Como no todas las x_i son iguales, las columnas de A son linealmente independientes. Ahora bien

$$A^\top A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}$$

Si $A^\top A$ no es invertible, entonces det $A^\top A = 0$. Esto significa que

$$n\sum_{i=1}^{n} x_i^2 = \left(\sum_{i=1}^{n} x_i\right)^2 \quad \textbf{(6.2.13)}$$

Sea $\mathbf{u} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ y $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$. Entonces

$$|\mathbf{u}|^2 = \mathbf{u} \cdot \mathbf{u} = n, \quad |\mathbf{x}|^2 = \sum_{i=1}^{n} x_i^2 \quad \text{y} \quad \mathbf{u} \cdot \mathbf{x} = \sum_{i=1}^{n} x_i$$

de manera que la ecuación (6.2.13) se puede establecer como

$$|\mathbf{u}||\mathbf{x}|^2 = |\mathbf{u} \cdot \mathbf{x}|^2$$

y sacando raíz cuadrada se obtiene

$$|\mathbf{u} \cdot \mathbf{x}| = |\mathbf{u}||\mathbf{x}|$$

Ahora, la desigualdad de Cauchy-Schwarz (página 418) dice que $|\mathbf{u} \cdot \mathbf{x}| \leq |\mathbf{u}||\mathbf{x}|$ en donde la igualdad se cumple si y sólo si $\mathbf{x}$ es una constante múltiplo de $\mathbf{u}$. Pero $\mathbf{u}$ y $\mathbf{x}$ son, por hipótesis, las columnas de A que son linealmente independientes. Esta contradicción prueba el teorema.

RESUMEN 6.2

- Sea $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ un conjunto de datos. Si se quiere representar estos datos por la recta $y = mx + b$, entonces el **problema de mínimos cuadrados** es encontrar los valores de m y b que minimizan la suma de los cuadrados

$$[y_1 - (b + mx_1)]^2 + [y_2 - (b + mx_2)]^2 + \ldots + [y_n - (b + mx_n)]^2$$

La solución a este problema es establecer

$$\begin{pmatrix} b \\ m \end{pmatrix} = \bar{\mathbf{u}} = (A^\top A)^{-1} A^\top \mathbf{y}$$

donde

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \text{y} \quad A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

Resultados similares se aplican cuando se quiere representar los datos usando un polinomio de grado > 1.