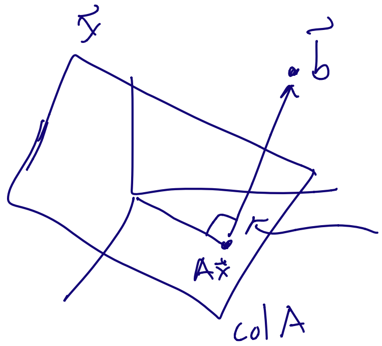
Numerical Analysis 2020-03-24

Last Time: Least Squares Problems

$$\underset{\vec{x}}{\text{minimize}} \quad \| A\vec{x} - \vec{b} \|_2^2 \longleftarrow \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right)^2$$



$$(\vec{b} - A\vec{x}) \perp \text{col } A$$

$$\Rightarrow \quad \underbrace{A^T}\,\underbrace{A\vec{x}} = A^T \vec{b}.$$

Normal Equations

Solving the normal equations is bad idea numerically:
The condition number (relative ℓ_2) of A^TA
is the square of that of A.

Instead solve a consistent system:

$$A\vec{x} = \text{proj}_{\text{col}A}\, b \quad \Rightarrow \quad \text{If } \hat{u}_1 \cdots \hat{u}_n \text{ is an}$$

orthonormal basis for col A,

$$\text{then } UU^T\vec{b} = (\hat{u}_1 \cdots \hat{u}_n) \begin{pmatrix} \hat{u}_1^T \\ \vdots \\ \hat{u}_n^T \end{pmatrix} \vec{b}$$

is the projection of $\vec{b}$ onto col A.

$\Rightarrow \quad A\vec{x} = UU^T\vec{b}$ is consistent.

The columns of U are orthonormal, so U is an orthogonal matrix $\Rightarrow$ $U^TU = I$.

Use Gram-Schmidt process.

G-S Process:

Goal: Given some set of vectors $\vec{a}_1, \dots \vec{a}_n$, form another sequence of vectors $\hat{q}_1, \hat{q}_2 \dots \hat{q}_n$ that are orthonormal.

Let: $\hat{q}_1 = \vec{a}_1 / \|\vec{a}_1\| \Rightarrow \|\hat{q}_1\| = 1$ All norms are l_2 norms.

$$\vec{q}_2' = \vec{a}_2 - \underbrace{(\vec{a}_2, \hat{q}_1)}_{\text{inner product}} \hat{q}_1$$

is not normalized

inner product $(\vec{x}, \vec{y}) = \vec{x}^T \vec{y}$.

$$\hat{q}_2 = \vec{q}_2' / \|\vec{q}_2'\|$$

^ means unit vector.

Proceed ...

$$\vec{q}_3' = \vec{a}_3 - (\vec{a}_3, \hat{q}_2)\hat{q}_2 - (\vec{a}_3, \hat{q}_1)\hat{q}_1$$

$$\hat{q}_3 = \vec{q}_3' / \|\vec{q}_3'\|$$

So when we are done, by construction,

$Q = (\hat{q}_1 \ \hat{q}_2 \ \cdots \ \hat{q}_n)$ then $Q^T Q = I$

And the projection of $\vec{x}$ onto the column space of A, i.e. $\text{Span}\{\vec{a}_1, \dots, \vec{a}_n\}$ is just $\text{proj}_A \vec{x} = Q Q^T \vec{x}$

Furthermore, this automatically yields a factorization of the matrix A:

$$\text{col } Q = \text{col } A$$

So any $\hat{q}_j = c_1 \vec{a}_1 + \dots + c_n \vec{a}_n$

and likewise $\vec{a}_j = d_1 \hat{q}_1 + \dots + d_n \hat{q}_n$

This gives us the QR factorization:

$$A = Q\,R$$

$$\begin{pmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{pmatrix} = \begin{pmatrix} \hat{q}_1 & \hat{q}_2 & \cdots & \hat{q}_n \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ 0 & r_{22} & & & \\ 0 & & \ddots & & \\ \vdots & & & & r_{nn} \end{pmatrix}$$

The question is: what are the r_{ij}'s?

r_{ij}'s can be obtained from G-S process.

