Back to least squares

$$\left.\begin{aligned} 2x &= b_1 \\ 3x &= b_2 \\ 4x &= b_3 \end{aligned}\right\}$$ This system is solvable if b is on the line through $\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$

Suppose we have a vector "b" that leads to an "inconsistent" system.

We could pick a subset of equations & solve it exactly.

Problem with this approach: large errors in some inputs & no error in others

Reasonable alternative: minimize average error

$$E^2 = (2x-b_1)^2 + (3x-b_2)^2 + (4x-b_3)^2$$

Want to minimize the sum of squares E^2.

$$\frac{dE^2}{dx} = 0 \iff 2\left[2(2x-b_1) + 3(3x-b_2) + 4(4x-b_3)\right] = 0$$

leading to $\hat{x} = \dfrac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2} = \dfrac{a^T b}{a^T a}$ with $a = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$

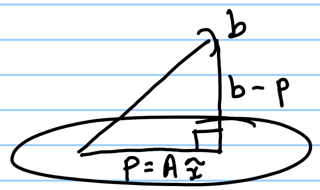
connects to projections

Bottom line: Taking derivative & finding the minima turns out to be the same as performing a projection.

Projection onto a subspace:

$Ax = b$, A is $m \times n$, $m > n$

Want: projection of b onto column space $C(A)$



S = span (columns of A)

Projection of b onto S is $p = A\hat{x}$

Orthogonal vector $e = b - p = b - A\hat{x}$

Q: How to find $\hat{x}$?

Observe $e \perp$ every vector in $C(A)$

Recall that $C(A) \perp N(A^T)$, i.e., $N(A^T)$ is the orthogonal complement of $C(A)$, i.e., every vector in $C(A)$ is orthogonal to every vector in $N(A^T)$ & any given vector is in either $C(A)$ or $N(A^T)$

Where does e belong? $e \in N(A^T) \Rightarrow A^T e = 0 \Rightarrow A^T(b - A\hat{x}) = 0$

Note Title 4/19/2021

leading to $\boxed{A^T A \hat{x} = A^T b}$ → equation to solve to obtain the projection of b onto $C(A)$.

(Note: Even if $Ax=b$ is not solvable, $A^T A \hat{x} = A^T b$ has a solution)

Alternative route to the above equation:

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

$$\left.\begin{matrix} a_1^T e = 0 \\ \vdots \\ a_n^T e = 0 \end{matrix}\right\} \Leftrightarrow \begin{matrix} a_1^T (b - A\hat{x}) = 0 \\ \vdots \\ a_n^T (b - A\hat{x}) = 0 \end{matrix}$$

$$\Updownarrow$$

$$\begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} \left[b - A\hat{x} \right] = 0 \Leftrightarrow A^T (b - A\hat{x}) = 0$$

leading to $\boxed{A^T A \hat{x} = A^T b}$

Note Title 4/19/2021

Bottomline: Solving $A^T A \hat{x} = A^T b$ leads to an $\hat{x}$ that minimizes $\|Ax - b\|^2$

connection of projections to least squares

Remarks:

① Suppose columns of A are linearly independent (l.i.)

Then, $\underbrace{A^T A}_{\text{square, symmetric}}$ is invertible (Why?)

Solving $A^T A \hat{x} = A^T b$ when $(A^T A)$ is invertible

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

Projection $P = A\hat{x} = A(A^T A)^{-1} A^T b$

Note Title 4/19/2021

(2) $b \in C(A)$ i.e., $b = Ax$

$$p = A(A^TA)^{-1}A^Tb = A\underbrace{(A^TA)^{-1}A^TA}_{=I}x = Ax = b$$

(3) $b \in N(A^T)$

$$p = A(A^TA)^{-1}A^Tb = 0 \quad \text{since } A^Tb = 0$$

(4) A is square & invertible $\Leftrightarrow C(A) = \mathbb{R}^n$

$$p = A(A^TA)^{-1}A^Tb = AA^{-1}(A^T)^{-1}A^Tb = b$$

(5) A is rank one i.e., $A = \begin{bmatrix} | \\ a \\ | \end{bmatrix}$

Then, $\hat{x} = \dfrac{a^Tb}{a^Ta}$ $\rightarrow$ coincides with what we derived earlier for projection onto a line.

Note Title 4/19/2021

Projection matrix $\mathbb{P} = A(A^T A)^{-1} A^T$

Symmetric $\mathbb{P}^T = \mathbb{P}$

$$\left(A (A^T A)^{-1} A^T\right)^T = (A^T)^T \left((A^T A)^{-1}\right)^T A^T$$
$$= A (A^T A)^{-1} A^T = \mathbb{P}$$

$\mathbb{P}^2 = \mathbb{P}$

$$\mathbb{P}^2 = A (A^T A)^{-1} A^T A (A^T A)^{-1} A^T$$
$$= A (A^T A)^{-1} A^T = \mathbb{P}$$

So, Projection matrix is symmetric & satisfies $\mathbb{P}^2 = \mathbb{P}$

The converse is also true.

Claim: If $\mathbb{P}^2 = \mathbb{P}$ & $\mathbb{P}$ is symmetric, then $\mathbb{P}$ is a projection matrix

$\mathbb{P} b =$ projection of b onto the column space of $\mathbb{P}$

$$(b - \mathbb{P} b)^T \mathbb{P} c = b^T (I - \mathbb{P})^T \mathbb{P} c = b^T (\mathbb{P} - \mathbb{P}^2) c = 0 \text{ since } \mathbb{P}^2 = \mathbb{P}.$$