Projections

Projection Matrix: A matrix $\underline{A}$ such that any given vector $\underline{b}$ can be projected onto $\underline{a}$ via

$$\underline{p} = \underline{A}\,\underline{b}$$

$$\underline{p} = \underbrace{\frac{\underline{a}^T\underline{b}}{\underline{a}^T\underline{a}}}_{\text{scalar}}\,\underline{a} = \underline{a}\,\frac{\overbrace{\underline{a}^T\underline{b}}^{\text{Inner product}}}{\underbrace{\underline{a}^T\underline{a}}_{\text{Inner product}}} = \underbrace{\frac{\overbrace{\underline{a}\,\underline{a}^T}^{\text{Outer product}}}{\underline{a}^T\underline{a}}}_{\underline{A}\ \text{(matrix)}}\,\underline{b}$$

Thus, the projection matrix for a vector $\underline{a}$ is given by

$$\underline{A} = \frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}}$$

$\underline{a}^T\underline{a}$: inner product (scalar)

$\underline{a}\,\underline{a}^T$: outer product (matrix)

$\underline{a} \otimes \underline{a}$

Example: In 2D, $\underline{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$

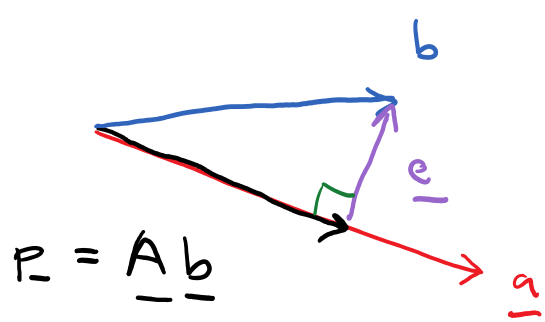
$$\underline{a}\,\underline{a}^T = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\begin{bmatrix} a_1 & a_2 \end{bmatrix} = \begin{bmatrix} a_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 \end{bmatrix}$$

$$\underline{a}^T \underline{a} = a_1^2 + a_2^2 \qquad \underline{A} = \frac{1}{a_1^2 + a_2^2}\begin{bmatrix} a_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 \end{bmatrix}$$

Note: $\underline{A}$ will project any vector $\underline{b}$ onto $\underline{a}$

Also, $\underline{A}$ is idempotent (ai·duhm·pow·tnt): repeated application of $\underline{A}$ has no effect

Example: $\underline{A}(\underline{A}\,\underline{b}) = \underline{A}\,\underline{b}$



$$\underline{A}^2 = \underline{A}\,\underline{A} = \left(\frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}}\right)\left(\frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}}\right) = \frac{\underline{a}\,\cancel{(\underline{a}^T\underline{a})}\,\underline{a}^T}{\cancel{(\underline{a}^T\underline{a})}(\underline{a}^T\underline{a})}$$

$$= \frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}} = \underline{A}$$

Furthermore, $\underline{I} - \underline{A}$ can now project onto the perpendicular space of $\underline{a}$

$$(\underline{I}-\underline{A})\underline{b} = \underline{b}-\underline{A}\,\underline{b} = \underline{b}-\underline{p} = \underline{e}$$

Also, $\underline{I}-\underline{A}$ is idempotent

$$\begin{aligned}(\underline{I}-\underline{A})^2 &= (\underline{I}-\underline{A})(\underline{I}-\underline{A}) \\ &= \underline{I}^2 - \underline{I}\underline{A} - \underline{A}\underline{I} + \underline{A}^2 \\ &= \underline{I} - \underline{A} - \underline{A} + \underline{A} \\ &= \underline{I}-\underline{A}\end{aligned}$$

Example: Find the projection matrix of $\underline{a} = \begin{bmatrix}1\\2\end{bmatrix}$ and then project $\underline{b} = \begin{bmatrix}1\\1\end{bmatrix}$ onto $\underline{a}$

$$\underline{a} = \begin{bmatrix}1\\2\end{bmatrix}, \quad \underline{A} = \frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}}$$

