Wed Apr 11

- 6.6 Fitting data to "linear" models.

Announcements: 6.6 HW: (1) (7) (and graph the points and best parabolic fit in #7, with technology

+ another

- handout

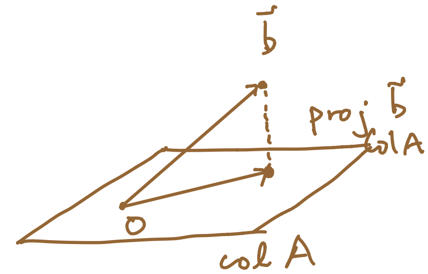
'til 12:56

Warm-up Exercise: a) Find the least squares sol'n to

$$\begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} = \begin{bmatrix}2\\0\\0\end{bmatrix}$$

$\uparrow A$

b) What is $\text{proj}_{col A}\begin{bmatrix}2\\0\\0\end{bmatrix}$?



$A\hat{x} = \text{proj}_{ColA}\vec{b}$

$A\vec{x} = \vec{b}$

$A^T A\vec{x} = A^T\vec{b}$

gives least squares sol'n

$$\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} = A^T\begin{bmatrix}2\\0\\0\end{bmatrix}$$

$$\begin{bmatrix}2 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} = \begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}\begin{bmatrix}2\\0\\0\end{bmatrix} = \begin{bmatrix}2\\0\end{bmatrix}$$

$2x_1 = 2$

$x_2 = 0$

$\hat{x} = \begin{bmatrix}1\\0\end{bmatrix}.$

b) $A\hat{x} = A\begin{bmatrix}1\\0\end{bmatrix} = \text{proj}_{ColA}\vec{b} = \begin{bmatrix}1\\0\\1\end{bmatrix}$

$$\text{Col}\,A = \text{span}\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$$

$$= \text{span}\left\{\underbrace{\tfrac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}}_{\vec{u}_1}, \underbrace{\begin{bmatrix}0\\1\\0\end{bmatrix}}_{\vec{u}_2}\right\}$$

Note Col's $\perp$ in this example.

$$\text{proj}_{ColA}\vec{b} = (\vec{b}\cdot\vec{u}_1)\vec{u}_1 + (\vec{b}\cdot\vec{u}_2)\vec{u}_2$$

$$= \left(\begin{bmatrix}2\\0\\0\end{bmatrix}\cdot\tfrac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}\right)\tfrac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}$$

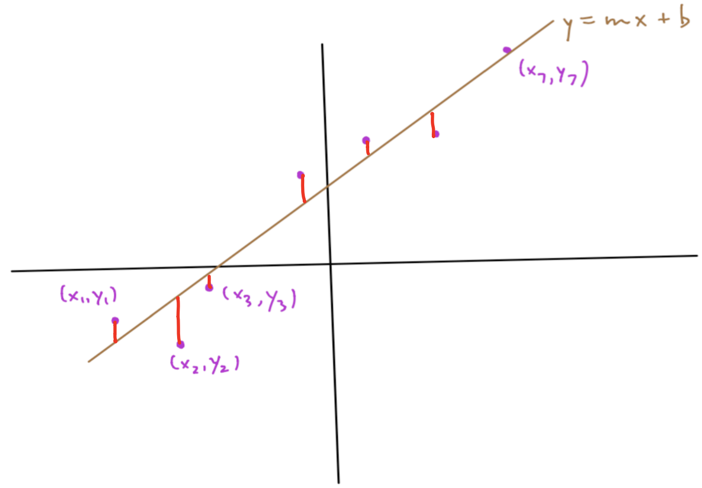
$$+ \left(\begin{bmatrix}2\\0\\0\end{bmatrix}\cdot\begin{bmatrix}0\\1\\0\end{bmatrix}\right)\begin{bmatrix}0\\1\\0\end{bmatrix} = \frac{2}{2}\begin{bmatrix}1\\0\\1\end{bmatrix} + \vec{0}$$

Applications of least-squares to data fitting.

- Find the best line formula $y = m\,x + b$ to fit n data points $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. We seek $\begin{bmatrix} m \\ b \end{bmatrix}$ so that

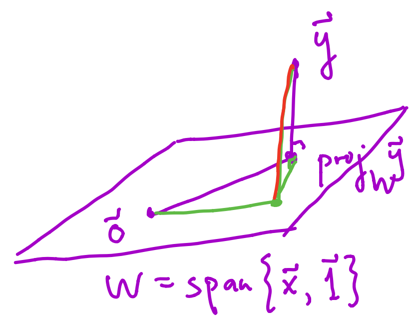
$y_i = m x_i + b$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = m \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$



In matrix form, find $\begin{bmatrix} m \\ b \end{bmatrix}$ so that

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}. \qquad A \begin{bmatrix} m \\ b \end{bmatrix} = \mathbf{y}.$$



There is no exact solution unless all the data points are actually on a single line!

