ORTHOGONAL PROJECTIONS WORKSHEET

NOVEMBER 3, 2017

Suppose we are given data points

$$(x_1, y_1) = (-3, 1), \quad (x_2, y_2) = (-2, 2), \quad (x_3, y_3) = (1, 0), \quad (x_4, y_4) = (4, 3)\,.$$

In this problem, you will find the least-squares regression line (sometimes called the line of best fit) for this data. The goal is determine values of m and b such that the line $y = mx + b$ best approximates the data.

(a) Let

$$A = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} m \\ b \end{pmatrix}, \qquad \text{and} \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}.$$

Show that the system $A\mathbf{x} = \mathbf{y}$ is inconsistent.

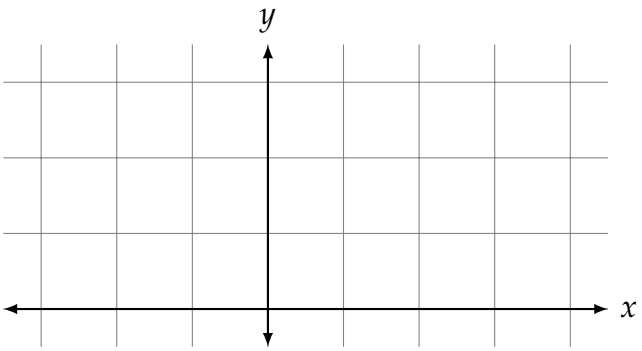
(b) Show that the columns of A form an orthogonal basis for $\mathrm{Col}(A)$.

(c) Compute the projection $\widehat{\mathbf{y}} = \text{proj}_{\text{Col}(A)}(\mathbf{y})$.

(d) Find a solution to the system $A\widehat{\mathbf{x}} = \widehat{\mathbf{y}}$. (*Hint*: You shouldn't have to row reduce! Use your calculations from the previous part.)

(e) Compute the distance $\|A\widehat{\mathbf{x}} - \mathbf{y}\|$.

(f) Plot the data points and your line on the grid below. (*Hint*: Your line should pass through the point (x_1, y_1).)



(g) Explain why the values of m and b computed in part (c) minimize the sum

$$\sum_{i=1}^{4}(y_i - (mx_i + b))^2$$

of the squared residuals. It is in this sense that the line $y = mx + b$ is the line of best fit for the data.