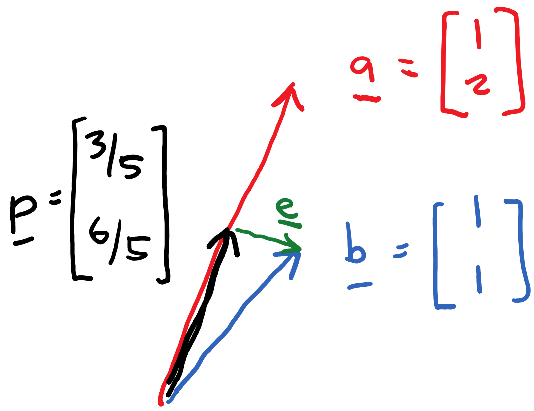
outer product
inner product

$$\underline{a}^T\underline{a} = \begin{bmatrix}1 & 2\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix} = 1^2 + 2^2 = 5$$

$$\underline{a}\,\underline{a}^T = \begin{bmatrix}1\\2\end{bmatrix}\begin{bmatrix}1 & 2\end{bmatrix} = \begin{bmatrix}1 & 2\\2 & 4\end{bmatrix}$$

$$\underline{A} = \frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}} = \begin{bmatrix}1/5 & 2/5\\2/5 & 4/5\end{bmatrix}$$

$$\underline{p} = \underline{A}\,\underline{b} = \begin{bmatrix}1/5 & 2/5\\2/5 & 4/5\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}3/5\\6/5\end{bmatrix}$$



$$\underline{A}^2\underline{b} = \underline{A}(\underline{A}\,\underline{b})$$
$$= \underline{A}\,\underline{p} = \underline{p}$$

Projects $\underline{p}$ onto $\underline{a}$ or onto itself!

$\therefore$ $\underline{A}$ is idempotent

$$(\underline{I}-\underline{A})\underline{b} = \underline{e}$$

$$\underline{I}-\underline{A} = \begin{bmatrix}4/5 & -2/5\\-2/5 & 1/5\end{bmatrix}$$

$$\underline{e} = \begin{bmatrix}2/5\\-1/5\end{bmatrix}$$

$$(\underline{I}-\underline{A})^2\underline{b} = (\underline{I}-\underline{A})\underline{e} = \underline{e}$$ Projects $\underline{e}$ onto the

$(\underline{I}-\underline{A})^2\underline{b} = (\underline{I}-\underline{A})\underline{e} = \underline{e}$

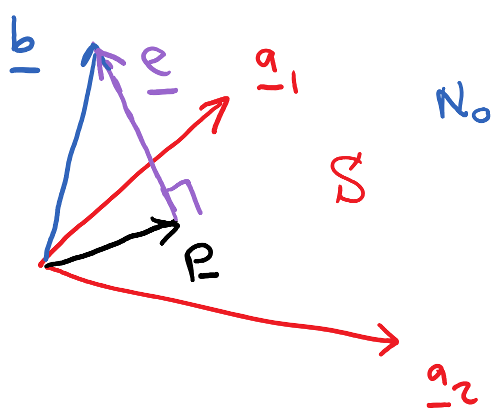
Projects $\underline{e}$ onto the perpendicular space of $\underline{a}$ or onto itself!

$\therefore$ $\underline{I}-\underline{A}$ is idempotent

Projection onto Subspaces

Next consider the projection of a vector $\underline{b}$ in $\mathbb{R}^m$ onto a subspace S in $\mathbb{R}^n$

Let S be spanned by vectors $\underline{a}_1$ and $\underline{a}_2$



Note: $\underline{e} \perp S$

$\underline{p} + \underline{e} = \underline{b}$

$\underline{e} = \underline{b} - \underline{p}$

Also, $\underline{e} \cdot \underline{p} = 0$ and $\underline{e} \cdot \underline{a}_1 = 0$, $\underline{e} \cdot \underline{a}_2 = 0$

More generally, let the subspace S be spanned by $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$

We want to find $\underline{A}$, such that

$$\underline{p} = \underline{A}\,\underline{\hat{x}}$$

where the columns of $\underline{A}$ span the subspace and $\underline{\hat{x}}$ is the coordinates ("weights") of the column space of $\underline{A}$

As before, the error vector $\underline{e}$ is perpendicular to subspace S

$$\Rightarrow \quad \underline{e} \cdot (\text{any vector in } S) = 0$$

Thus, since $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ are in S, then

$$\underline{a}_1 \cdot \underline{e} = \underline{a}_1^T \underline{e} = 0$$
$$\underline{a}_2 \cdot \underline{e} = \underline{a}_2^T \underline{e} = 0$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$\underline{a}_n \cdot \underline{e} = \underline{a}_n^T \underline{e} = 0$$

which can be written as

$$\begin{bmatrix} - & \underline{a}_1^T & - \\ - & \underline{a}_2^T & - \\ & \vdots & \\ - & \underline{a}_n^T & - \end{bmatrix} \underline{e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\Rightarrow \underline{A}^T \underline{e} = \underline{0}$$

