1.4.8 Multiple regression fits two-dimensional data by a plane $y = C + Dt + Ez$. For the data below, write down the 4 equations in the 3 unknowns C, D, E. What is the least squares solution from the normal equations?

	$z=0$	$z=1$
$t=0$	$b_1 = 2$	$b_2 = 2$
$t=1$	$b_3 = 1$	$b_4 = 5$

$\longrightarrow$ System:

$$\begin{aligned} 2 &= C \\ 2 &= C + E \\ 1 &= C + D \\ 5 &= C + D + E \end{aligned}$$

$$\overset{A}{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}} \overset{x}{\begin{bmatrix} C \\ D \\ E \end{bmatrix}} = \overset{b}{\begin{bmatrix} 2 \\ 2 \\ 1 \\ 5 \end{bmatrix}} \rightarrow \overset{A^TA}{\begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}} \overset{x}{\begin{bmatrix} C \\ D \\ E \end{bmatrix}} = \overset{A^Tb}{\begin{bmatrix} 10 \\ 6 \\ 7 \end{bmatrix}} \longrightarrow x = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

$$\underline{y = 1 + t + 2z}$$

1.4.9 For a matrix A with more columns than rows the matrix A^TA is not positive definite. Why is it impossible for the columns of A^TA to be independent?

Compute A^TA for the following A matrices and check that A^TA is singular.

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

If A has more columns than rows, the columns must be dependent. The columns of A are dependent iff $\exists\, x \neq 0$ s.t. $Ax = 0$. But this means $\exists\, x \neq 0$ s.t. $A^TAx = A^T0 = 0$, which is the case iff A^TA is singular iff the columns of A^TA are dependent.

Also, each column of A^TA is a linear combination of the columns of A^T, with coefficients given by one of the columns of A. Since this column of A could be written as a linear combination of the other columns of A, so could any column of A^TA be written as a linear combination of the other columns.

$$A^TA = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 5 & 2 \\ 5 & 5 & 2 \\ 2 & 2 & 1 \end{bmatrix} \text{ is singular (column 1 = column 2)}$$

$$A^TA = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \text{ is singular (column 2 = } 2 \cdot \text{column 1)}$$

1.4.4 Find the best least squares solution to

$$Ax = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix} = b$$

By 1H, the best line is $y = \bar{b} + \alpha (t - \bar{t})$ with $\alpha = \dfrac{\sum (t_i - \bar{t}) b_i}{\sum (t_i - \bar{t})^2}$.

$\bar{b} = (0+4+2)/3 = 2$, $\bar{t} = (1+2+3)/3 = 2$

$$\alpha = \frac{\sum (t_i - \bar{t}) b_i}{\sum (t_i - \bar{t})^2} = \frac{(1-2)\cdot 0 + (2-2)\cdot 4 + (3-2)\cdot 2}{(1-2)^2 + (2-2)^2 + (3-2)^2} = 1$$

The best least squares solution is $y = t$ $(C = 0, D = 1)$

1.4.7 For the measurements $b = 0, 3, 12$ at times $t = 0, 1, 2$, find

(i) the best horizontal line $y = C$

$$Ax = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [C] = \begin{bmatrix} 0 \\ 3 \\ 12 \end{bmatrix} = b \longrightarrow A^T A x = 3C = 15 = A^T b \longrightarrow \underline{y = 5}$$

(ii) the best straight line $y = Ct + D$

$$Ax = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 12 \end{bmatrix} = b \rightarrow A^T A x = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 15 \\ 27 \end{bmatrix} = A^T b$$

$$\rightarrow x = \begin{bmatrix} -1 \\ 6 \end{bmatrix} \quad \underline{y = 6t - 1}$$

(iii) the best parabola $y = C + Dt + Et^2$

$$Ax = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 12 \end{bmatrix} = b \rightarrow A^T A x = \begin{bmatrix} 3 & 3 & 5 \\ 3 & 5 & 9 \\ 5 & 9 & 17 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 15 \\ 27 \\ 51 \end{bmatrix}$$

$$\rightarrow x = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} \quad \underline{y = 3t^2}$$

Exercises

1.4.2 What equations determine the minimizing x for

(i) $P = \frac{1}{2}x^TAx - x^Tb$?

