ME170b Lecture 7 3/8/24

Experimental Techniques

Last time:

> Rejection of data
> Weighted Averages
> Least Squares

Today:

> Ch.9
> > Finish least squares
> > correlation and covariance

What is the purpose?

$$y = \underbrace{A}_{\text{intercept}} + \underbrace{B}_{\text{slope}}x \leftarrow \text{fit this to data}$$

1. We want to estimate the coefficients A and B

2. Another important determination is whether the data (x_i, y_i) rally are linear — "how <u>well</u> does the data fit our model?" (Ch.9)

How to estimate A and B?

$$(x_1, y_1), \ldots, (x_N, y_N)$$

assume y suffer appreciable uncertainty, the uncertainty in our measurements of x is negligible.

let's use ML. first proceed as if we know A and B:

$$(\text{true value for } y_i) = A + Bx_i$$

$$Prob_{A,B}(y_i) \propto \frac{1}{\sigma_y} e^{-(y_i - A - Bx_i)^2/2\sigma_y{}^2}$$

$$Prob_{A,B}(y_1, \ldots, y_N) = Prob_{A,B}(y_1) \cdots Prob_{A,B}(y_N) \propto \frac{1}{\sigma_y{}^N} e^{-\chi^2/2},$$

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - A - Bx_i)^2}{\sigma_y{}^2}$$

Best estimates of A and B maximize the probability, which corresponds to minimizing the CHI^2 term (hence least squares)

minimize χ^2

maximize $Prob(y_1 \cdots y_n)$

How to estimate A and B?

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - A - Bx_i)^2}{\sigma_y{}^2}$$

How to find and expression for the minimum?

$$\frac{\partial \chi^2}{\partial A} = 0 \rightarrow \frac{-2}{\sigma_y^2} \sum_{i=1}^{N} (y_i - A - Bx_i) = 0$$

$$\frac{\partial \chi^2}{\partial B} = 0 \rightarrow \frac{-2}{\sigma_y^2} \sum_{i=1}^{N} x_i (y_i - A - Bx_i)$$

Solve for B & A

How to estimate A and B?

$$A = \frac{\sum x^2 \sum y - \sum x \sum xy}{\Delta}$$

$$\Delta = N \sum x^2 - \left(\sum x\right)^2$$

$$B = \frac{N \sum xy - \sum x \sum y}{\Delta}$$

How to estimate uncertainty in y?

Remember that the numbers y_1, y_2, … y_N are not N measurements of the same quantity. (They might, for instance, be the times for a stone to fall from N different heights.)

The measurement of each y, is (we are assuming) normally distributed about its true value A + Bx,, with width parameter sigma.

$$\sigma_y = \sqrt{\frac{1}{N}\sum(y_i - A - Bx_i)^2}$$

How to estimate uncertainty in A and B?

The uncertainties in A and B are given by simple error propagation in terms of those in y_1 … y_N

