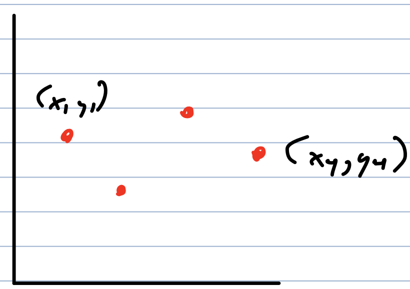
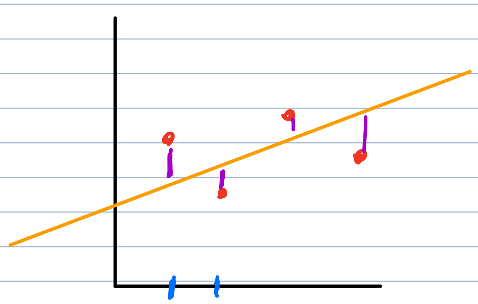
Section 6.6

Recall intro from last time. (Will introduce some different notation along the way)

Have data (measurements/observations) with coordinates (x_i, y_i)



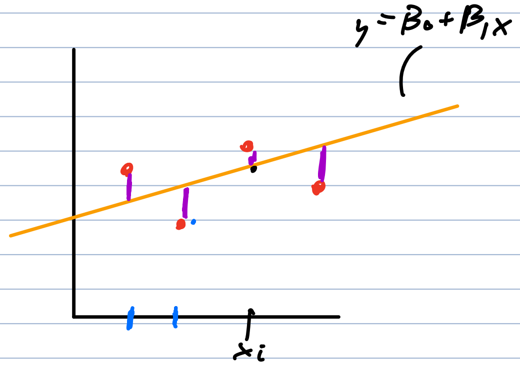
Would like a line (function) that best explains data



Instead of $y = mx + b$ will use notation

$$y = \beta_0 + \beta_1 x$$

Distance from line to point is called the residual



As in last section we want to minimize error

Again measure error using square terms

line at specific pt x_i

$$\bigstar \quad (\beta_0 + x_i \beta_1) - (y_i) = \text{residual}_i$$

i.e. predicted y value at x_i — Actual/observed value of y at x_i

$$\text{Error} = \sum (\text{residual}_i)^2$$

Goal: Choose line (so choose β_0, β_1) so as to minimize error, i.e. find the least squares line / linear regression

Note: β_0, β_1 are called regression coefficients

<u>If</u> every point was on a line, we would have system of equations

predicted y-values — observed y-values

$$\begin{cases} \beta_0 + \beta_1 x_1 = y_1 \\ \beta_0 + \beta_1 x_2 = y_2 \\ \vdots \\ \beta_0 + \beta_1 x_n = y_n \end{cases}$$

Write as matrix equation

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Inconsistent?

$n \times 2$ matrix

2×1 vector
regression coefficients

$n \times 1$ vector
observed y's

New Notation: $X\vec{\beta} = \vec{y}$ instead of $A\vec{x} = \vec{b}$

Real world situation: Will have data, no line
Want to find line of best fit, i.e. find
β_0, β_1

Ideally, want β_0, β_1 that solve matrix equation:

$$\star \quad \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

1's known inputs known outputs

What if system is inconsistent? Section 6.5!
Now this is a least squares problem we
know how to solve

Short Story: Solve system $X^T X \vec{\beta} = X^T \vec{y}$

☆ Example / Problem break

Survey of Topics

A modification: In **design matrix** X, may change the first column to get "nicer" matrix

① Let $\bar{x}$ = average of observed <u>inputs</u> <u>$x_1, x_2 \ldots x_n$</u>

② Then consider equations

$$y = \beta_0 \bar{x} + \underline{\beta_1} (\bar{x} - x_i) \quad ☆$$

③ Get matrix equation

$$\begin{bmatrix} \bar{x} & (\bar{x} - x_1) \\ \bar{x} & (\bar{x} - x_2) \\ \bar{x} & \vdots \\ \vdots & \\ \bar{x} & (\bar{x} - x_n) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Why? Columns will be <u>orthogonal</u>, and orthogonal nice to work with.

General Linear Model

Have been considering $\vec{y} = X\vec{\beta}$ as an approximation, and this is accurate, want to minimize error

Another way to write this:

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon}$$

error term
called "residual vector"

Then by rearranging see that ε is

$$\vec{\varepsilon} = (\vec{y} - X\vec{\beta})$$

residual vector = observed values − predicted values

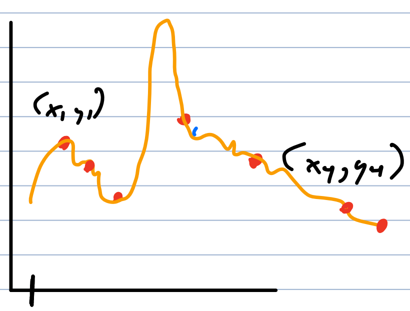
error = ||residual vector|| = $\|\vec{y} - X\vec{\beta}\|$ minimize this

If we start with system of equations

$$y_1 = \beta_0 + \beta_1 x_1$$
$$\vdots$$
$$y_n = \beta_0 + \beta_1 x_n$$

then nothing has changed. It is exactly set up we had before, just slightly different notation.

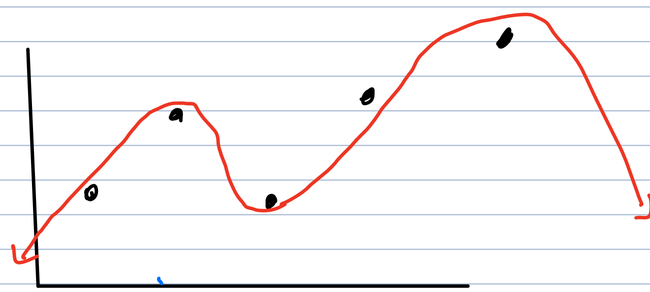
But maybe we want to begin from different set up.



Sometimes no line is a good approximation to the "real" function that governs our data a D

outputs Input

$$\left.\begin{array}{l} y_1 = \beta_0 + \beta_1 f(x_1) + \underline{\beta_2}\, g(x_1) + \underline{\beta_3}\, h(x_1) + \beta_4\, i(x_1) \\ \quad \vdots \\ y_n = \beta_0 + \beta_1 \quad + \beta_2 x_n^2 + \beta_3 x_n^3 + \beta_4 x_n^4 \end{array}\right]$$



$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 \\ 1 & x_3 & x_3^2 & x_3^3 & x_3^4 \\ 1 & x_4 & x_4^2 & x_4^3 & x_4^4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

polynomial of form

$$y = \underline{\beta_0} + \underline{\beta_1}\, x + \underline{\beta_2}\, x^2 + \underline{\beta_3}\, x^3 + \underline{\beta_4}\, x^4$$

$$\left[X^T X \mid X^T y \right]$$

Weierstrass Approximation Theorem:

Any continuous function on interval $[a, b]$ can be approximated arbitrarily well by a polynomial.

$$(x_1, y_1)$$

$$(x_1, y_1, z_1)$$