The Use of Linear Models in Teaching Linear Algebra

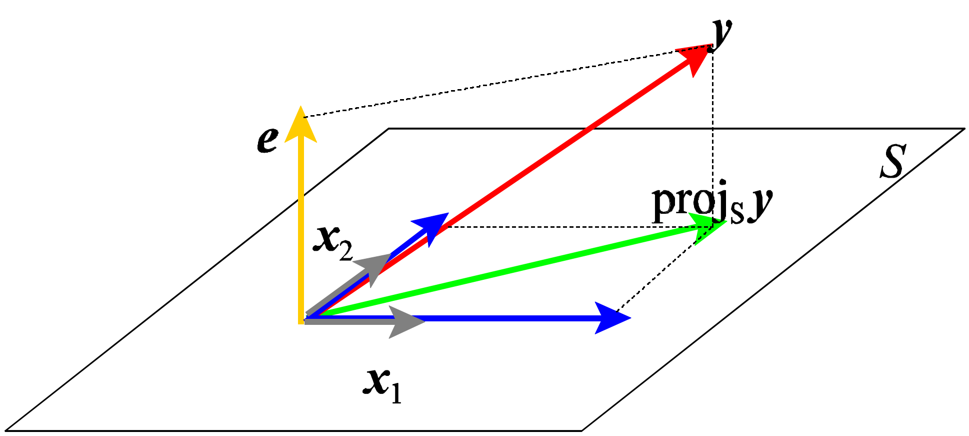
Jan Zijlstra

Department of Mathematical Sciences
Middle Tennessee State University

Linear Model:

$$\mathbf{y} = c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k + \mathbf{e} \quad (1)$$

where **y**: data vector
x: effect vectors
e: remainder/error vector



Linear Model:

$$\mathbf{y} = c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k + \mathbf{e} \quad (1)$$

where **y**: data vector

x: effect vectors

***e*:** remainder/error vector

In component form:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = c_1\begin{bmatrix} | \\ \mathbf{x}_1 \\ | \end{bmatrix} + \cdots + c_k\begin{bmatrix} | \\ \mathbf{x}_k \\ | \end{bmatrix} + \begin{bmatrix} | \\ \mathbf{e} \\ | \end{bmatrix}$$

Since

$$c_1\begin{bmatrix} | \\ \mathbf{x}_1 \\ | \end{bmatrix} + \cdots + c_k\begin{bmatrix} | \\ \mathbf{x}_k \\ | \end{bmatrix} = \begin{bmatrix} | & & | \\ \mathbf{x_1} & \cdots & \mathbf{x_k} \\ | & & | \end{bmatrix}\begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} = \mathbf{Xc}$$

Matrix form of model:

$$\mathbf{y} = \mathbf{Xc} + \mathbf{e},$$

an *over-determined* system - **X** is *not* invertible.

Requirement: c_1 , ... , c_k must be chosen such that the length of the error vector e is minimal:

$$\min_{\{c_1,\ldots,c_k\}} \|e\|_2 \text{ where } \|e\|_2 = \sqrt{e_1^2 + \cdots + e_n^2}$$

Since $\|e\|$ is the distance of y from S = span$\{x_1, \ldots ,x_k\}$, the vector *e* must be *perpendicular to effect vectors*:

$$\mathbf{e} \perp \mathbf{x}_i \quad \Leftrightarrow \quad \mathbf{e} \,.\, \mathbf{x}_i = 0, 1 \le i \le k \quad \Leftrightarrow \quad \mathbf{Xe} = \mathbf{0}$$