But $\underline{e} = \underline{b} - \underline{p} = \underline{b} - \underline{A}\hat{\underline{x}}$

$$\underline{A}^T \underline{e} = \underline{A}^T (\underline{b} - \underline{A}\hat{\underline{x}}) = \underline{0}$$

$$\Rightarrow \boxed{\underline{A}^T \underline{A} \hat{\underline{x}} = \underline{A}^T \underline{b}}$$

or $\hat{\underline{x}} = (\underline{A}^T \underline{A})^{-1} \underline{A}^T \underline{b}$

We wanted a matrix such that the matrix times $\underline{b}$ gives $\underline{p}$

$$\underline{p} = \underline{A}\hat{\underline{x}} = \underbrace{\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T}_{\text{This is the matrix that will project}} \underline{b}$$

any vector $\underline{b}$ in $\mathbb{R}^m$ onto the space spanned by the columns of $\underline{A}$ (in $\mathbb{R}^n$)

What if S is spanned by one vector, say $\underline{a}$?

$$\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T = \underline{a}(\underline{a}^T\underline{a})^{-1}\underline{a}^T$$

$$= \underbrace{(\underline{a}^T\underline{a})^{-1}}_{\text{scalar}}\underline{a}\,\underline{a}^T = \frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}}$$

Yes, exactly the previous result!

Question: Why is the following not true in general?

$$\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T \overset{?}{=} \underline{A}\,\underline{A}^{-1}\underline{A}^{-T}\underline{A}^T = \underline{I}\,\underline{I} = \underline{I}$$

Because we do not know if $\underline{A}^{-1}$ and $\underline{A}^{-T}$ exist! For example, $\underline{A}$ may not even be square, as in case above with $\underline{A} = \underline{a}$

But why is $(\underline{A}^T\underline{A})^{-1}$ ok?

$\underline{A} \in M_{mn}$ $\quad\quad$ $\underline{A}^T \in M_{nm}$

$\underline{A}^T\underline{A} \Rightarrow (n\times m)(m\times n) \Rightarrow n\times n$

where columns of $\underline{A}$ span subspace S in $\mathbb{R}^n$

What if $\underline{A}^{-1}$ does exist?

① $\underline{A} \in M_{nn}$

② The columns of $\underline{A}$ span $\mathbb{R}^n$

$\Rightarrow$ A vector $\underline{b}$ in $\mathbb{R}^n$ projected onto $\mathbb{R}^n$ is nothing but $\underline{b}$ itself

$\Rightarrow$ If $\underline{A}^{-1}$ exists, then

$$\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T \text{ must equal } \underline{I},$$

where

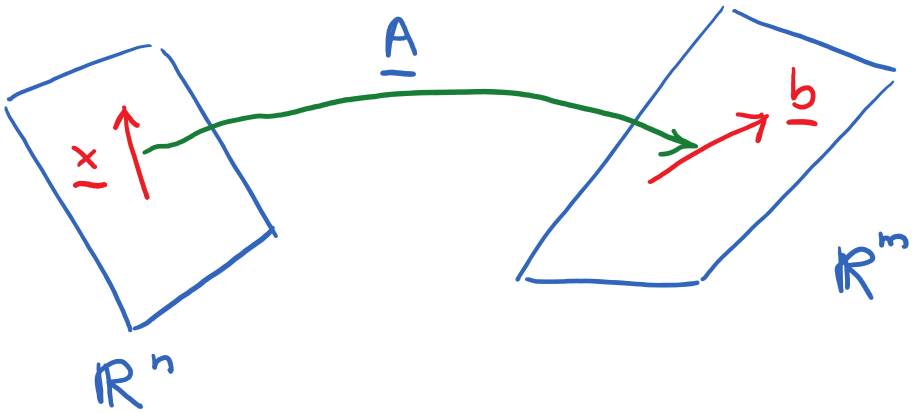
$$\underline{p} = \underline{I}\,\underline{b} = \underline{b}$$

