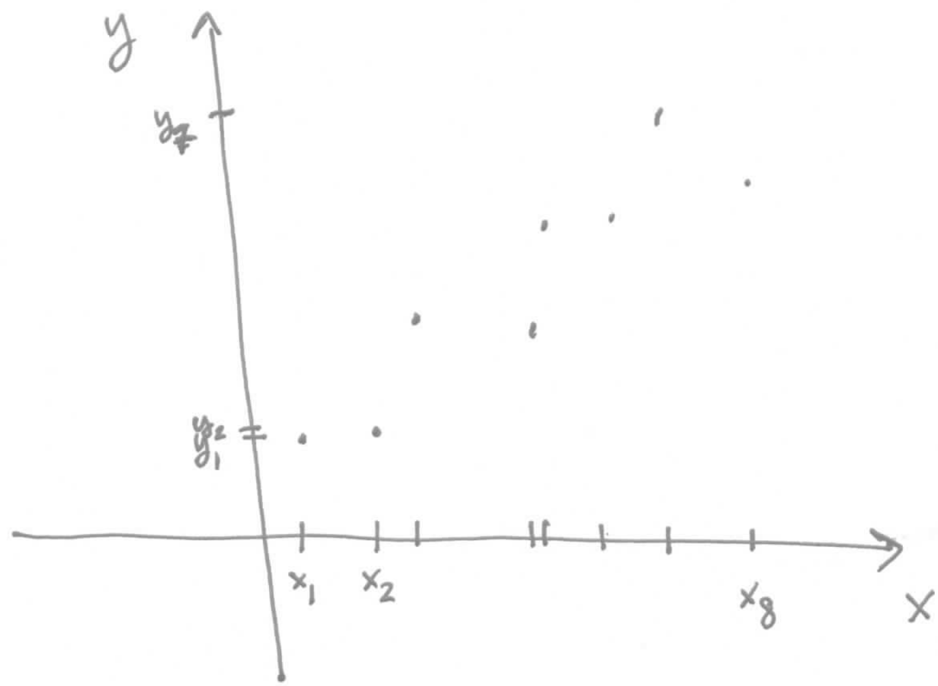
Least squares



find the line

$$y = ax + b$$

that best fits the data?

One answer:

minimize $\displaystyle\sum_{i=1}^{8} (y_i - (ax_i + b))^2$

with respect to a & b.

if A denotes the matrix

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_8 & 1 \end{bmatrix} \quad \text{and } ~~ag~~ \quad \underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_8 \end{bmatrix}$$

then this is equivalent to

$$\boxed{\text{minimize } \left\| A \begin{bmatrix} a \\ b \end{bmatrix} - \underline{y} \right\|_2^2 \quad \text{wrt} \quad \begin{bmatrix} a \\ b \end{bmatrix}}$$

In general:

(1) $$\text{minimize } \| A\underline{c} - \underline{y} \|_2^2 \quad \text{wrt to } \underline{c}$$

where $A = m \times n$, $\underline{c} \in \mathbb{R}^n$ & $\underline{y} \in \mathbb{R}^m$

$$\| A\underline{c} - \underline{y} \|_2^2 = (A\underline{c} - \underline{y})^T (A\underline{c} - \underline{y}) =$$

$$= c^T A^T A c - y^T A c - c^T A^T y - y^T y$$

$$= c^T A^T A c - 2 c^T A^T y - y^T y \qquad \text{(why?)}$$

$c_i (A^T A)_{ij} c_j$ $\qquad$ $2 c_i (A^T y)_i$ $\qquad$ with summation convention!

now require $\frac{\partial}{\partial c_k} = 0$, resulting in

$$\underline{(A^T A)_{kj} c_j + c_i (A^T A)_{ik} - 2 (A^T y)_k = 0} \qquad \underline{k = 1 \dots n}$$

or

$$2 \sum_{j=1}^{m} (A^T A)_{kj} c_j - 2 (A^T y)_k = 0 \qquad \underline{k = 1 \dots n}$$

<u>i.e.</u>

$$(2) \quad \boxed{A^T A \underline{c} = A^T \underline{y}}$$

the so-called normal equation for (1).

In terms of the linear regression problem we started out with

$$A^T A = \begin{bmatrix} \sum_{i=1}^{8} x_i^2 & \sum_{i=1}^{8} x_i \\ \sum_{i=1}^{8} x_i & 8 \end{bmatrix} \qquad A^T \underline{y} = \begin{bmatrix} \sum_{i=1}^{8} x_i y_i \\ \sum_{i=1}^{8} y_i \end{bmatrix}$$

$$A^T A \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix}$$

has the solution

$$\left[\begin{array}{l} a = \dfrac{1}{\left(8 \sum x_i^2 - \left(\sum x_i\right)^2\right)} \left(8 \sum x_i y_i - \sum y_i \sum x_i\right) \\ b = \dfrac{1}{8}\left(\sum y_i - \sum x_i \, a\right) \end{array} \right.$$

Note

$$\left(\sum x_i\right)^2 = \left|\sum x_i\right|^2 \le \left(\sum |x_i|\right)^2$$

$$\underset{\substack{\uparrow \\ \text{why?}}}{\le} \sum |x_i|^2 \cdot 8 = 8 \sum x_i^2$$

and equality only holds if
all x_i are the same (why?)

Going back to 2 — this
equation always has a solution
because $R(A^TA) = R(A^T) \subseteq \mathbb{R}^n$

Here is why:

$$R(A^TA)^\perp = N(A^TA) = N(A)$$
$$= R(A^T)^\perp$$

and so

$$R(A^TA) = R(A^T)$$

In terms of our linear
regression problem — we could
take $\boxed{b = \frac{1}{8}\sum y_i \text{ and } a = 0}$ when x_i are all the same

When is the solution to (2) unique

Answer : when $N(A^TA) = \{0\}$, i.e.,

when $N(A) = \{0\}$ or when

all columns of A are linearly independent

(since we typically have $m \geq n$

this happens when A has full rank).

In our linear regression problem this happens exactly when the x_i's are not all the same.

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_n & 1 \end{bmatrix}$$

let $\underline{c}$ denote a solution to 2

then and $\tilde{\underline{c}}$ an arbitrary vector

$$\|A\tilde{\underline{c}} - \underline{y}\|_2^2 = \|A(\tilde{\underline{c}} - \underline{c}) + A\underline{c} - \underline{y}\|_2^2$$

$$= \|A(\tilde{\underline{c}} - \underline{c})\|_2^2 + \|A\underline{c} - \underline{y}\|_2^2$$

$$+ 2[A(\tilde{\underline{c}} - \underline{c})]^T [A\underline{c} - \underline{y}]$$

$$= \underset{\geq 0}{\|A(\tilde{\underline{c}} - \underline{c})\|_2^2} + \|A\underline{c} - \underline{y}\|_2^2$$

why?

this is another way to see that Solutions to 2 minimize the ~~least~~ $\|\cdot\|_2$ fit of $A\underline{c}$ to $\underline{y}$