<u>Ex:</u> From G-S:

$$\hat{q}_1 = \vec{a}_1 / \|\vec{a}_1\| \implies \vec{a}_1 = \|\vec{a}_1\|\,\hat{q}_1 \implies r_{11} = \|\vec{a}_1\|$$

$$\vec{q}_2\,' = \vec{a}_2 - (\vec{a}_2, \hat{q}_1)\,\hat{q}_1$$

$$\vec{q}_2\,' = \|\vec{a}_2 - (\vec{a}_2, \vec{q}_1)\,\hat{q}_1\|\,\hat{q}_2$$

$$\implies \vec{a}_2 = \underbrace{(\vec{a}_2, \hat{q}_1)}\,\hat{q}_1 + \vec{q}_2\,'$$

$$= r_{12}\,\hat{q}_1 + r_{22}\,\hat{q}_2$$

With sufficient bookkeeping, the G-S process yields the factorization $A = QR$.

So now: to solve least squares:

① $A\vec{x} = Q\underbrace{Q^T\vec{b}}$

Alternatively: ② Given that $A = QR$

Solve $\underbrace{R\vec{x}} = \underbrace{Q^T\vec{b}}$

↑

upper triangular system

Preferred method

Show later on HW or exam...

Applications: Linear Regression

Given data: (x_i^1, x_i^2, y_i) for $i = 1 \ldots n$

Generative model: $y = a + bx_1 + cx_2 + \epsilon$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ $\uparrow$ $N(0,\sigma^2)$ error or measurement noise.

You observe $\underbrace{y_i}_{\text{dependent variable}} = a + b\underbrace{x_i^1}_{\text{independent}} + c\underbrace{x_i^2}_{\text{variables}} + \epsilon_i$

Assumption is that noise only appears in y_i.

Given an estimate of a, b, c: $\hat{a}, \hat{b}, \hat{c}$, then the residuals are given as

$$r_i = \underbrace{y_i}_{\text{observed value}} - \underbrace{(\hat{a} + \hat{b}x_i^1 + \hat{c}x_i^2)}_{\text{predicted value, given estimates } \hat{a}, \hat{b}, \hat{c}.}$$

Question: How do we estimate a, b, c?

One option: minimize the squared residuals:

$$\min_{a,b,c} \|\vec{r}\|_2^2 = \min_{a,b,c} \sum_{i=1}^{n} (y_i - a - bx_i^1 - cx_i^2)^2$$

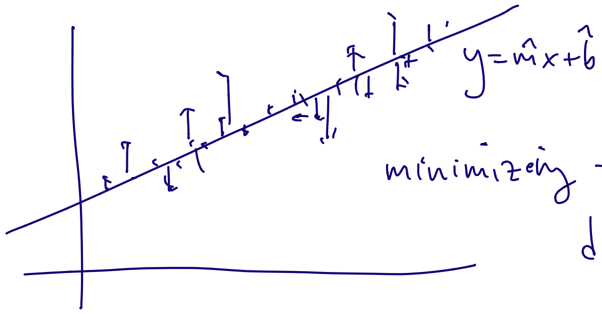
This is a least squares problem.

Form the least square system:

$$\min_{a,b,c} \left\| \underbrace{\begin{pmatrix} 1 & x_1^1 & x_1^2 \\ 1 & x_2^1 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n^1 & x_n^2 \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} a \\ b \\ c \end{pmatrix}}_{\vec{a}} - \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{\vec{y}} \right\| = \min_{\vec{a}} \| X\vec{a} - \vec{y} \|$$

design matrix

Graphically



Fit $y = mx + b$ to data.

minimizing the vertical errors between the data y_i and the line $y = mx + b$.

One additional method to solve least squares problems:

Using the SVD.

Recall: If $A \in \mathbb{R}^{m \times n}$, $m \geq n$, rank $A = n$.

Then we can write $A = U S V^T$

U: $m \times n$ orthogonal; S: diagonal $n \times n$; V^T: $n \times n$ orthogonal.

Special case: If A is invertible $(m = n)$ then

$$A^{-1} = (U S V^T)^{-1} = V^{-T} S^{-1} U^{-1}$$

but U, V are orthogonal $\Rightarrow V^T = V^{-1}$

$$= V S^{-1} U^T$$

If A is not invertible $(m > n)$

Define the pseudo-inverse of A to be:

$$A^{+} = V S^{-1} U^T$$

$\Rightarrow$ Even though A is not invertible,

$$A^{+}A = (V S^{-1} U^T)(U S V^T)$$

$$= V \underbrace{S^{-1} \underbrace{U^T U}_{I} S}_{I} V^T$$

$$= V V^T = I \quad \text{since } V \text{ is square.}$$

HW/Exam problem:

Show that the least squares solution to $A\vec{x} = \vec{b}$ is $\vec{x} = V S^{-1} U^T \vec{b}$.