$$\underline{e} = (\underline{I} - \underline{I})\underline{b} = \underline{0}$$

Least Squares Approximations

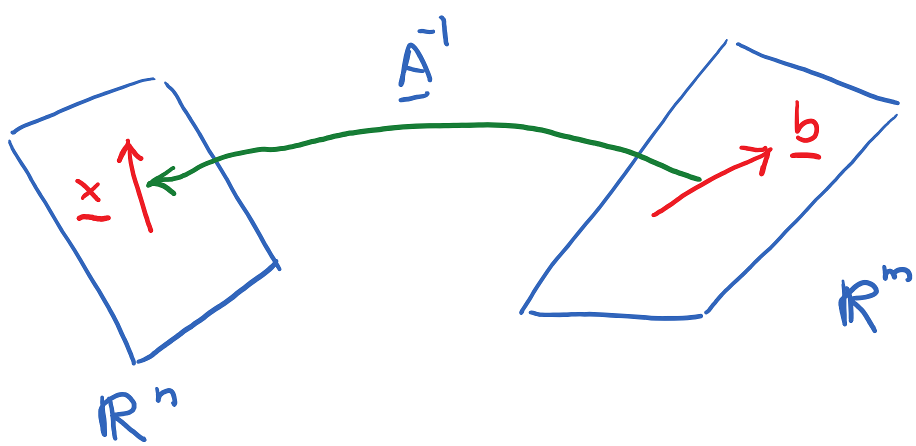
Consider what a linear operator $\underline{A}$ does to $\underline{x}$

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



Given $\underline{A}$ & $\underline{x}$, there definitely is always a vector $\underline{b}$

Is the reverse true?



This mapping only exists, if $\underline{A}^{-1}$ exists, which is not always true

If $\underline{A}^{-1}$ does not exist, then can we find an approximate solution, called $\hat{\underline{x}}$, that does lie in $\mathbb{R}^n$?

Let's try to minimize $\underline{e} = \underline{b} - \underline{p} = \underline{b} - \underline{A}\hat{\underline{x}}$

This is a projection onto a subspace! Using $\underline{I} - \underline{A}$

The solution that minimizes $\underline{e}$ is

$$\underline{A}^T \underline{A} \hat{\underline{x}} = \underline{A}^T \underline{b}$$

Here $\hat{\underline{x}}$ is the solution that minimizes

$$\underline{e} = \underline{A}\hat{\underline{x}} - \underline{b} \quad \text{or} \quad \|\underline{e}\|_2 = \|\underline{A}\hat{\underline{x}} - \underline{b}\|_2$$

Consider the following:

If $\underline{A}^{-1}$ does not exist, then $\underline{A}\underline{x} = \underline{b}$ has a solution iff $\underline{b}$ is in the $C(\underline{A})$

If $\underline{b}$ is not in $C(\underline{A})$, then project $\underline{b}$

onto $C(\underline{A})$

Don't forget, $\hat{\underline{x}}$ does not solve $\underline{A}\underline{x} = \underline{b}$

Note: The columns of $\underline{A}$ must still be independent for this to work

The set of equations given by

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

are called the Normal Equations

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

$$\Rightarrow \underline{x} = \underbrace{(\underline{A}^T \underline{A})^{-1}} \underline{A}^T \underline{b}$$

This is very difficult to compute

Why? Consider the condition number of $\underline{A}^T \underline{A}$

$$\kappa(\underline{A}^T \underline{A}) = ?$$

First, look at $\kappa(\underline{A}\,\underline{B})$

$$\kappa(\underline{A}\,\underline{B}) = \|\underline{A}\,\underline{B}\|\,\|(\underline{A}\,\underline{B})^{-1}\| = \|\underline{A}\,\underline{B}\|\,\|\underline{B}^{-1}\underline{A}^{-1}\|$$

$$\le \|\underline{A}\|\,\|\underline{B}\|\,\|\underline{B}^{-1}\underline{A}^{-1}\|$$

$$\le \|\underline{A}\|\,\|\underline{B}\|\,\|\underline{B}^{-1}\|\,\|\underline{A}^{-1}\| = \kappa(\underline{A})\kappa(\underline{B})$$

$$\Rightarrow \kappa(\underline{A}\,\underline{B}) \sim \kappa(\underline{A})\,\kappa(\underline{B})$$

Also $\kappa(\underline{A}^T) = \kappa(\underline{A})$ (Will show later)

Then,

$$\kappa(\underline{A}^T\underline{A}) = \kappa(\underline{A}^T)\kappa(\underline{A}) = \kappa^2(\underline{A})$$

$\Rightarrow$ if $\underline{A}$ is not well-conditioned, then

$\underline{A}^T\underline{A}$ is even worse!

