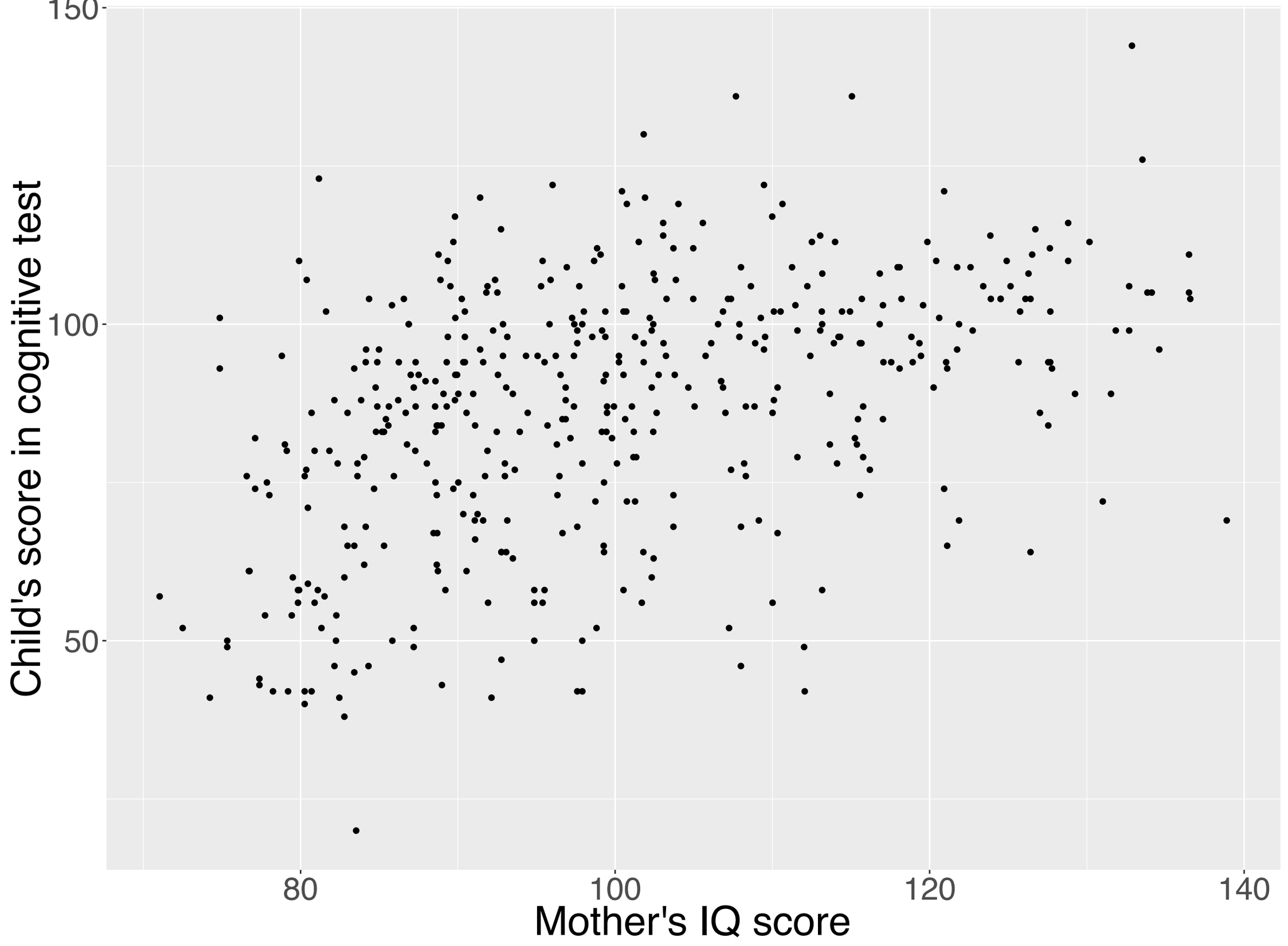

Child's score in cognitive test
150
100
50
80
100
120
140
Mother's IQ score


Child's score in cognitive test
150
100
50
80
100
120
140
Mother's IQ score

Toy Example: Linear regression in statistics

Find the straight line that best fits given points (x, b_x): (0,1), (1,9), (3, 9) and (4,21)