By 1E,

P is minimized at the point x where $Ax = b$.

The minimum value is $P(A^{-1}b) = -\frac{1}{2}b^TA^{-1}b$.

(ii) $P = \frac{1}{2}x^TA^TAx - x^TA^Tb$?

Replacing A with A^TA and b with A^Tb in 1E,

P is minimized by the point x where $A^TAx = A^Tb$.

The minimum value is $P((A^TA)^{-1}A^Tb) = -\frac{1}{2}b^TA(A^TA)^{-1}A^Tb = -\frac{1}{2}b^Tb$

(iii) $E = \|Ax-b\|^2$?

$E = \|Ax-b\|^2 = (Ax-b)^T(Ax-b) = x^TA^TAx - x^TA^Tb - b^TAx + b^Tb$

$= x^TA^TAx - 2x^TA^Tb + b^Tb$ ($x^TA^Tb = b^TAx$ since they are scalars)

Since b^Tb does not depend on x, minimizing $x^TA^TAx - 2x^TA^Tb + b^Tb$ is equivalent to minimizing $x^TA^TAx - 2x^TA^Tb$. But then minimizing $x^TA^TAx - 2x^TA^Tb$ is equivalent to minimizing $\frac{1}{2}x^TA^TAx - x^TA^Tb$, which is P of part (ii). Therefore:

E is minimized by the point x where $A^TAx = A^Tb$

The minimum value is $E((A^TA)^{-1}A^Tb) = 0$

Check: $E((A^TA)^{-1}A^Tb) = b^TA(A^TA)^{-1T}A^TA(A^TA)^{-1}A^Tb - 2b^TA(A^TA)^{-1T}A^Tb + b^Tb$

$= b^TA(A^TA)^{-1}A^Tb - 2b^TA(A^TA)^{-1}A^Tb + b^Tb$

$= b^TAA^{-1}A^{-1T}A^Tb - 2b^TAA^{-1}A^{-1T}A^Tb + b^Tb$

$= b^Tb - 2b^Tb + b^Tb = 0$

Alternatively, reverse the logic of the previous paragraph, applied to the minimum P of part (ii), to find:

$$\min_x E = 2(-\tfrac{1}{2}b^Tb) + b^Tb = -b^Tb + b^Tb = 0$$

$$\left[\begin{array}{cc|c} m & \Sigma t_i & \Sigma b_i \\ \Sigma t_i & \Sigma t_i^2 & \Sigma t_i b_i \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & \bar{t} & \bar{b} \\ \bar{t} & \frac{1}{m}\Sigma t_i^2 & \frac{1}{m}\Sigma t_i b_i \end{array}\right]$$

$$\sim \left[\begin{array}{cc|c} 1 & \bar{t} & \bar{b} \\ 0 & \frac{1}{m}\Sigma t_i^2 - \bar{t}^2 & \frac{1}{m}\Sigma t_i b_i - \bar{b}\bar{t} \end{array}\right]$$

$$\sim \left[\begin{array}{cc|c} 1 & \bar{t} & \bar{b} \\ 0 & 1 & \alpha \end{array}\right] \quad (\alpha \text{ simplified below})$$

$$\sim \left[\begin{array}{cc|c} 1 & 0 & \beta \\ 0 & 1 & \alpha \end{array}\right] \quad (\beta = \bar{b} - \alpha\bar{t} \text{ simplified below})$$

$$\alpha = \frac{\frac{1}{m}\Sigma t_i b_i - \bar{b}\bar{t}}{\frac{1}{m}\Sigma t_i^2 - \bar{t}^2} = \frac{\Sigma t_i b_i - \frac{1}{m}\Sigma b_i \Sigma t_i}{\Sigma t_i^2 - \frac{1}{m}\Sigma t_i \Sigma t_i} = \frac{\Sigma t_i b_i - \Sigma \bar{t} b_i}{\Sigma t_i^2 - \Sigma \bar{t} t_i} = \frac{\Sigma(t_i - \bar{t})b_i}{\Sigma(t_i^2 - \bar{t}t_i)}$$