$$\sigma_A = \sigma_y \sqrt{\frac{\sum x^2}{\Delta}}$$

$$\sigma_B = \sigma_y \sqrt{\frac{N}{\Delta}}$$

Some caveats

1. What if the uncertainty of y is <u>not</u> equal for all measurements?

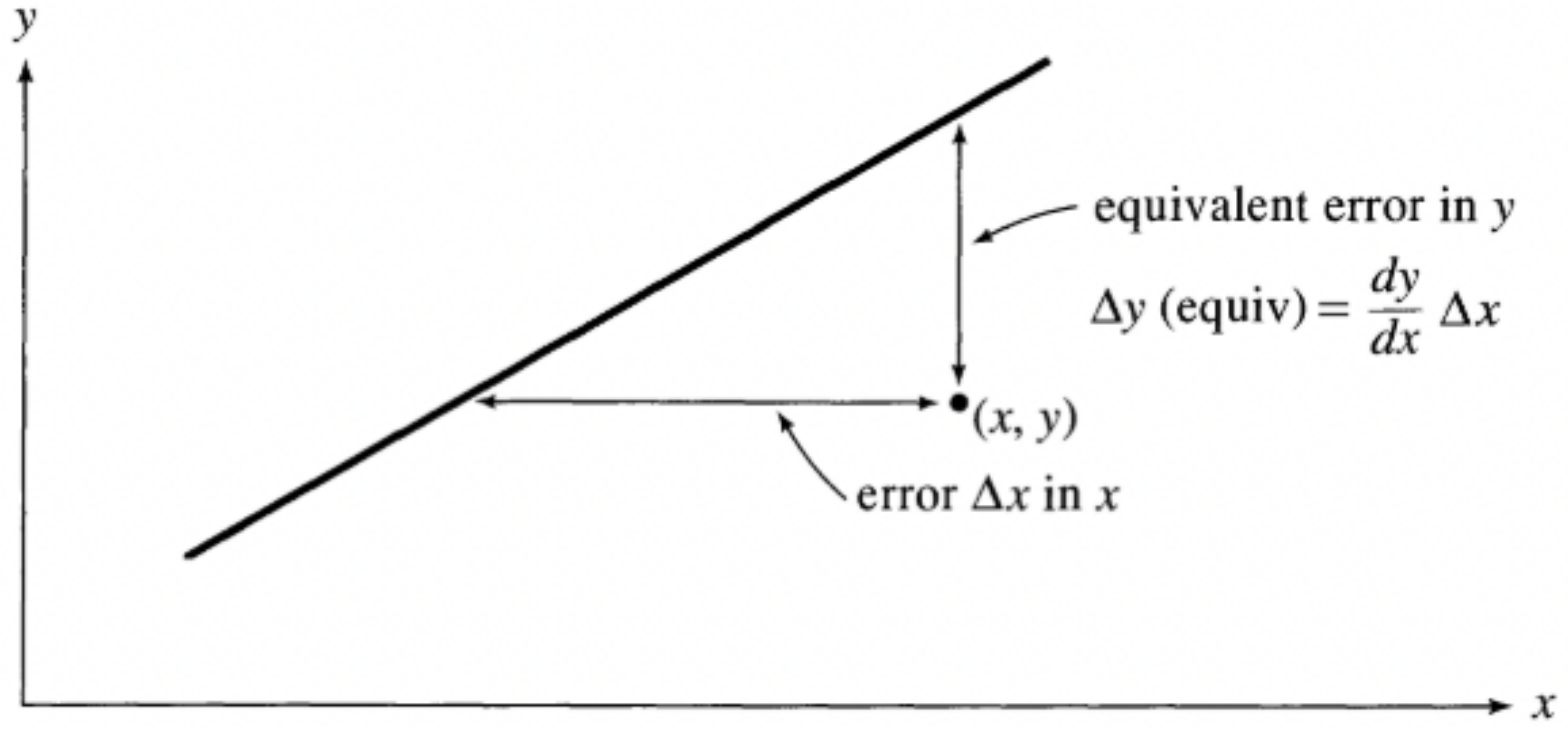
we can use the method of weighted least squares,
(ex. in Prob. 8.9)

2. What if both x and y have uncertainties

actually doesn't make a bog difference

What if both x and y have uncertainties

Assume error in x only



$$\sigma_y(\text{equiv}) = \frac{dy}{dx}\sigma_x. \qquad \sigma_y(\text{equiv}) = B\sigma_x$$

if all the uncertainties sigma_x, are equal, the same is true of the equivalent uncertainties simga_y(equiv).

What if both x and y have uncertainties

Now for the case that both x and y have uncertainties.

$$\sigma_y(\text{equiv}) = \sqrt{\sigma_y^2 + (B\sigma_x)^2} \quad \checkmark$$

If both x and y have uncertainties, we can combine in quadrature and replace with a single uncertainty

The most complicated case is when each measurement x_i and y_i have their own uncertainties, then we need to use the equivalence and a weighted least squares

We can use least squares to fit nonlinear curves!

$$y = A + Bx + Cx^2$$ polynomial

$$\mathrm{Prob}_{A,B,C}(y_1 \cdots y_N) \propto e^{-\chi^2/2}$$

$$\chi^2 = \sum_{i=0}^{N} \left(\frac{y_i - A - Bx_i - Cx_i^2}{\sigma_y^2} \right)^2$$

General case when least squares can fit

problems in which the function y = f(x) depends linearly on the parameters A, B, C, …

$$y = A\sin x + B\cos x.$$

$$y = Ae^{Bx} \qquad \ln y = \ln A + Bx$$

Another look at least squares in matrix form

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

model: beta are our parameters
y and x measurements

$$y_i = \beta_1 x_i + \beta_0$$

$$\mathbf{Y} = \mathbf{X}\beta$$

$$\mathbf{Y} = \begin{bmatrix} y_i \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_i & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_0 \end{bmatrix}$$

$y = \beta_1 x + \beta_0$ $\quad$ β_1 : slope

β_0 : intercept

model

ML

$y_i = \beta_1 x_i + \beta_0$ $\quad$ for all data $\{y_i, x_i\}$

rewrite as a matrix?