For example, if $\kappa(\underline{A}) \sim 10^5$, then

$$\kappa(\underline{A}^T\underline{A}) \sim 10^{10}$$

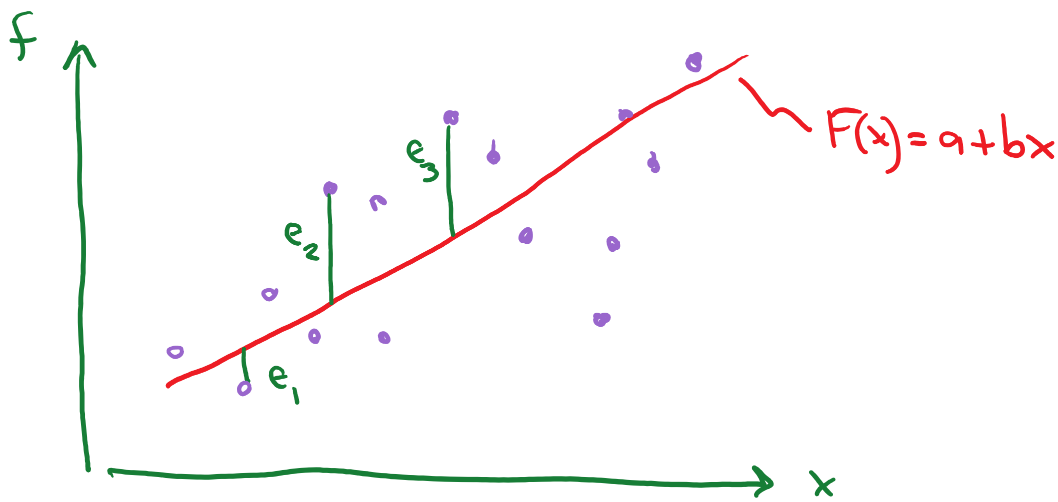
Note: In a few weeks, we will discuss matrix decompositions that allow one to solve

$$\underline{A}^T \underline{A}\, \underline{x} = \underline{A}^T \underline{b} \qquad \text{(or similar)}$$

without issue

Example: Curve Fitting

Consider many data points: (x_i, f_i)



Want to find an approximate relation

$F(x) = a + bx$ that describes the data

Want to minimize the error written as

$$e_1^2 + e_2^2 + \dots + e_m^2 \qquad \text{for } m \text{ points}$$

At each point, we wish that

$$a + bx_i = f_i$$

would hold, or

$$\begin{cases} a + b x_1 = f_1 \\ a + b x_2 = f_2 \\ \quad \vdots \quad \vdots \quad \vdots \\ a + b x_m = f_m \end{cases}$$

Here x_i & f_i are known,

while a & b are the unknowns

$$\Rightarrow \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}}_{\underline{A} \;\rightarrow\; \text{rank} = 2} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}$$

i.e., $\underline{A}\,\hat{\underline{x}} = \underline{b}$

If at least one of the x_i are distinct, then columns of $\underline{A}$ are independent

If there are more than two data points, then $\underline{A}^{-1}$ does not exist, and this is called an over-constrained system,

but $\underbrace{(\underline{A}^T \underline{A})^{-1}}$ does exist   Why?

size? 2×2 is full rank (2)

$\Rightarrow$ Solve

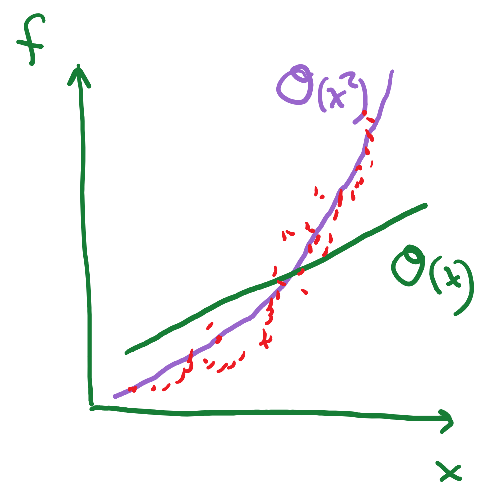
$$\underline{A}^T \underline{A} \begin{bmatrix} a \\ b \end{bmatrix} = \underline{A}^T \underline{f} \qquad \text{(Normal equations)}$$

for a & b to minimize $\|\underline{e}\|_2$

Generalize to higher order:

$$F(x) = a + bx + \underbrace{cx^2}$$

$$a + bx_1 + cx_1^2 = f_1$$
$$a + bx_2 + cx_2^2 = f_2$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$a + bx_m + cx_m^2 = f_m$$



$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{bmatrix}}_{\underline{A}} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}$$

Notice that $F(x)$ can be non-linear,

Solve $A^T A \begin{bmatrix} a \\ b \end{bmatrix} = A^T f$

Solve $\underline{\bar{A}}^T \underline{A} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \underline{A}^T \underline{f}$

can be non-linear, but least squares is still a linear problem!