$$= \frac{\Sigma(t_i - \bar{t})b_i}{\Sigma(t_i^2 - 2\bar{t}t_i + \bar{t}^2) + \Sigma(\bar{t}t_i - \bar{t}^2)} = \frac{\Sigma(t_i - \bar{t})b_i}{\Sigma(t_i - \bar{t})^2 + \Sigma(\bar{t}t_i) - \bar{t}^2} = \frac{\Sigma(t_i - \bar{t})b_i}{\Sigma(t_i - \bar{t})^2 + \bar{t}^2 - \bar{t}^2}$$

$$\alpha = \frac{\Sigma(t_i - \bar{t})b_i}{\Sigma(t_i - \bar{t})^2}$$

<u>or</u>

$$\alpha = \frac{\frac{1}{m}\Sigma t_i b_i - \bar{b}\bar{t}}{\frac{1}{m}\Sigma t_i^2 - \bar{t}^2} = \frac{\Sigma t_i b_i - m\bar{b}\bar{t}}{\Sigma t_i^2 - m\bar{t}^2}$$

$$\beta = \bar{b} - \alpha\bar{t} = \bar{b} - \frac{\bar{t}\Sigma t_i b_i - m\bar{b}\bar{t}^2}{\Sigma t_i^2 - m\bar{t}^2} = \frac{\bar{b}\Sigma t_i^2 - m\bar{b}\bar{t}^2 - \bar{t}\Sigma t_i b_i + m\bar{b}\bar{t}^2}{\Sigma t_i^2 - m\bar{t}^2}$$

$$\beta = \frac{\bar{b}\Sigma t_i^2 - \bar{t}\Sigma t_i b_i}{\Sigma t_i^2 - m\bar{t}^2}$$

The line of best fit is $y = \alpha t + \beta$ with

$$\boxed{\alpha = \frac{\Sigma(t_i - \bar{t})b_i}{\Sigma(t_i - \bar{t})^2} = \frac{\Sigma t_i b_i - m\bar{b}\bar{t}}{\Sigma t_i^2 - m\bar{t}^2} \qquad \beta = \frac{\bar{b}\Sigma t_i^2 - \bar{t}\Sigma t_i b_i}{\Sigma t_i^2 - m\bar{t}^2}}$$

<u>Alternatively</u> we could write this using $\beta = \bar{b} - \alpha\bar{t}$ as

$$\boxed{y = \alpha t + \bar{b} - \alpha\bar{t} = \bar{b} + \alpha(t - \bar{t}) \text{ with } \alpha = \frac{\Sigma(t_i - \bar{t})b_i}{\Sigma(t_i - \bar{t})^2} = \frac{\Sigma t_i b_i - m\bar{b}\bar{t}}{\Sigma t_i^2 - m\bar{t}^2}}$$

Linear Regression

1H Given measurements $b_1, \ldots, b_m$ at times $t_1, \ldots, t_m$, the line $y = \alpha t + \beta$ which minimizes error $\|e\|^2$ is determined by

$$A^T A x = A^T b \quad \text{or} \quad \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}$$

The best line is $y = \bar{b} + \alpha(t - \bar{t})$ with $\alpha = \dfrac{\sum (t_i - \bar{t}) b_i}{\sum (t_i - \bar{t})^2}$

Proof: The previous section explains how to set up a system that will minimize the error in describing the relationship between the t_i and b_i using a line $y = Ct + D$. First write out the system we want to solve

$$\begin{aligned} \beta + \alpha t_1 &= b_1 \\ \beta + \alpha t_2 &= b_2 \\ &\vdots \\ \beta + \alpha t_m &= b_m \end{aligned} \longrightarrow \underset{(A)}{\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}} \underset{(x)}{\begin{bmatrix} \beta \\ \alpha \end{bmatrix}} = \underset{(b)}{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}} \longrightarrow Ax = b$$

If $Ax = b$ has a solution this is fine – it means all the b_i lie along a line and error $e = 0$. But in the case we mainly want to consider, the measurements will not be perfectly linear and $Ax = b$ has no solution:

$$\text{solve } A^T A x = A^T b \text{ for } x$$

Unless $t_1 = t_2 = \ldots = t_m$ the columns of A are independent, $A^T A$ is positive definite and thus a unique solution $x = (A^T A)^{-1} A^T b$ can be found.*

$$A^T A x = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = A^T b$$

$$A^T A x = \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum b_i t_i \end{bmatrix} = A^T b$$

This shows that to find the α, β such that the error in fitting $y = \alpha t + \beta$ to b is minimized by solving this 2×2 system. Use either the formula for the inverse of a 2×2 matrix or use Gaussian elimination. We use the second approach next to find $x = \begin{bmatrix} \beta \\ \alpha \end{bmatrix}$.

