Note Title 2024-09-27

Linear least-squares problem

Given
$v_1, v_2, \dots v_n \in \mathbb{R}^m$ $y \in \mathbb{R}^m$

we look for coefficients $x_1, x_2, \dots x_n$ s.t.

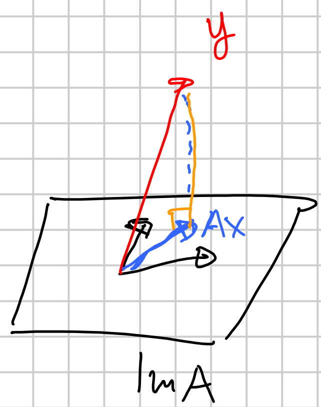
$$v_1 x_1 + v_2 x_2 + \dots + v_n x_n = y \iff Ax = y$$

$m \times n$ $n \times 1$ $m \times 1$

Food 1:	Food2:
10 fats	20 fats
20 proteins	10 prot
30 sugars	0 sugars

A mixture has
50 fats
40 protein
60 sugars

$$\begin{bmatrix} 4 \\ 3 \\ 0 \end{bmatrix} x_1 + \begin{bmatrix} 5 \\ 2 \\ 0 \end{bmatrix} x_2 = \begin{bmatrix} 6 \\ 6 \\ 2 \end{bmatrix}$$



$$\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} x_1 + \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} x_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

sum of entries $= 0$ $\neq$ sum of entries $\neq 0$

When the problem is not solvable, it makes sense to ask:
what is the closest I can get to y?

$$\min_{x \in \mathbb{R}^n} \|Ax-y\|_2 = \min_{x \in \mathbb{R}^n} \sqrt{\sum_{i=1}^{m} (Ax-y)_i^2}$$

Division operators in Matlab:

$5/2 \to \frac{5}{2}$

$5 \backslash 2 \to \frac{2}{5}$

$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ $b = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$ $Ax = b \iff x = A^{-1}b$

⚠ NOT bA^{-1}

$A \backslash b$ $\to$ sugar for $A^{-1}b$

$X / Y \to$ sugar for $X \cdot Y^{-1}$

$\underset{n\times 1}{b} / \underset{n\times n}{A}$ doesn't make sense b^T / A

Also to solve least squares problems: $A \backslash y$

Applications: Machine learning:

$\min \|Xw - y\|$ X, y data w weights (unknown)

NBA salary estimation:

salary $\approx$ rebounds $\cdot w_1$ + fouls $\cdot w_2$ + points $\cdot w_3$

Best estimate:

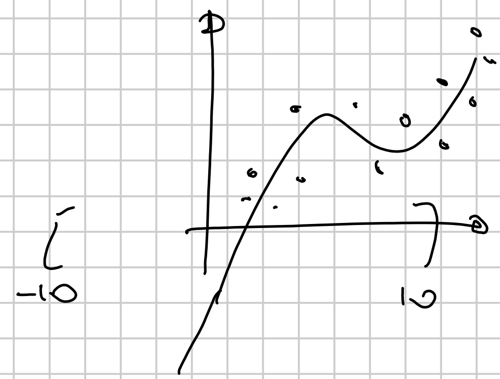
$$\min_{w \in \mathbb{R}^3} \sum_{i \in \text{Players}} (\text{salary}_i - \text{rebounds}_i w_1 - \text{fouls}_i w_2 - \text{points}_i w_3)^2$$

$$X = \begin{bmatrix} reb_1 & fouls_1 & pts_1 \\ | & | & | \\ reb_m & fouls_m & pts_m \end{bmatrix} \qquad y = \begin{bmatrix} salary_1 \\ | \\ salary_m \end{bmatrix}$$

$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \qquad \min_{w \in \mathbb{R}^3} \| Xw - y \|^2$$