Multidimensional:

Fit $F(x, y) = a + bx + cy + dxy$

to the data (x_i, y_i, f_i)

$$\begin{bmatrix} 1 & x_1 & y_1 & x_1 y_1 \\ 1 & x_2 & y_2 & x_2 y_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & y_m & x_m y_m \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}$$

Generic functions:

Imagine data (x_i, g_i) for $i = 1, 2, \ldots, m$

Let the interpolent be

$$F(x) = a f_0(x) + b f_1(x) + c f_2(x) + \ldots$$

where $f_0(x), f_1(x), f_2(x), \ldots$ are some functions

e.g. $f_0(x) = x$

$f_1(x) = 1 - x$

$$f_2(x) = x^2 - x$$

$$\vdots$$

Then,

$$\underbrace{\begin{bmatrix} f_0(x_1) & f_1(x_1) & f_2(x_1) & \cdots \\ f_0(x_2) & f_1(x_2) & f_2(x_2) & \cdots \\ \vdots & \vdots & \vdots & \\ f_0(x_m) & f_1(x_m) & f_2(x_m) & \cdots \end{bmatrix}}_{\underline{A}} \begin{bmatrix} a \\ b \\ c \\ \vdots \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix}$$

Solve $\underline{A}^T \underline{A} \begin{bmatrix} a \\ b \\ c \\ \vdots \end{bmatrix} = \underline{A}^T \underline{g}$ (Normal equations)

Aside: Many processes exhibit power law behavior $F(x) \cong a x^\beta$

Then, $\underbrace{\ln F}_{G} \cong \underbrace{\ln a}_{\alpha} + \beta \underbrace{\ln x}_{y}$

Use least squares to estimate $\alpha + \beta$

Normal Equations: Error Minimization

Consider fitting $f(x) = a + bx$ to a set of data (x_i, f_i)

Resulting system

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}$$

$$\underline{A} \quad \underline{x} \quad = \quad \underline{b}$$

It was stated that the solution to

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

minimizes the error $\|\underline{e}\|_2$

Let's show the connection. Here,

$$e = \sum \left((a + bx_i) - f_i \right)^2$$

$$= \sum \left(a + bx_i - f_i \right)^2$$

For the minimum, the gradient wrt a & b must be zero

$$\frac{\partial e}{\partial a} = \sum (a + bx_i - f_i) = 0$$

$$\frac{\partial e}{\partial b} = \sum x_i (a + bx_i - f_i) = 0$$

Rewrite as a linear system:

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum f_i \\ \sum x_i f_i \end{bmatrix}$$

for n points (x_i, f_i)

Does $\underline{A}^T \underline{A}\, \underline{x} = \underline{A}^T \underline{b}$ give the same

2×2 system?

For simplicity, let $n = 3$

$(x_1, f_1), (x_2, f_2), (x_3, f_3)$

$$\underline{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix}, \quad \underline{b} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}, \quad \underline{x} = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\underline{A}^T \underline{A} = \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \end{bmatrix} = \begin{bmatrix} 3 & x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 & x_1^2 + x_2^2 + x_3^2 \end{bmatrix}$$

$$\underline{A}\,\underline{A} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} 1 & x_2 \\ 1 & x_3 \end{bmatrix} = \begin{bmatrix} 3 & x_1+x_2+x_3 \\ x_1+x_2+x_3 & x_1^2+x_2^2+x_3^2 \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$\underline{A}^T \underline{b} = \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} f_1 + f_2 + f_3 \\ x_1 f_1 + x_2 f_2 + x_3 f_3 \end{bmatrix}$$

$$= \begin{bmatrix} \sum f_i \\ \sum x_i f_i \end{bmatrix}$$

$\Rightarrow$ $\underline{A}^T \underline{A}\, \underline{x} = \underline{A}^T \underline{b}$ results in the same system that minimizes $\underline{e}$ ✓