LECTURE 8

Q. Suppose I want to model the relationship between

- independent variable t
- dependent variable P

using a cubic model (ie $P(t) = at^3 + bt^2 + ct + d$)

Suppose I run the experiment 30 times (ie 30 data points.)

In the matrix equation ($Ax = b$) describing the overdetermined system,

- what are the dimensions of A?
- What is x explicitly?

* residual vector: $b - Ax$; we want to minimize $\|b - Ax\|_2$.

Solving the least squares problem with normal equations

residual vector $b - Ax$

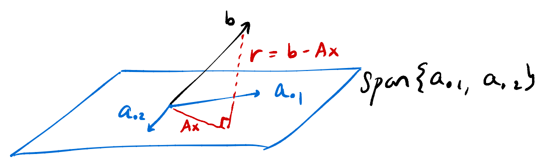
$A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$ (variable), $b \in \mathbb{R}^m$ (data)

We want to find a vector $x \in \mathbb{R}^n$ that solves the minimization problem $\min_x \|b - Ax\|_2$ i.e. find the min value, and where it occurs

e.g. (running example with fixed dimensions)

$$Ax = \begin{array}{c} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \\ a_{\cdot 1} \quad a_{\cdot 2} \end{array} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Intuition:



- $Ax \in$ colspace of A

- r minimized if we make it orthogonal to colspace(A)

  $r^T a_{\cdot j} = 0 \quad \forall j$

  i.e. in general: $r^T (a_{\cdot 1} \; a_{\cdot 2} \; \dots \; a_{\cdot n}) = r^T A = 0$

$$r^T A = [r_1 \; r_2 \; r_3] \begin{bmatrix} a_{11} & & \\ \vdots & & \\ & & a_{32} \end{bmatrix} = 0 \quad \longleftrightarrow \quad A^T r = \begin{bmatrix} a_{11} & \dots & \\ & \ddots & a_{32} \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = 0$$

- Normal Equations (Normal = orthogonal, but for ↗ & planes)

  $r = b - Ax$

  $r^T = (b - Ax)^T = b^T - (Ax)^T = b^T - x^T A^T$

  $\Rightarrow r^T A = (b^T - x^T A^T) A = b^T A - x^T A^T A$ (ew)

Transpose the whole (# equation) $x^T A^T A = b^T A$

$\rightsquigarrow$ $A^T A x = A^T b$ The normal equations.

HW: Read §3.6, specifically Example 3.11

In our small running example:

$$Ax = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$A^T A x = A^T b$ * x is a variable!

$$\begin{bmatrix} & A^T & \end{bmatrix} \begin{bmatrix} A \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} & A^T & \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \cdot 4 \end{bmatrix} \rightsquigarrow \begin{array}{l} x_1 = 3 \\ 4x_2 = 8 \end{array} \rightsquigarrow \text{unique solution} \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

This works out quite often (Unique sol!):

<u>thm 3.10</u> If the col vects of A are linearly indep, then the normal eqns $A^T A x = A^T b$ are nonsingular $\Rightarrow$ have a unique solution!

Q. What does it mean, in practice, that the col vects of A are linearly independent?

$$\begin{bmatrix} & a_{21} \\ A & \\ & \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} \\ b \\ \\ \end{bmatrix}$$

a_{21} = coeff of param x_2 in experiment 1.

$a_{\cdot j}$ = coeffs of param x_j through all the experiments.

* If you really expect the coeffs of x_2 to be multiples of the coeffs of x_1, then you're using the wrong model...

$e + KF = l$ $\qquad x = \begin{bmatrix} e \\ K \end{bmatrix}$

if $K \approx 17 \cdot e$ then $e + (17e)F = l$
$= e(1 + 17F)$

only one parameter; use a simpler model

$\Longrightarrow$ So this always should work if you did your modeling HW before setting up the experiment...

Pf. of thm 3.10 (Sketch)

Claim 1 A^TA is positive definite.

let $x \neq 0$. cols of A are lin indp $\Longrightarrow$ $Ax \neq 0$ (x_i are coeffs)

So let $y = Ax \neq 0$.

$\Longrightarrow$ $x^TA^TAx = y^Ty = \sum y_i^2 > 0$ $\Longrightarrow$ A^TA is pos. def. (Review if needed!)

$\Longrightarrow$ A^TA (square matrix) is nonsingular, $\Rightarrow$ get a unique solution to the normal equations, called $\hat{x}$.

Note $A^T(A\hat{x}) = A^T(b)$ $\Longrightarrow$ $A^T(\underbrace{b - A\hat{x}}_{\hat{r}}) = A^T\hat{r} = 0$

$\hat{r}$ optimal residual vect.

Claim 2 $\hat{x}$ actually is the solution to the least squares problem (minimized residual vector $\|\cdot\|_2$ length)

WTS $\|\hat{r}\|_2 < \|r\|_2$ $\forall$ $r = b - Ax$. $\quad$ • $\hat{r} = b - A\hat{x}$

• $r = b - Ax = b - Ax + A\hat{x} - A\hat{x} = b - A\hat{x} + A(\hat{x} - x) = \hat{r} + A(\hat{x} - x)$

• $\|r\|_2^2 = r^Tr = [\hat{r} + A(\hat{x} - x)]^T[\hat{r} + A(\hat{x} - x)]$

expand: $= [\hat{r}^T + (\hat{x} - x)^TA^T][\hat{r} + A(\hat{x} - x)]$

$= \hat{r}^T\hat{r} + \underbrace{\hat{r}^TA}_{0}(\hat{x} - x) + (\hat{x} - x)^T\underbrace{A^T\hat{r}}_{0} + \underbrace{(\hat{x} - x)^TA^T}_{y^T}\underbrace{A(\hat{x} - x)}_{y}$

$\geq \|\hat{r}\|_2^2$

//

HW03: fill in the gaps; explain