<u>Application</u>: polynomial fitting

$$y \approx ax^3 + bx^2 + cx + d$$



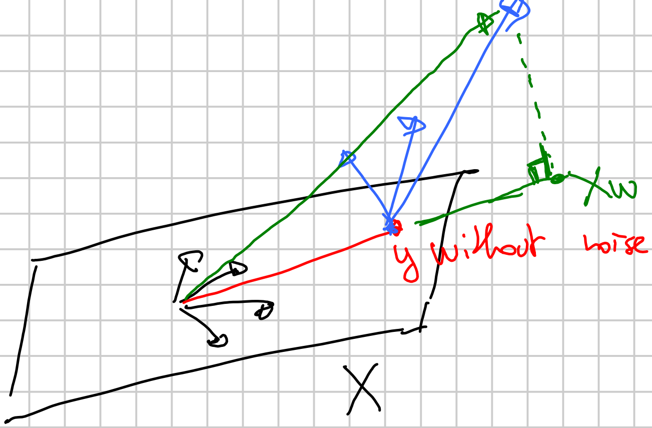
Given $y_1, \dots y_{1000}$, $x_1, \dots x_{1000}$, find

$$\min \sum_{i=1}^{1000} \left(y_i - (ax_i^3 + bx_i^2 + cx_i + d) \right)^2 =$$

Unknowns: $\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$

$$= \min_{\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \in \mathbb{R}^4} \Bigg\| \underbrace{\begin{bmatrix} x_1^3 & x_1^2 & x_1 & 1 \\ x_2^3 & x_2^2 & x_2 & 1 \\ \vdots & & & \\ x_m^3 & x_m^2 & x_m & 1 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}}_{w} - \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_{y} \Bigg\|^2$$

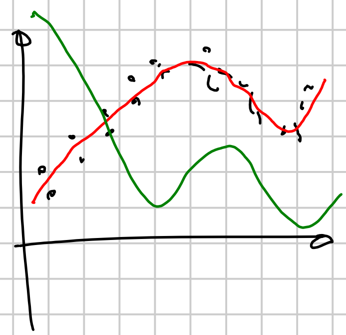
bias column



Maximum likelihood estimator:

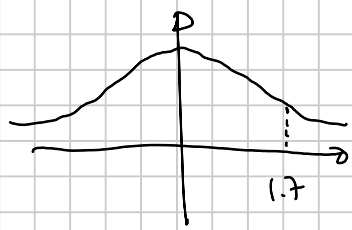
error ~ Gaussian

If one assumes $y_i = ax_i^3 + bx_i^2 + cx_i + d + \boxed{e_i}$



max. likelihood with Gaussian errors:

$$\Leftrightarrow \quad \min \|Xw - y\|^2$$



Solvability of Least squares problems

$$\left\| \begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} x - \begin{bmatrix} 4 \\ 4 \\ 1 \\ 2 \end{bmatrix} \right\|^2$$

In some cases, more than one solution!

We cannot reach $\begin{bmatrix} 4 \\ 4 \\ 1 \\ 2 \end{bmatrix}$, but we can reach $Ax = \begin{bmatrix} 4 \\ 4 \\ 0 \\ 0 \end{bmatrix}$

$x = \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix}$ or $x = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix}$ $\quad x = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ ....

This happens because there is a vector $z \neq 0$ such that $Az = 0$ $\quad z \in \ker A$

$$z = \begin{bmatrix} 2 \\ 4 \\ -4 \end{bmatrix} \text{ is such that } Az = 0$$

Non uniqueness: if x is a solution, so is $x + z$ $\quad x + 2z$

$x - 37.4z$

$$A(x + z\alpha) = Ax + \cancel{Az\alpha}$$

Def: we say that A has full column rank

if $\operatorname{rk} A =$ number of columns or equivalently its columns are linear independent or equivalently $\ker A = \{z : Az = 0\} = \{0\}$.

Result: A has full column rank if and only if $A^T A$ is positive definite.