* If the columns of A were dependent we would need to use a pseudoinverse.

1G The x that minimizes $\|Ax-b\|^2$ is the solution to the <u>normal equations</u>:

$$A^TAx = A^Tb$$

This vector $x = (A^TA)^{-1}A^Tb$ is the least squares solution to $Ax=b$.

Proof: For this section it is assumed that the columns of A are independent. Then A^TA is positive definite by 1F. It follows that A^TA is invertible since

$A^TAx = 0 \rightarrow x^TA^TAx = 0 \rightarrow x = 0$ shows that if x is in the null space of A, then $x = 0$. In other words, the null space of A^TA is trivial.

The error $e = b - Ax$ is minimized when $\|Ax-b\|^2$ is minimized.

$$\|Ax-b\|^2 = (Ax-b)^T(Ax-b) = (x^TA^T - b^T)(Ax-b) = x^TA^TAx - x^TA^Tb - b^TAx + b^Tb$$

But $\|Ax-b\|^2$ is minimized when $\frac{1}{2}\|Ax-b\|^2$ is minimized and $x^TA^Tb = b^TAx$ since both are scalars so $x^TA^Tb = (x^TA^Tb)^T = b^TAx$. Finally note that since b^Tb is constant wrt x, it does not affect our choice of x in the minimization and should therefore be ignored in the minimization.

Thus we want to find the x that minimizes $\frac{1}{2}x^TA^TAx - x^TA^Tb$. Recall from 1E that the solution x of $Ax=b$ minimizes $P(x) = \frac{1}{2}x^TAx - x^Tb$. Replace the matrix A by the matrix A^TA and the vector b by the vector A^Tb. Then 1E applies here since A^TA is positive definite. This tells us that the x that minimizes $\frac{1}{2}x^TA^TAx - x^TA^Tb$ is the solution to $A^TAx = A^Tb$.

$$\therefore\ x = (A^TA)^{-1}A^Tb \text{ minimizes the error } e = b - Ax$$

We could call the error minimizing point $p = (A^TA)^{-1}A^Tb$ for 'projection' since Ap is the closest point to b in the column space of A. Note

$$A^Te = A^T(b - Ap) = A^Tb - A^TAp = 0$$

The inner (dot) product of e with each column of A is zero by this calculation — the error is orthogonal to the column space of A when we choose $x = p$, the projection of b onto the column space of A.

The takeaway from this section is simple:

<u>If $Ax=b$ has no solution, multiply by A^T and solve $A^TAx = A^Tb$</u>

Consider now the case that A is not assumed to be symmetric positive definite. The coefficient matrix for a physical problem often gets assembled as A^TCA. The matrices A^TA and A^TCA are always symmetric (when C is). We want to know when they are positive definite.

1F (i) If A has linearly independent columns — it can be square or rectangular — then the product A^TA is positive definite.

(ii) If C is symmetric positive definite, so is the triple product A^TCA.

Note that if A is $m\times n$, the columns of A can only be independent if $n \le m$. If $n > m$ we have no hope of independence since the number of independent rows is the same as the number of independent columns by the fundamental theorem of linear algebra.

Proof (1F): Suppose the columns of A are linearly independent.

$x^TA^TAx = (Ax)^T(Ax) = \|Ax\|_2^2 \geq 0$ with equality iff $Ax = 0$. Since the columns of A are independent, $Ax = 0$ iff $x = 0$. Thus x^TA^TAx is positive except when $x = 0$. Conclude A^TA is positive definite.

Suppose C is positive definite and A has linearly independent columns.

$x^TA^TCAx = (Ax)^TC(Ax) \geq 0$ with equality iff $Ax = 0$ iff $x = 0$.

Therefore A^TCA is positive definite.