Multiplying (1) by $\mathbf{x}_i$:

$$\mathbf{y}.\mathbf{x}_i = c_1\mathbf{x}_1.\mathbf{x}_i + \cdots + c_k\mathbf{x}_k.\mathbf{x}_i + \underbrace{\mathbf{e}.\mathbf{x}_i}_{=0}, 1 \le i \le k$$

yields a square linear system

$$\begin{bmatrix} \mathbf{x}_1.\mathbf{x}_1 & \mathbf{x}_2.\mathbf{x}_1 & \cdots & \mathbf{x}_k.\mathbf{x}_1 \\ \mathbf{x}_2.\mathbf{x}_1 & \mathbf{x}_2.\mathbf{x}_2 & \cdots & \mathbf{x}_k.\mathbf{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_1.\mathbf{x}_k & \mathbf{x}_2.\mathbf{x}_k & \cdots & \mathbf{x}_k.\mathbf{x}_k \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} = \begin{bmatrix} \mathbf{y}.\mathbf{x}_1 \\ \vdots \\ \mathbf{y}.\mathbf{x}_k \end{bmatrix} \qquad (2)$$

In matrix form:

$$\begin{bmatrix} - \mathbf{x}_1 - \\ \vdots \\ - \mathbf{x}_k - \end{bmatrix} \begin{bmatrix} | & & | \\ \mathbf{x}_1 & \cdots & \mathbf{x}_k \\ | & & | \end{bmatrix} \mathbf{c} = \begin{bmatrix} - \mathbf{x}_1 - \\ \vdots \\ - \mathbf{x}_k - \end{bmatrix} \mathbf{y}$$

or

$$\mathbf{X}^T\mathbf{X}\mathbf{c} = \mathbf{X}^T\mathbf{y}$$

Solution of the system $\mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{c} = \mathbf{X}^{\mathbf{T}}\mathbf{y}$, provided $\mathbf{X}^{T}\mathbf{X}$ invertible

$$\mathbf{c} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}, \qquad (3)$$

' Least Squares Estimate' for vector ***c***.

Necessary and sufficient condition for $\boldsymbol{X}^{T}\boldsymbol{X}$ *to be invertible.*

LEMMA: *The homogeneous k × k linear system* $\boldsymbol{Ac} = \boldsymbol{0}$ *has a non-zero solution* $\boldsymbol{c} \neq \boldsymbol{0}$ *if and only if* $\det(\mathbf{A}) = 0$.

PROOF:

"⇒": Suppose $\mathbf{Ac} = \mathbf{0}$:

$$\mathbf{Ac} = \begin{bmatrix} | & & | \\ \mathbf{a_1} & \cdots & \mathbf{a_n} \\ | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = c_{\mathbf{1}}\mathbf{a_1} + \cdots + c_{\mathbf{n}}\mathbf{a_n} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

for $c_1, \ldots, c_n$ not all zero. But then, by definition, the columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ of **A** are linearly dependent, so det(**A**) must be zero.

"⇐": Suppose $\det(\mathbf{A}) = 0$. Then the columns of A are *linearly dependent*: there exist $c_1, \ldots, c_n$, not all zero, such that

$$c_{\mathbf{1}}\mathbf{a_1} + \cdots + c_{\mathbf{n}}\mathbf{a_n} = \begin{bmatrix} | & & | \\ \mathbf{a_1} & \cdots & \mathbf{a_n} \\ | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{Ac} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

□

THEOREM: For any $n \times k$ matrix $\mathbf{X} = \begin{bmatrix} | & & | \\ \mathbf{x_1} & \cdots & \mathbf{x}_k \\ | & & | \end{bmatrix}$, $n \geq k$,

det($\mathbf{X}^T\mathbf{X}$) is non-zero if and only if the column vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ of $\mathbf{X}$ are *linearly independent.*

"⇐": *If the set of column vectors* $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ *of matrix* $\mathbf{X}$ *is linearly independent, then* det($\mathbf{X}^T\mathbf{X}$) $\neq 0$.

PROOF: If the set $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is linearly independent, then

$$c_1\mathbf{x_1} + \cdots + c_k\mathbf{x}_k = 0 \quad \Rightarrow \quad c_1 = c_2 = \cdots = c_k = 0$$

in vector notation:

$$\mathbf{Xc} = \mathbf{0} \quad \Rightarrow \quad \mathbf{c} = \mathbf{0}$$

Multiplying the antecedent by $\mathbf{X}^T$:

$$\mathbf{X^T X c}\left(= \mathbf{X^T 0}\right) = \mathbf{0} \quad \Rightarrow \quad \mathbf{c} = \mathbf{0}$$

and so, by Lemma, det($\mathbf{X}^T\mathbf{X}$) $\neq 0$.