$Y = X\beta$ $\quad$ $Y : \begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_N \end{bmatrix}$ $\quad$ $X : \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix}$ $\quad$ $\begin{bmatrix} \beta_1 \\ \beta_0 \end{bmatrix}$

$N \times 1$ $\quad$ $N \times 2$ $\quad$ 2×1

$$Y = X\beta$$

$$\min_\beta \underbrace{\|Y - X\beta\|^2}_{J}$$ (optimization problem)

$$J = (Y - X\beta)^T (Y - X\beta)$$

$$= Y^T Y - \underbrace{(X\beta)^T Y - Y^T (X\beta)}_{\text{combine}} + (X\beta)^T (X\beta)$$

$$J = Y^T Y - 2\beta^T X^T Y - \beta^T X^T X \beta$$

$$\frac{\partial J}{\partial \beta} = 0$$

$$= -2X^T Y + 2X^T X \beta$$

$$\boxed{\beta^* = \left[(X^T X)^{-1} X^T \right] Y}$$

psuedo-inverse

pinv($\cdot$)

$$Y = X\beta \quad \text{if } N = 2$$

$N \times 1$ $\quad$ $N \times 2$ $\quad$ 2×1

Another look at least squares in matrix form

$$\min_{\beta} ||\mathbf{Y} - \mathbf{X}\beta||^2 = \min_{\beta} \mathcal{J}$$

$$\begin{aligned}
\mathcal{J} =& (\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta) \\
=& \mathbf{Y}^T\mathbf{Y} - (\mathbf{X}\beta)^T\mathbf{Y} - \mathbf{Y}^T(\mathbf{X}\beta) + (\mathbf{X}\beta)^T(\mathbf{X}\beta) \\
=& \mathbf{Y}^T\mathbf{Y} - 2\beta^T\mathbf{X}^T\mathbf{Y} - \beta^T\mathbf{X}^T\mathbf{X}\beta
\end{aligned}$$

Another look at least squares in matrix form

$$\frac{\partial \mathcal{J}}{\partial \beta} = 0$$

$$= -2\mathbf{X}^T\mathbf{Y} + 2\mathbf{X}^T\mathbf{X}\beta$$

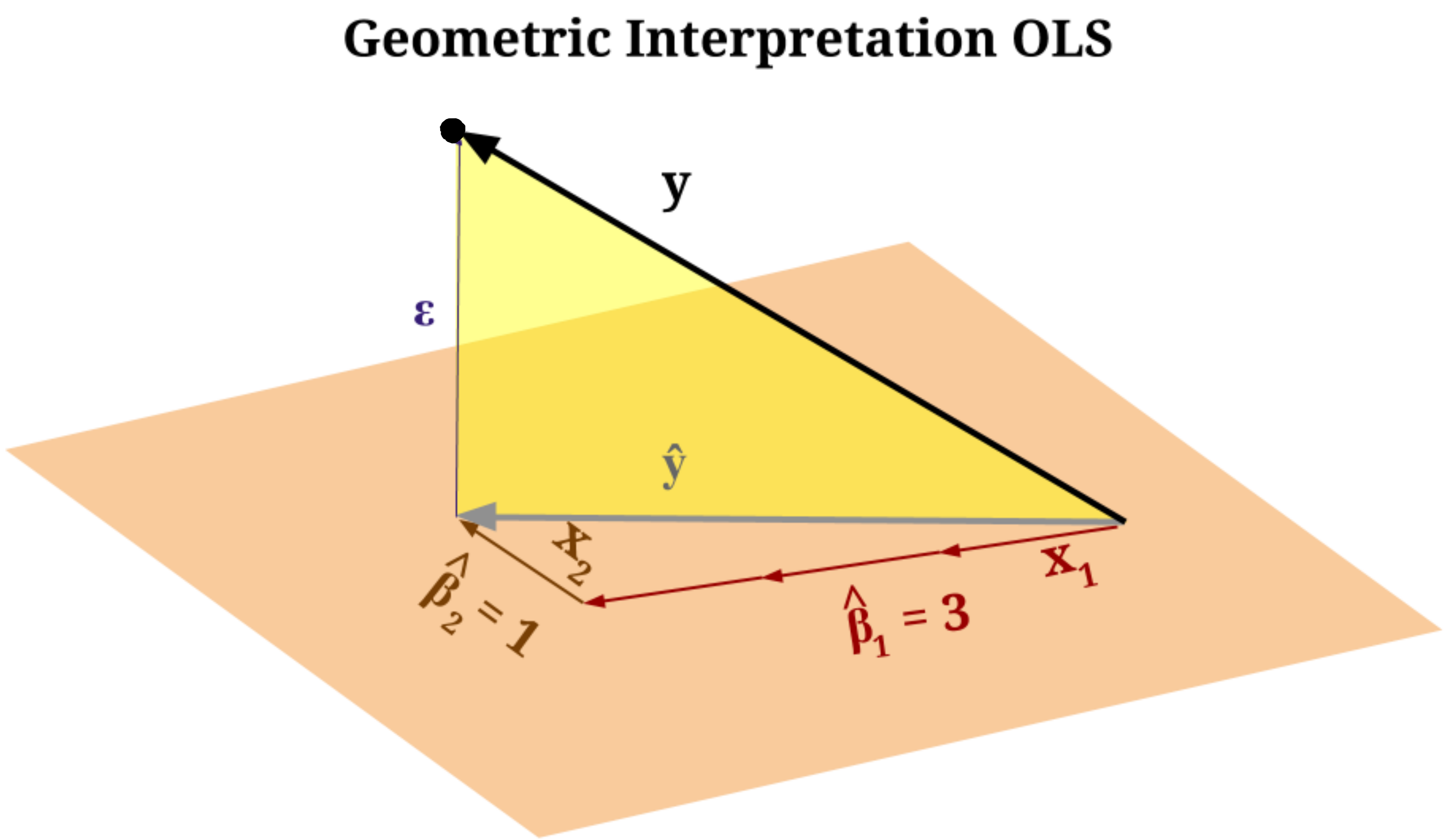
$$\beta^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

