit a dataset. Our goal, as always, must be to minimize the two-norm (squared) of the error vector: $e = ||\Lambda \mathbf{p} - \mathbf{q}||^2$, for which we must set the derivatives $\frac{\partial e}{\partial p_k}$ equal to 0 for all $k = 1, 2, ..., m$ following the arguments presented in the previous section. This leads to the following set of equations.

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \lambda_{ik}\lambda_{ij}p_j = \sum_{i=1}^{n} \lambda_{ik}q_i, \tag{6.17}$$

where $k = 1, 2, ..., n$ designates the equation for the k^{th} data point. The system of equations 6.17 may be written more compactly in matrix-vector form as follows.

$$(\Lambda^T\Lambda)\mathbf{p} = \Lambda^T\mathbf{q}. \tag{6.18}$$

Equations 6.18 is known as the *normal system* whose solution $\mathbf{p}$ is the least squares solution to the linear system $\Lambda\mathbf{p} = \mathbf{q}$.[7]

[7] If Λ is of dimensions $(n \times m)$ and if $rank(\Lambda) = m$, then $\Lambda^T\Lambda$ is always invertible and there exists a unique least squares solution of the normal system 6.18, namely, $\mathbf{p} = (\Lambda^T\Lambda)^{-1}\Lambda^T\mathbf{q}$.

Example: quadratic least squares model

Consider the following data from an experiment. Find a quadratic least squares model that best fits this data.

x	y
−1	0
0	1
1	3
2	9
3	19

Solution: Our objective here is to fit a model $M(x) = a + bx + cx^2$. We will begin by constructing the entries of the matrix Λ. $\lambda_{12} = x_1 = -1$, $\lambda_{13} = 1^2$. In fact, $\lambda_{i2} = x_i$ and

This will be (x_i)^2

$\lambda_{i3} = x_i^3$ for all $i = 1, 2, ..., n = 5$. Therefore, $\Lambda = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}$ and $\mathbf{q} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 9 \\ 19 \end{pmatrix}$. The matrix Λ has rank $m = 3$ and hence there exists a unique least squares solution $\mathbf{p} = (\Lambda^T\Lambda)^{-1}\Lambda^T\mathbf{q} = \begin{pmatrix} 0.2286 \\ 1.4571 \\ 1.5714 \end{pmatrix}$. This model and the data are plotted in Figure 6.10.

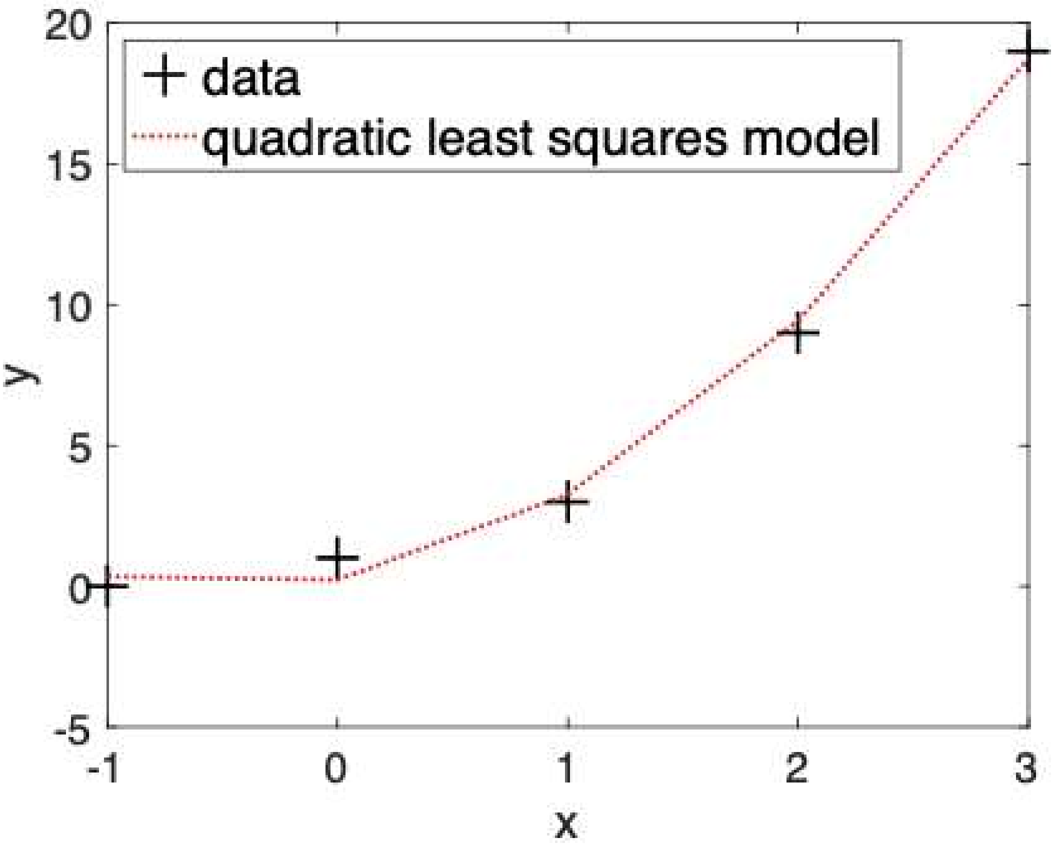

Figure 6.10: Quadratic least squares fit to the data in the example of section 6.4.2.