Least Squares Solution of $Ax=b$

Suppose A is $m\times n$ with $m>n$. The problem $Ax = b$ is an overdetermined system of m equations in n unknowns. The vectors b that can be solved for form the n-dimensional subspace of m-dimensional space. This subspace is called the column space of A. $Ax = b$ has a solution iff b is in the column space of A. With $n < m$ that is unlikely.

In general there will be an error $e = b - Ax$. We emphasize that the components e_i of e are the 'vertical' distances between each component of b and Ax, not the shortest distances.

To minimize error is equivalent to minimizing $\|Ax-b\|^2 = (Ax-b)^T(Ax-b)$ if we measure a vector in the most common way.

The question is which x will minimize $\|Ax-b\|^2$. The answer is given in 1G, which we prove using 1E.

1.4 Minimum Principles

Notes

1E If A is positive definite, the quadratic $P(x) = \frac{1}{2}x^TAx - x^Tb$ is minimized at the point where $Ax = b$. The minimum value is

$$P(A^{-1}b) = -\frac{1}{2}b^TA^{-1}b$$

Proof: Suppose x is the solution to $Ax = b$ and let y be any point.

$$\begin{aligned}
P(y) - P(x) &= \tfrac{1}{2}y^TAy - y^Tb - \tfrac{1}{2}x^TAx + x^Tb \\
&= \tfrac{1}{2}y^TAy - y^TAx - \tfrac{1}{2}x^TAx + x^TAx \\
&= \tfrac{1}{2}y^TAy - y^TAx + \tfrac{1}{2}x^TAx \\
&= \tfrac{1}{2}y^TAy - \tfrac{1}{2}y^TAx - \tfrac{1}{2}y^TAx + \tfrac{1}{2}x^TAx \\
&= \tfrac{1}{2}(y^TAy - y^TAx - y^TAx + x^TAx) \\
&= \tfrac{1}{2}(y^TAy - x^TAy - y^TAx + x^TAx) \\
&= \tfrac{1}{2}((y^T - x^T)Ay - (y^T - x^T)Ax) \\
&= \tfrac{1}{2}((y-x)^TAy + (y-x)^TA(-x)) \\
&= \tfrac{1}{2}(y-x)^TA(y-x)
\end{aligned}$$

Use: $y^TAx = y^TA^Tx = x^TAy$

$P(y) - P(x) = \frac{1}{2}(y-x)^TA(y-x) \geq 0$ with equality iff $y - x = 0$ since A is positive definite. Conclude that $x = A^{-1}b$ is the unique minimizer of P and

$$\begin{aligned}
P(x_{min}) = P(A^{-1}b) &= \tfrac{1}{2}(A^{-1}b)^TAA^{-1}b - (A^{-1}b)^Tb \\
&= \tfrac{1}{2}b^T(A^{-1})^Tb - b^T(A^{-1})^Tb \\
&= \tfrac{1}{2}b^TA^{-1}b - b^TA^{-1}b \\
&= -\tfrac{1}{2}b^TA^{-1}b
\end{aligned}$$

Use: $I = AA^{-1} = A^TA^{-1} \rightarrow (A^T)^{-1} = A^{-1}$

Alternatively,

$$\tfrac{1}{2}(x - A^{-1}b)^TA(x - A^{-1}b) - \tfrac{1}{2}b^TA^{-1}b$$

$-\frac{1}{2}b^TA^{-1}b$ constant wrt x. Writing $P(x)$ in this way shows $P(x)$ minimized when $x = A^{-1}b$ and $P_{min} = -\frac{1}{2}b^TA^{-1}b$.

$$= \tfrac{1}{2}x^TAx - \tfrac{1}{2}b^T(A^{-1})^TAx - \tfrac{1}{2}x^TAA^{-1}b + \tfrac{1}{2}b^T(A^{-1})^TAA^{-1}b - \tfrac{1}{2}b^TA^{-1}b$$

$$= \tfrac{1}{2}x^TAx - \tfrac{1}{2}b^T(A^{-1})^Tb - \tfrac{1}{2}x^Tb + 0 = \tfrac{1}{2}x^TAx - \tfrac{1}{2}x^Tb - \tfrac{1}{2}x^Tb = \tfrac{1}{2}x^TAx - x^Tb = P(x)$$