Matrix Least-Squares Symbolic Solution

Matrix Formalism for a Straight Line

Consider the least-squares fit to a straight line, $y = a_0 + a_1x$. The matrix solution is given by,

$$
\begin{aligned}
&\mathbf{y} = \mathbf{X}\mathbf{a} \\
&\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\mathbf{X}\mathbf{a} \\
&\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{a}
\end{aligned}
\tag{1}
$$

Symbolic Solution

For simplicity in "symbol pushing" assume that only three points are used: (x_1,y_1), (x_2,y_2) and (x_3,y_3). Start by writing out the two vectors and one matrix appearing in the first expression of Eq. 1.

$$
\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{pmatrix} \qquad \mathbf{a} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}
\tag{2}
$$

Second, define the transpose of $\mathbf{X}$, and evaluate $\mathbf{X}^T\mathbf{X}$ (N is the number of data points).

$$
\begin{aligned}
&\mathbf{X}^T = \begin{pmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \end{pmatrix} \\
&\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \end{pmatrix}\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{pmatrix} = \begin{pmatrix} 1+1+1 & x_1+x_2+x_3 \\ x_1+x_2+x_3 & x_1^2+x_2^2+x_3^2 \end{pmatrix} = \begin{pmatrix} N & \sum x \\ \sum x & \sum x^2 \end{pmatrix}
\end{aligned}
\tag{3}
$$

Third, take the inverse of Eq. 3. To do this use the well-known symbolic inverse of a 2×2 matrix (known as an "abcd" matrix).

$$
\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}
\tag{4}
$$

Now plug the appropriate terms into Eq. 4 to evaluate the inverse, $(\mathbf{X}^T\mathbf{X})^{-1}$.

$$\begin{pmatrix} N & \sum x \\ \sum x & \sum x^2 \end{pmatrix}^{-1} = \frac{1}{N\sum x^2 - \left(\sum x\right)^2}\begin{pmatrix} \sum x^2 & -\sum x \\ -\sum x & N \end{pmatrix} \tag{5}$$

Fourth, evaluate $\mathbf{X}^T\mathbf{y}$.

$$\mathbf{X}^T\mathbf{y} = \begin{pmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} y_1 + y_2 + y_3 \\ x_1y_1 + x_2y_2 + x_3y_3 \end{pmatrix} = \begin{pmatrix} \sum y \\ \sum xy \end{pmatrix} \tag{6}$$

Finally, evaluate the left side of the last expression in Eq. 1 to obtain the least-squares coefficients.

$$\begin{aligned}
\mathbf{a} &= \frac{1}{N\sum x^2 - \left(\sum x\right)^2}\begin{pmatrix} \sum x^2 & -\sum x \\ -\sum x & N \end{pmatrix}\begin{pmatrix} \sum y \\ \sum xy \end{pmatrix} \\
\mathbf{a} &= \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \frac{1}{N\sum x^2 - \left(\sum x\right)^2}\begin{pmatrix} \sum y \sum x^2 - \sum x \sum xy \\ N\sum xy - \sum x \sum y \end{pmatrix} \\
\Delta &= N\sum x^2 - \left(\sum x\right)^2 \\
\hat{a}_0 &= \frac{\sum y \sum x^2 - \sum xy \sum x}{\Delta} \\
\hat{a}_1 &= \frac{N\sum xy - \sum x \sum y}{\Delta}
\end{aligned} \tag{7}$$

This is the same result derived in the handout, “The Statistical Underpinning of Least-Squares”, and given as Eq. 11. The “hat” on the coefficients indicates that they are experimental estimates of the true value.

Coefficient Errors

When multiplied by σ^2, Eq. 5 is called the variance-covariance equation, $\mathbf{C} = \sigma^2\,(\mathbf{X}^T\mathbf{X})^{-1}$, where $c_{i,i}$ is the variance of the i^{th} coefficient and $c_{i,j}$ is the co-variance of the i^{th} and j^{th} coefficients. For a least-squares fit to a straight line, Eq. 5 yields the following variances and covariances.

$$\sigma_{a_0}^2 = \frac{\sigma^2\sum x^2}{\Delta} \qquad \sigma_{a_1}^2 = \frac{\sigma^2 N}{\Delta} \qquad \operatorname{cov}\left(a_0, a_1\right) = \frac{-\sum x}{\Delta} \tag{8}$$

Since σ is seldom known, it is replaced by the experimental estimate, s,

$$s^2 = \frac{1}{N-n}\sum_{i=1}^{N}\left(y_i - \hat{a}_0 - \hat{a}_1 x_i\right)^2 \tag{9}$$

where n is the number of coefficients and $(N - n)$ the degrees of freedom. Again, Eq. 8 agrees with the result derived in the handout, "The Statistical Underpinning of Least-Squares". Note that the covariance is negative and significant. This implies that if there is a positive error in the slope, the corresponding error is the intercept will be negative (and *vice versa*).

Minimizing Coefficient Errors with a Straight-Line Fit

Before examining strategies, rewrite Δ from Eq. 7 by multiplying the right hand side by N^2/N^2 and rearranging.

$$\begin{aligned}\Delta &= N\sum x^2 - \left(\sum x\right)^2 = N^2\left[\frac{1}{N}\sum x^2 - \left(\frac{1}{N}\sum x\right)^2\right] \\ \Delta &= N^2\sigma_x^2\end{aligned} \tag{10}$$

The term within brackets is the variance of the error-free x-axis values. If the x-axis values can be selected by the experimentalist, this variance can be controlled. Using this new notation for Δ, the errors of the coefficients can be rewritten.

$$\sigma_{a_0}^2 = \frac{\sigma^2}{N^2\sigma_x^2}\sum x^2 \qquad \sigma_{a_1}^2 = \frac{\sigma^2}{N\sigma_x^2} \tag{11}$$

Looking at these equations, there are three strategies available to minimize both errors and one that works only with the intercept.

1. decrease σ
2. increase N
3. increase σ_x by increasing the range of x-values
4. for the intercept, center the x-values so that the sum is zero

To examine in more detail strategy 4, set the sum of x in Δ equal to zero, this results in the following,

$$\begin{aligned}\Delta &= N\sum x^2 - \left(\sum x\right)^2 = N\sum x^2 \\ \sigma_{a_0}^2 &= \frac{\sigma^2\sum x^2}{\Delta} = \frac{\sigma^2\sum x^2}{N\sum x^2} = \frac{\sigma^2}{N}\end{aligned} \tag{12}$$

where the error in the intercept is now the same as that for the error of the average. This is the smallest error achievable for the intercept. Unfortunately, x-values cannot always be centered, e.g. concentrations.