$A\begin{bmatrix} \hat{m} \\ \hat{b} \end{bmatrix} = \text{proj}_W \vec{y}$

<u>Least squares solution</u>:

$$A^T A \begin{bmatrix} m \\ b \end{bmatrix} = A^T \mathbf{y}.$$

$\|\vec{z}\|^2 = \left\| y - A\begin{bmatrix} \hat{m} \\ \hat{b} \end{bmatrix} \right\|^2$

is the smallest that $\|\vec{y} - A\vec{x}\|^2$ can be.

minimized the sum of the squared vertical deviating

$$\vec{y} - A\begin{bmatrix} \hat{m} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - m \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - b \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} y_1 - m x_1 - b \\ y_2 - m x_2 - b \\ \vdots \\ y_n - m x_n - b \end{bmatrix}$$

$$A^T A \begin{bmatrix} m \\ b \end{bmatrix} = A^T \underline{y}$$

As long as the columns of A are linearly independent (i.e.at least two different values for x_j) there is a unique solution $[m, b]^T$. Furthermore, you are actually solving

$$A\,\underline{x} = proj_W \underline{y}$$

where

$$W = span\left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right\},$$

so

$$\left\| \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - m \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - b \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right\|^2$$

is as small as possible. In other words, you've minimized the sum of the *squared vertical deviations* from points on the line to the data points,

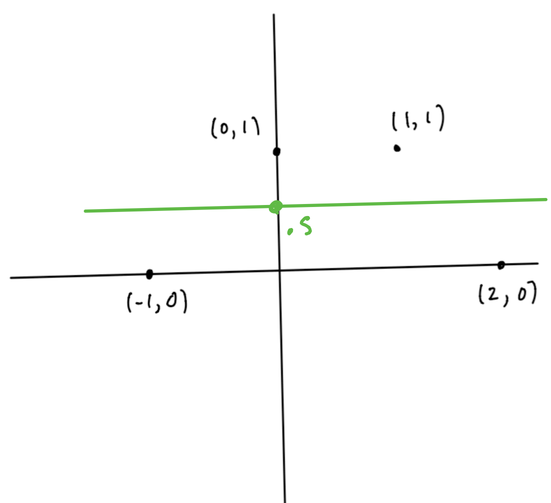
$$\sum_{i=1}^{n} \left(y_i - mx_i - b_i\right)^2.$$

$x_1 \quad x_2 \quad x_3 \quad x_4$

Exercise 1   Find the least squares line fit for the 4 data points $\{(-1, 0), (0, 1), (1, 1), (2, 0)\}$. Sketch.

$y_1 \quad y_2 \quad y_3 \quad y_4$

$$m \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$



$$\begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\boxed{\begin{aligned} m &= 0 \\ b &= .5 \\ y &= 0x + .5 \end{aligned}}$$

$$\begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} ; \quad \begin{bmatrix} m \\ b \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 4 & -2 \\ -2 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 0 \\ 10 \end{bmatrix} = \begin{bmatrix} 0 \\ .5 \end{bmatrix}$$

Example 2 Find the best quadratic fit to the same four data points. This is still a "linear" model!! In other words, we're looking for the best quadratic function

$$p(x) = c_0 + c_1 x + c_2 x^2$$

to fit to the four data points

$$\begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ \{(-1,0), & (0,1), & (1,1), & (2,0)\}. \\ y_1 & y_2 & y_3 & y_4 \end{array}$$

We want to solve

$$c_0 \begin{bmatrix} 1 \\ 1 \\ : \\ 1 \end{bmatrix} + c_1 \begin{bmatrix} x_1 \\ x_2 \\ : \\ x_n \end{bmatrix} + c_2 \begin{bmatrix} x_1^2 \\ x_2^2 \\ : \\ x_n^2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ : \\ y_n \end{bmatrix}.$$

For our example this is the system

$$\begin{array}{ccccccc} & & & x_i & & x_i^2 & & y_i \\ c_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & + & c_1 & \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} & + c_2 & \begin{bmatrix} 1 \\ 0 \\ 1 \\ 4 \end{bmatrix} & = & \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}. \end{array}$$

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

$A\vec{c} = \vec{y}$

$(A^T A)\vec{c} = A^T\vec{y}$

$\vec{c} = (A^T A)^{-1} A^T \vec{y}$

I used technology (Maple, with which I write these notes), and the least squares normal equation , see next page...