Another look at least squares in matrix form

$$\beta^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

Solution is the ‘projection’ of the b on the space that matrix A spans

X



$$y = \beta_1 x + \beta_2 x^2 + \beta_0$$

$$Y = \begin{bmatrix} x_1 & x_1^2 & 1 \\ \vdots & \vdots & \vdots \\ x_N & x_N & \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_0 \end{bmatrix}$$

$$y = \beta_1 \sin(x) + \beta_2 \log(x)$$

$$Y = \begin{bmatrix} \sin(x_1) & \log(x_1) \\ \vdots & \vdots \\ \sin(x_n) & \log(x_N) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Another look at least squares in matrix form

$$y = \beta_1 cos(x) + \beta_2 sin(\cancel{y}) + \beta_3 x^2$$

Write the matrix X, Y, beta

$$X = \begin{bmatrix} cos(x_1) & sin(x_1) & x_1^2 \\ \vdots & \vdots & \vdots \\ cos(x_n) & sin(x_n) & x_N^2 \end{bmatrix}$$

$\min \|y - X\beta\| + \lambda|\beta|$

$$X = \begin{bmatrix} x^2 & sin(x) & \dot{x}\,x^2 & \dot{x} & \cdots \\ & & & & \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}$$

Covariance and Correlation

First let's review the principles of error propagation

If we measuring two quantities x and y to calculate some function q(x, y):

$$\delta q \approx \left|\frac{\partial q}{\partial x}\right| \delta x + \left|\frac{\partial q}{\partial y}\right| \delta y$$

"naive".

they could cancel errors in x y.

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x}\delta x\right)^2 + \left(\frac{\partial q}{\partial y}\delta y\right)^2}$$

Covariance and Correlation

First let's review the principles of error propagation

If we measuring two quantities x and y to calculate some function q(x, y):

$$\delta q \approx \left|\frac{\partial q}{\partial x}\right| \delta x + \left|\frac{\partial q}{\partial y}\right| \delta y.$$

our naive uncertainty

there may be partial cancellation of the errors in x and y.

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x}\,\delta x\right)^2 + \left(\frac{\partial q}{\partial y}\,\delta y\right)^2}.$$

we can prove this assuming Gaussians

Covariance and Correlation

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x}\delta x\right)^2 + \left(\frac{\partial q}{\partial y}\delta y\right)^2}.$$

if the measurements of x and y are governed by independent normal distributions, with standard deviations sigma_x and sigma_y the values of q(x, y) are also normally distributed, with standard deviation

$$\sigma_q = \sqrt{\left(\frac{\partial q}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial q}{\partial y}\sigma_y\right)^2}.$$

✓ as bound even if non-normal

This result provides the justification for the claim

But what if we don't meet the assumptions?

does it still apply whether or not the errors in x and y are independent and normally distributed.

Claim: the estimate <u>always</u> is upper bound estimate of uncertainty!