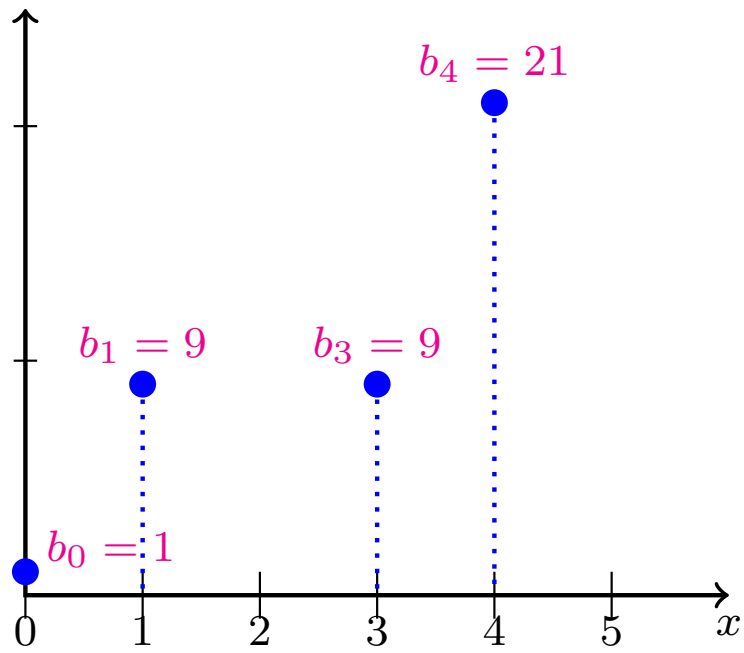
The equation for a straight line is

$$b_x = u_1 + u_2 x$$

so have 4 equations and two unknowns, u_1 and u_2.

$$\mathbf{Ax} = \mathbf{u} \Leftrightarrow \begin{array}{rcl} u_1 + u_2 \cdot 0 & = & 1 \\ u_1 + u_2 \cdot 1 & = & 9 \\ u_1 + u_2 \cdot 2 & = & 9 \\ u_1 + u_2 \cdot 3 & = & 21 \end{array} \Leftrightarrow$$

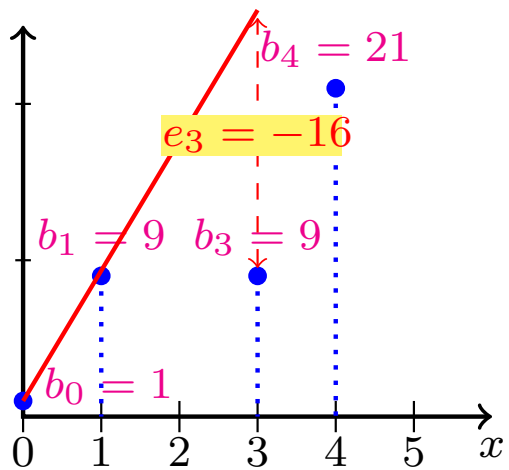
$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 9 \\ 21 \end{bmatrix}$$



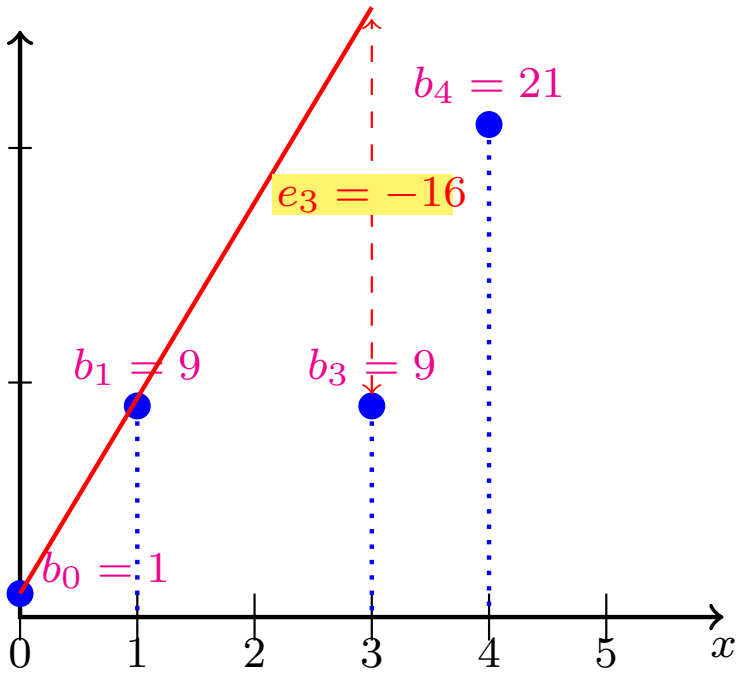
No line fits all the points since $\mathbf{b}$ is not a linear combination of the two column vectors from $\mathbf{A}$:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 9 \\ 9 \\ 21 \end{bmatrix}$$