$$A^T A\, \underline{\boldsymbol{c}} = A^T \underline{\boldsymbol{b}}\,.$$

> $with(LinearAlgebra):$

$$C := \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} : b := \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} :$$

$c := (Transpose(C).C)^{-1}.Transpose(C).b;$

$$c := \begin{bmatrix} 1 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} \begin{array}{l} c_0 \\ c_1 \\ c_2 \end{array} \qquad p(x) = c_0 + c_1 x + c_2 x^2 \qquad (1)$$

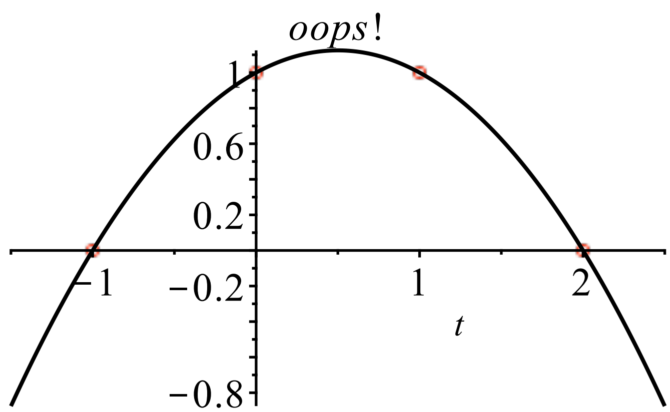
$p(x) = 1 + .5x - .5x^2$

> $with(plots):$

$plot1 := plot\left(1 + .5 \cdot t - .5 \cdot t^2, t = -1.5 ..2.5, color = black\right):$

$plot2 := pointplot(\{[-1, 0], [0, 1], [1, 1], [2, 0]\}, color = red, symbol = circle, symbolsize = 18):$

$display(\{plot1, plot2\}, title = 'oops!');$



$$BMI = C \cancel{\frac{\text{height}^2}{\text{weight}}}$$
$$= C \, \frac{\text{weight}}{\text{height}^2}$$

Math 2270-004
April 11, 2018
Applying least squares linear regression to obtain power law fits

How do you test for power laws?

Suppose you have a collection of n data points

$$[[x_1, y_1], [x_2, y_2], [x_3, y_3], \ldots, [x_n, y_n]]$$

and you expect there may be a good power-law fit

$$\boxed{y = b\, x^m}$$

which approximately explains how the y_i's are related to the x_i's. You would like to find the "best possible" values for b and m to make this fit. It turns out, if you take the ln-ln data, your power law question is actually just a best-line fit question:
Taking (natural) logarithms of the proposed power law yields

$$\ln(y) = \ln(b) + m \ln(x).$$

So, if we write $Y = \ln(y)$ and $X = \ln(x)$, $B = \ln(b)$, this becomes the equation of a line in the new variables X and Y:

$$Y = m X + B \qquad e^Y = y \quad e^B = b$$

Thus, in order for there to be a power law for the original data, the ln-ln data should (approximately) satisfy the equation of a line, and vise verse. If we get a good line fit to the ln-ln data, then the slope m of this line is the power relating the original data, and the exponential e^B of the Y-intercept is the proportionality constant b in the original relation $y = b\, x^m$. With real data it is not too hard to see if the ln-ln data is well approximated by a line, in which case the original data is well-approximated by a power law.

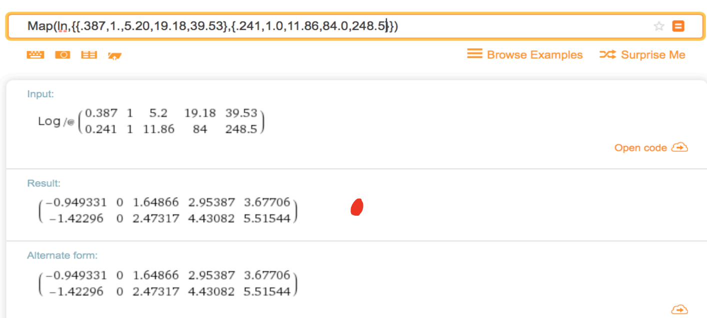
Astronomical example As you may know, Isaac Newton was motivated by Kepler's (observed) Laws of planetary motion to discover the notions of velocity and acceleration, i.e. differential calculus and then integral calculus, along with the inverse square law of planetary acceleration around the sun.....from which he deduced the concepts of mass and force, and that the universal inverse square law for gravitatonal attraction was the ONLY force law depending only on distance between objects, which was consistent with Kepler's observations! Kepler's three observations were that