Proof: A has full column rank $\Leftrightarrow Az \neq 0$ for all $z \neq 0$

$\Leftrightarrow \|Az\| \neq 0$ for all $z \neq 0$ $\Leftrightarrow \|Az\|^2 \neq 0$ for all $z \neq 0$

$\Leftrightarrow (Az)^T (Az) = z^T A^T A z \neq 0$ for all $z \neq 0 \Leftrightarrow A^T A$ is pos. def.

Note that

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - y\|^2 = \min \frac{1}{2} (Ax - y)^T (Ax - y)$$

$$= \min \frac{1}{2} (Ax)^T Ax - \underbrace{\frac{1}{2} (Ax)^T y}_{\text{equal}} - \underbrace{\frac{1}{2} y^T (Ax)} + \frac{1}{2} y^T y$$

$$= \min_{x \in \mathbb{R}^n} \frac{1}{2} x^T \underbrace{A^T A}_{Q} x - \underbrace{\cancel{2}\, y^T A}_{q^T} x + \underbrace{\frac{1}{2} y^T y}_{\text{constant}}$$

gradient: $A^T A x - A^T y$

Hessian: $A^T A$

For a quadratic function, the minimum is unique <u>if and only if</u> $Q = A^T A \succ 0$ is pos. definite

$\iff$ A has full column rank

minimum: $x = -Q^{-1} q = (A^T A)^{-1} (A^T y)$

closed-form expression for the solution!

Solution algorithm:

<u>Input</u>: $A \in \mathbb{R}^{m \times n}$ full col. rank, $y \in \mathbb{R}^m$

<u>Output</u>: $x \in \mathbb{R}^n$ that solves $x = \arg\min \| Ax - y \|$

1. Compute $\boxed{A^T A}$ $n \times m$ $m \times n \to n \times n$ $\sim \cancel{2} mn^2$
2. Compute $A^T y$ $n \times m$ $m \times 1 \to n \times 1$ $\sim 2mn$
3. Solve the square lin. system $(A^T A) x = A^T y$ $\sim \frac{\cancel{2}}{3} n^3$

$x = (A^T A)^{-1} (A^T y)$

<u>Linear</u> in the larger dimension, <u>quadratic</u> in the smaller

<u>Trick</u>: $A^T A$ is symmetric: $2mn^2$ becomes mn^2

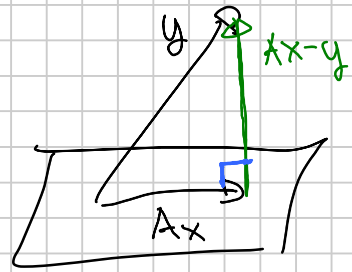
<u>Trick</u>: the linear system has a symmetric pos. definite matrix

We can use a better algorithm than Gaussian elimination / LU factorization: <u>Cholesky</u> factorization

$\frac{2}{3}n^3$ becomes $\frac{1}{3}n^3$.

This algorithm is known as <u>method of normal equations</u>

$$A^TAx = A^Ty \iff A^T(Ax-y) = 0$$



The residual $Ax-y$ is perpendicular to the plane spanned by the columns of A.

<u>Consequence</u>: the solution of $\min \|Ax-y\|$ is given by $x = \boxed{(A^TA)^{-1}A^T}\,y$

Def: $(A^TA)^{-1}A^T =: A^+$ is called the <u>pseudoinverse</u> of $A \in \mathbb{R}^{m\times n}$, $m > n$ with full column rank.

Note that $A^+A = (A^TA)^{-1}(A^TA) = I$

but $AA^+ \neq I$

$\underset{m\times n}{\square}\cdot\underset{n\times m}{\square}$ is an $m\times m$ matrix with rank $n<m$ and so cannot be the identity.

If A square $(A^TA)^{-1} = A^{-1}(A^T)^{-1}$

$$\Rightarrow A^+ = (A^TA)^{-1}A^T = A^{-1}\cancel{(A^T)^{-1}A^T} = A^{-1}.$$