Try candidate line through first two points, $1 + 8x = b_x$, i.e., $u_1 = 1, u_2 = 8$.



- Let $e_x = b_x - (1 + 8x)$ between the observed point and position of line at x



$$\underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}}_{\mathbf{u}} = \underbrace{\begin{bmatrix} 1 \\ 9 \\ 9 \\ 21 \end{bmatrix}}_{\mathbf{b}} \Rightarrow \underbrace{\mathbf{e} = \mathbf{b} - \mathbf{A}\mathbf{u}}_{\text{residual error}}$$

Total squared error $E(\mathbf{u}) = \mathbf{e}^{\mathsf{T}}\mathbf{e} \equiv \| \mathbf{e} \|^2$

$$= (\mathbf{b} - \mathbf{A}\mathbf{u})^{\mathsf{T}}(\mathbf{b} - \mathbf{A}\mathbf{u})$$

For this line squared error is $E(\mathbf{u}) = 0 + 0 + 256 + 144 = \mathbf{400}$.

Finding best fit line

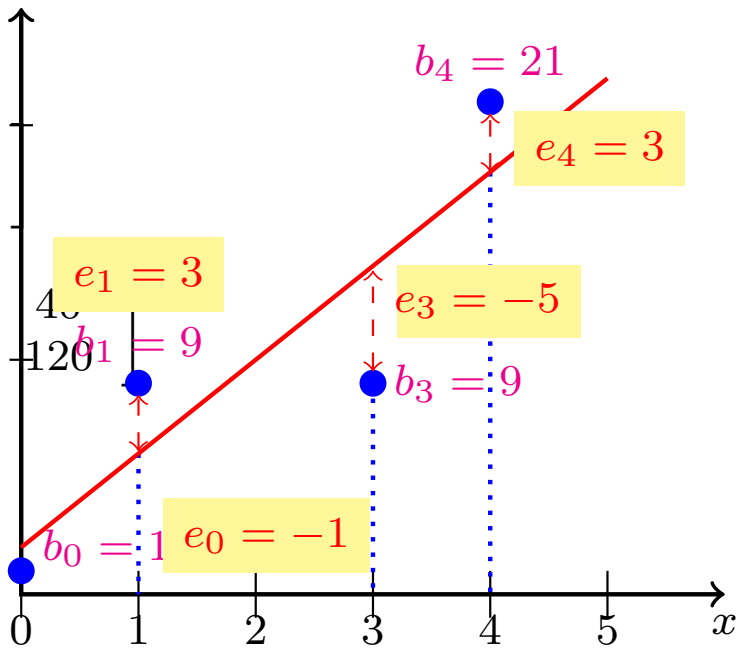
The normal equation $\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{u}^* = \mathbf{A}^\mathsf{T}\mathbf{b}$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} u_1^* \\ u_2^* \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 9 \\ 9 \\ 21 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix} \begin{bmatrix} u_1^* \\ u_2^* \end{bmatrix} = \begin{bmatrix} 40 \\ 120 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} u_1^* \\ u_2^* \end{bmatrix} = \frac{1}{40} \begin{bmatrix} 26 & -8 \\ -8 & 4 \end{bmatrix} \begin{bmatrix} 40 \\ 120 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} u_1^* \\ u_2^* \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$



Error for the best line: $E(\mathbf{u}^*) = 1 + 9 + 25 + 9 = 44$