"⇒": *If* det($\mathbf{X}^T\mathbf{X}$) ≠ 0*, then the column vectors* $\mathbf{x}_1, \ldots, \mathbf{x}_k$ *of the* $n \times k$ *matrix* $\mathbf{X}$ ($n \geq k$) *are linearly independent*.

PROOF: Since det($\mathbf{X}^T\mathbf{X}$) ≠ 0, the square homogeneous system $\mathbf{X}^T\mathbf{X}\mathbf{c} = \mathbf{0}$ has, by lemma, only the trivial solution:

$$\mathbf{X}^T\mathbf{X}\mathbf{c} = \mathbf{0} \;\Rightarrow\; \mathbf{c} = \mathbf{0}$$

Writing out this implication:

$$\begin{bmatrix} - & \mathbf{x}_1 & - \\ & \vdots & \\ - & \mathbf{x}_k & - \end{bmatrix}\begin{bmatrix} | & & | \\ \mathbf{x}_1 & \cdots & \mathbf{x}_k \\ | & & | \end{bmatrix}\mathbf{c} = \mathbf{0} \;\Rightarrow\; \mathbf{c} = \mathbf{0}$$

$$\mathbf{X}^T\mathbf{X}\mathbf{c} = \begin{bmatrix} \mathbf{x_1}.\mathbf{x_1} & \mathbf{x_1}.\mathbf{x_2} & \cdots & \mathbf{x_1}.\mathbf{x}_k \\ \mathbf{x_2}.\mathbf{x_1} & \mathbf{x_2}.\mathbf{x_2} & \cdots & \mathbf{x_2}.\mathbf{x}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_k.\mathbf{x_1} & \mathbf{x}_k.\mathbf{x_2} & \cdots & \mathbf{x}_k.\mathbf{x}_k \end{bmatrix}\begin{bmatrix} c_1 \\ \vdots \\ \vdots \\ c_k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \;\Rightarrow\; \mathbf{c} = \mathbf{0},$$

Equivalent system of equations:

$$c_1(\mathbf{x_i}.\mathbf{x_1}) + \cdots + c_k(\mathbf{x_i}.\mathbf{x}_k) = 0, \text{ for } 1 \leq i \leq k \Rightarrow \mathbf{c} = \mathbf{0}$$

Factoring out $\mathbf{x}_i$:

$$\mathbf{x_i}.(c_1\mathbf{x_1} + \cdots + c_k\mathbf{x}_k) = 0, \quad 1 \leq i \leq k, \Rightarrow \mathbf{c} = \mathbf{0}$$

Since $\mathbf{x}_i \neq \mathbf{0}$, and

$$\mathbf{x}_i \in \text{span}(S) \text{, so } \mathbf{x_i} \not\perp c_1\mathbf{x_1} + \cdots + c_k\mathbf{x}_k,$$

it must be that

$$c_1\mathbf{x_1} + \cdots + c_k\mathbf{x}_k = \mathbf{0} \;\Rightarrow\; c_1 = c_2 = \cdots = c_k = 0,$$

and so, by definition, the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$. are *linearly independent.* □

(Alternate Proof:

Suppose $\det(\mathbf{X}^T\mathbf{X}) = 0$. Then, by lemma, the square homogeneous system $\mathbf{X}^T\mathbf{X}\mathbf{c} = \mathbf{0}$ for some $\mathbf{c} \neq \mathbf{0}$.

$$\mathbf{X}^T\mathbf{X}\mathbf{c} = \begin{bmatrix} \mathbf{x_1}.\mathbf{x_1} & \mathbf{x_1}.\mathbf{x_2} & \cdots & \mathbf{x_1}.\mathbf{x}_k \\ \mathbf{x_2}.\mathbf{x_1} & \mathbf{x_2}.\mathbf{x_2} & \cdots & \mathbf{x_2}.\mathbf{x}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_k.\mathbf{x_1} & \mathbf{x}_k.\mathbf{x_2} & \cdots & \mathbf{x}_k.\mathbf{x}_k \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ \vdots \\ c_k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix},$$

equivalent system of equations:

$$c_1(\mathbf{x_i}.\mathbf{x_1}) + c_2(\mathbf{x_i}.\mathbf{x_2}) + \cdots + c_n(\mathbf{x_i}.\mathbf{x}_k) = 0, \quad \text{for } 1 \leq i \leq k$$