Recall STD

$$\sigma_x^{\ 2} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

If the measurements of x are normally distributed, then in the limit that N is large,

$$\frac{1}{\sigma_x\sqrt{2\pi}}e^{-(x-X)^2/2\sigma_x{}^2}$$

simga_x <u>is</u> the width parameter

If the underlying process is non-Gaussian — simga_x is <u>still</u> the STD, but this relationship is no longer available to us.

Covariance Propagation

$f =$

Suppose that to find a value for the function q(x, y), we measure the two quantities x and y several times, obtaining N pairs of data, (x1, y1)…(xN, yN).

We can still calculate:

> mean x and sigma_x

> mean y and sigma_y

> mean q and sigma_q

Covariance Propagation

Jacobian

$$q_i = q(x_i, y_i)$$

$$\approx q(\bar{x}, \bar{y}) + \frac{\partial q}{\partial x}(x_i - \bar{x}) + \frac{\partial q}{\partial y}(y_i - \bar{y})$$

$$\bar{q} = \frac{1}{N} \sum q_i$$

$$= \frac{1}{N} \sum \left[\underbrace{q(\bar{x}, \bar{y})}_{\substack{\text{from} \\ \bar{x} \\ \bar{y}}} + \frac{\partial q}{\partial x}(x_i - \bar{x}) + \frac{\partial q}{\partial y}(y_i - \bar{y}) \right]$$

$$\bar{q}$$

Covariance Propagation

$$\sigma_q^{\,2} = \frac{1}{N}\Sigma(q_i - \bar{q})^2$$

$$= \left(\frac{\partial q}{\partial x}\right)^2 \frac{1}{N}\sum(x_i - \bar{x})^2 + \left(\frac{\partial q}{\partial y}\right)^2 \frac{1}{N}\sum(y_i - \bar{y})^2$$

$$+ 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\boxed{\underbrace{\frac{1}{N}\sum(x_i - \bar{x})(y_i - \bar{y})}_{\sigma_{xy}}}$$

$$\sigma_q^{\,2} = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^{\,2} + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^{\,2} + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\sigma_{xy}$$

Covariance Propagation

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}).$$

$${\sigma_q}^2 = \left(\frac{\partial q}{\partial x}\right)^2 {\sigma_x}^2 + \left(\frac{\partial q}{\partial y}\right)^2 {\sigma_y}^2 + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\sigma_{xy}$$

This equation gives the standard deviation sigma_q, whether or not the measurements of x and y are independent or normally distributed.

Covariance Propagation

- If the measurements of x and y are not independent, the covariance sigma_xy is non zero.
- if measurements are independent the covariance is zero
- When the covariance is not zero (even in the limit of infinitely many measurements, we say that the errors in x and y are correlated.

Example: Two Angles with a Negative Covariance

Each of five students measures the same two angles α and β and obtains the results shown in the first three columns of Table 9.1.

Table 9.1. Five measurements of two angles α and β (in degrees).

Student	α	β	$(\alpha - \overline{\alpha})$	$(\beta - \overline{\beta})$	$(\alpha - \overline{\alpha})(\beta - \overline{\beta})$
A	35	50	2	−2	−4
B	31	55	−2	3	−6
C	33	51	0	−1	0
D	32	53	−1	1	−1
E	34	51	1	−1	−1

$$\sigma_{\alpha\beta} = \frac{1}{N} \sum (\alpha - \overline{\alpha})(\beta - \overline{\beta}) = \frac{1}{5} \cdot (-12)$$

$$= \underline{\underline{-2.4}}$$

Upper limit on sigma_q

Schwarz inequality

$$|\sigma_{xy}| \leqslant \sigma_x \sigma_y$$

$$\sigma_q{}^2 \leqslant \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x{}^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y{}^2 + 2\left|\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\right| \sigma_x \sigma_y$$

$$= \left[\left|\frac{\partial q}{\partial x}\right| \sigma_x + \left|\frac{\partial q}{\partial y}\right| \sigma_y\right]^2;$$

$$\sigma_q \leqslant \left|\frac{\partial q}{\partial x}\right| \sigma_x + \left|\frac{\partial q}{\partial y}\right| \sigma_y$$

Main Results on Covariance

$$\delta q \approx \left|\frac{\partial q}{\partial x}\right| \delta x + \left|\frac{\partial q}{\partial y}\right| \delta y$$

naive estimate is
still upper bound!

$$\sigma_q \leqslant \left|\frac{\partial q}{\partial x}\right| \sigma_x + \left|\frac{\partial q}{\partial y}\right| \sigma_y$$