(1) Planets orbit the sun in ellipses, with the sun at one of the ellipse foci.
(2) A planet sweeps out equal areas from the sun, in equal time intervals, independently of where it is in its orbit.
(3) The square of the period of a planetary orbit is directly proportional to the cube of the orbit's semi-major axis. $T^2 = kR^3$ $\quad T = kR^{3/2}$

So, for roughly circular orbits, Keplers third law translates to the statement that the period t is related to the radius r, by the equation $t = b\,r^{1.5}$, for some proportionality constant b. Let's see if that's consistent with the following data:

Planet	mean distance r from sun (in astronomical units where 1=dist to earth)	Orbital period t (in earth years)
Mercury	0.387	0.241
Earth	1.	1.
Jupiter	5.20	11.86
Uranus	19.18	84.0
Pluto	39.53	248.5

Taking the (natural) logarithm of the data points, as put into a matrix, using Wolfram alpha.



ln(radius) ln(period)

We want the least squares solution to the ln-ln data, $Y = m\,X + B$

$$\begin{bmatrix} -.9493 & 1 \\ 0 & 1 \\ 1.64866 & 1 \\ 2.95387 & 1 \\ 3.67706 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} -1.42296 \\ 0 \\ 2.47317 \\ 4.43082 \\ 5.51544 \end{bmatrix}.$$

$$A\,\hat{\underline{\mathbf{x}}} = \hat{\underline{\mathbf{b}}}$$

$$A^T A\,\underline{\mathbf{x}} = A^T\,\underline{\mathbf{b}}$$

$$\widehat{\begin{bmatrix} m \\ b \end{bmatrix}} \qquad \hat{\underline{\mathbf{x}}} = \left(A^T A\right)^{-1} A^T\,\underline{\mathbf{b}}$$

I didn't have time (yet) to do these steps neatly at Wolfram alpha. In Maple:

> $with(LinearAlgebra):$

$$A := \begin{bmatrix} -.9493 & 1 \\ 0 & 1 \\ 1.64866 & 1 \\ 2.95387 & 1 \\ 3.67706 & 1 \end{bmatrix}:$$

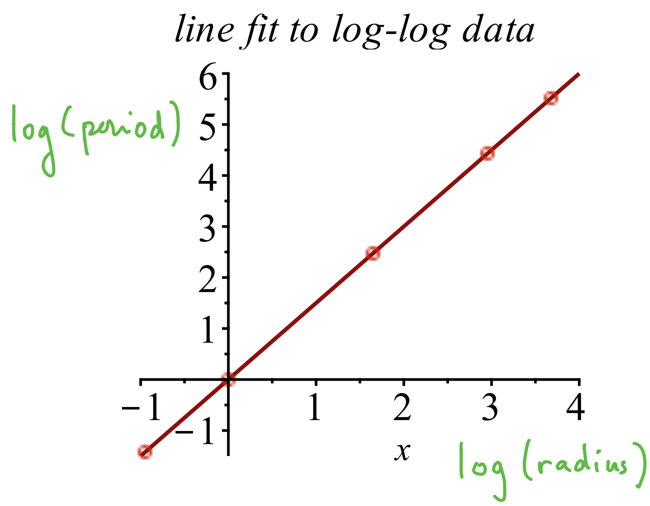
$$\left(Transpose(A).A\right)^{-1}.Transpose(A).\begin{bmatrix} -1.42296 \\ 0 \\ 2.47317 \\ 4.43082 \\ 5.51544 \end{bmatrix};$$

1.5

$$\begin{array}{cc} \begin{bmatrix} 1.49982355212829 \\ 0.000465682813906573 \end{bmatrix} & \begin{array}{l} m \\ b \end{array} \end{array} \qquad (1)$$

So we get essentially the correct power.

```
> with(plots) :
  plot1 := pointplot({[-.9493, -1.42296], [0, 0], [1.64866, 2.47317], [2.95387, 4.43082],
      [3.67706, 5.51544]}, color = red, symbol = circle, symbolsize = 18) :
  plot2 := plot(1.4998·x + .0005, x = -1 ..4) :
  display({plot1, plot2}, title = `line fit to log-log data`);
```




```
> plot3 := pointplot({[.387, .241], [1., 1.], [5.20, 11.86], [19.18, 84.0], [39.53, 248.5]}, color
      = red, symbol = circle, symbolsize = 18) :
  plot4 := plot(exp(0.00046568)·R^1.49982, R = 0 ..50) :
  display({plot3, plot4}, title = `Kepler's Laws`);
```