Factoring out $\mathbf{x}_i$:

$$\mathbf{x_i}.\underbrace{\left(c_1\mathbf{x_1} + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k\right)}_{\mathbf{w}} = \mathbf{x}_i.\mathbf{w} = 0, \quad \text{for all } 1 \leq i \leq n,$$

so either $\mathbf{w} = 0$ (not so since $\boldsymbol{c} \neq \mathbf{0}$ and $\mathbf{x}_i \neq \mathbf{0}$),

or

$$\mathbf{x}_i.\mathbf{w} = 0, \text{ for all } 1 \leq i \leq k$$

that is, **w** <u>is perpendicular to all of the vectors $\mathbf{x}_i$</u>.
But if $\mathbf{w} \perp \mathbf{x}_i$, then **w** is *not* in span$\{\mathbf{x}_1, ..., \mathbf{x}_k\}$.
Since **w** is *defined* as $\mathbf{w} = c_1\mathbf{x}_1 + ... + c_k\mathbf{x}_k$, a linear combination of the $\mathbf{x}_i$, it *must* be in their span.

Contradiction: the assumption $\det(\mathbf{X}^T\mathbf{X}) = 0$ is wrong, and therefore $\det(\mathbf{X}^T\mathbf{X})$ must be *non*-zero. □

EXAMPLE:

Model: Scores on a test are thought to depend linearly on effects:

- ❑ study time (hrs),
- ❑ course load (sem hrs),
- ❑ off-campus work (hrs / wk)
  and
- ❑ course related computer use (0/1)

Score of the i^{th} student, y_i , is modeled as a *linear combination* of effects x_i, plus error term e_i:

$$y_i = c_1 x_{1i} + c_2 x_{2i} + c_3 x_{3i} + c_4 x_{4i} + e_i$$

Data:

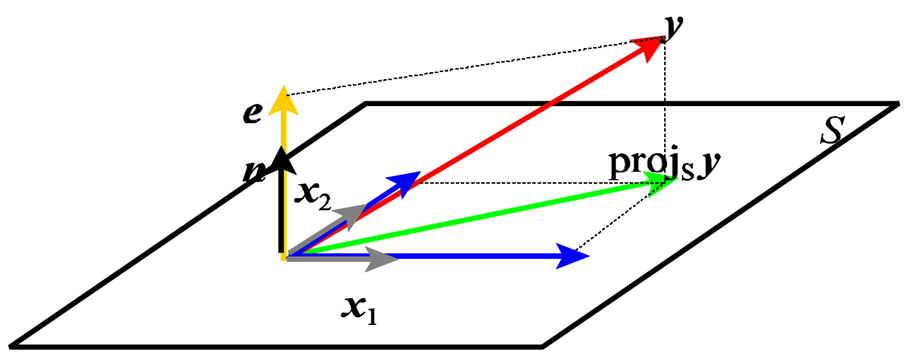
Test scores y	**Study time** x_1	**Course Load** x_2	**Work** x_3	**Computer Use** x_4	**Predicted grade**	**Error** e
78	2.5	18	30	0		
92	3.5	12	10	1		
57	1.5	18	35	0		
81	4.0	15	10	1		
89	3.5	9	12	1		
75	2.0	9	14	1		
89	3.5	12	15	1		
98	4.5	12	8	0		
73	2.0	18	20	0		
65	1.0	15	24	0		
74	2.5	15	18	0		
68	1.5	18	30	0		
Estimate for c_i	$c_1 \approx$	$c_2 \approx$	$c_3 \approx$	$c_4 \approx$		$e^t e =$

Results:

Test scores y	Study time x_1	Course Load x_2	Work x_3	Computer Use x_4	Predicted grade	Error e
78	2.5	18	30	0	83	-5
92	3.5	12	10	1	88	4
57	1.5	18	35	0	70	-13
81	4.0	15	10	1	99	-19
89	3.5	9	12	1	85	4
75	2.0	9	14	1	63	12
89	3.5	12	15	1	91	-2
98	4.5	12	8	0	92	6
73	2.0	18	20	0	68	5
65	1.0	15	24	0	51	14
74	2.5	15	18	0	71	3
68	1.5	18	30	0	67	1
Estimate for c_i	$c_1 \approx$ 15.78	$c_2 \approx$ 1.311	$c_3 \approx$ 0.6569	$c_4 \approx$ 10.17		$\mathbf{e}^t\mathbf{e} =$ 1002

$$\text{Coefficient estimates: } \mathbf{c} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} = \begin{bmatrix} 15.78 \\ 1311 \\ 0.6569 \\ 10.17 \end{bmatrix}$$

Algorithm for finding a vector *n* normal to the span of vectors $\{x_1 , ... , x_k\}$



We seek n such that $n \perp x_i \Rightarrow n \, . \, x_i = 0$ for $1 \leq i \leq k$

Equivalent system of equations:

$$\begin{array}{l} n_1 x_{11} + \cdots + n_k x_{1k} + n_{k+1} x_{1,k+1} + \cdots + n_m x_{1m} = 0 \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad \vdots \\ n_1 x_{k1} + \cdots + n_k x_{kk} + n_{k+1} x_{k,k+1} + \cdots + n_m x_{km} = 0 \end{array}$$

'Squaring' this system:

$$\begin{array}{l} n_1 x_{11} + \cdots + n_k x_{1k} = - n_{k+1} x_{1,k+1} - \cdots - n_m x_{1m} \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ n_1 x_{k1} + \cdots + n_k x_{kk} = - n_{k+1} x_{k,k+1} - \cdots - n_m x_{km} \end{array}$$

from which we may determine the normal vector n:

THEOREM: If the $m \times k$ matrix **X**, $m > k$:

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{k,1} & \cdots & x_{k,k} \\ x_{k+1,1} & \cdots & x_{k+1,m} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mk} \end{bmatrix}$$

has an invertible submatrix $\mathbf{X}_{trunc} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{k1} & \cdots & x_{kk} \end{bmatrix}$

then the vector n whose entries are given by:

$$\begin{bmatrix} n_1 \\ \vdots \\ n_k \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{k1} & \cdots & x_{kk} \end{bmatrix}^{-T} \begin{bmatrix} -\sum_{i=1}^{m-k} x_{k+i,1} \\ \vdots \\ -\sum_{i=1}^{m-k} x_{k+i,k} \end{bmatrix}, \quad n_{k+1} = \cdots = n_m = 1$$

is a normal vector to the columns of **X**.

EXAMPLE : Find a vector n normal to the columns of **X**:

$$X = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 2 & 4 \\ 2 & 4 & 1 \\ 3 & 9 & -2 \\ \underbrace{5}_{x_1} & \underbrace{-3}_{x_2} & \underbrace{6}_{x_3} \end{bmatrix} \rightarrow \quad n = \begin{bmatrix} 10 \\ -5 \\ -4 \\ 1 \\ 1 \end{bmatrix}$$

since

$$n_{1,2,3} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 2 & 4 \\ 2 & 4 & 1 \end{bmatrix}^{-T} \begin{bmatrix} -(3+5) \\ -(9-3) \\ -(-2+6) \end{bmatrix} = \begin{bmatrix} 10 \\ -5 \\ -4 \end{bmatrix}$$

and the remaining entries: $n_4 = n_5 = 1.$

$$\text{Check:} \quad n.x_1 = \begin{bmatrix} 10 \\ -5 \\ -4 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \\ 5 \end{bmatrix} = 0, \quad n.x_2 = \begin{bmatrix} 10 \\ -5 \\ -4 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 4 \\ 9 \\ -3 \end{bmatrix} = 0, \quad n.x_3 = \begin{bmatrix} 10 \\ -5 \\ -4 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 1 \\ -2 \\ 6 \end{bmatrix} = 0$$