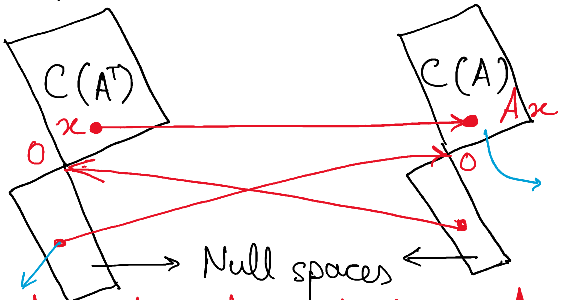
Least Squares

Tuesday, July 7, 2020 16:29

A: $m \times n$ matrix.

The pseudo-inverse of the matrix A: denoted A^+ is $n \times m$.

For a square matrix B, $B^+ = B^{-1}$ if the inverse exists.



C(A) denotes column space of A.

Here we have a nice basis $v_1, v_2, \ldots v_r$ for the nice space

These vectors transform to 0 as $Ax = 0$ for them.

The null space is what makes a matrix non-invertible.
In the upper space, every vector x transformed to Ax and can be transformed back

→ Thus, the pseudo-inverse should:

1. Take everything in $C(A)$ back to x: $A^+(Ax) = x$
2. Take everything in the null space on the right to 0.
   $A^+(y) = 0$ for y in this space.

<u>How to get a nice expression for A^+?</u>

Let $A = U \Sigma V^T$

↳ All singular values are non-0 if A is invertible.

↳ In this case, $A^{-1} = V \Sigma^{-1} U^T$

More generally, Σ may have 0 singular values: $\sigma_1, \ldots \sigma_r, 0, \ldots, 0$.

$$\boxed{A^+ = V \Sigma^+ U^T}$$

If $\underset{(m \times n)}{\Sigma} = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}$ then $\underset{(n \times m)}{\Sigma^+} = \begin{bmatrix} 1/\sigma_1 & & & & \\ & \ddots & & & \\ & & 1/\sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}$

So, $\Sigma^+ \Sigma = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$ (the best we can do, close to identity)

The least^2 problem

We have $Ax = b$ to solve for x.

If b is in the column space of A, then we can just find x with the pseudo-inverse.

Let A be an $m \times n$ matrix with rank r.

A is invertible if $m = n = r$

→ Typically b is some vector of measurements (maybe noisy)
e.g. for R^2, A would be two columns.

$$\underset{(m\times 2)}{Ax = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_m \end{bmatrix}} \underset{(2\times 1)}{\begin{bmatrix} C \\ D \end{bmatrix}} = \underset{(m\times 1)}{\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}}$$

where the line we're looking for is $y = C + Dx$.

→ Want to find $\begin{bmatrix} C \\ D \end{bmatrix}$. If points $b_1, \dots, b_m$ fit perfectly on a line then $\exists$ a solution and we have C, D.

But what if the points are not on a line?

We try to: minimize $\|Ax - b\|_2^2 = (Ax-b)^T(Ax-b)$ to get the best approximation.

$$\Rightarrow x^T A^T A x - b^T A x - x^T A^T b + b^T b$$

(in learning terms, this is the loss, and loss is being minimized).

$$= x^T A^T A x - \underbrace{2x^T A^T b}_{\text{bcoz } b^TAx \text{ is scalar}} + b^T b$$

In statistics, this is linear regression.

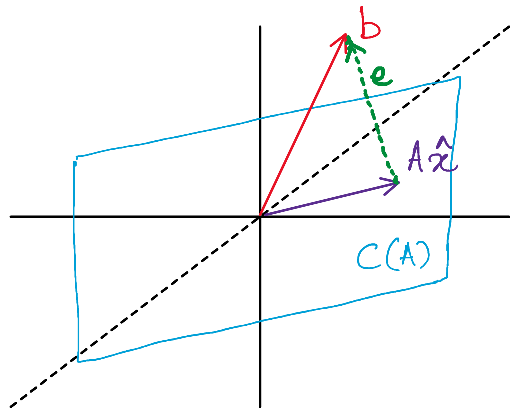
The best $x \rightarrow \hat{x}$ will come from:

$\boxed{A^T A \hat{x} = A^T b}$ → Can simply do this with some code.

↳ Take the derivative of loss and put $=0$ to get this $\hat{x}$.

We have $C(A) = \{\text{all possible } Ax\}$

Given $b \notin C(A)$ $\rightarrow$ otherwise we don't need any loss minimization.



e: Error $b - A\hat{x}$

b: The given data points

$A\hat{x}$: Projection of b on $C(A)$, the best x we get with this method.

Note: If A has independent columns, then $A^T A$ is invertible and we can directly find $\hat{x}$

Claim: $A^+ b = (A^T A)^{-1} A^T b$ when $N(A) = 0$, $r = n$

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

Side note: $((A^T A)^{-1} A^T) A = I$, but $(A^T A)^{-1} A^T$ is not the inverse of A. It doesn't work on the other side, i.e. $A(A^T A)^{-1} A^T \neq I$.

To check this, put $A = U \Sigma V^T$ on RHS and see that Σ^+ pops up!