# LEAST SQUARES

Example:

$A$ : 2x2 matrix (given matrix)

$\bar{b} \in \mathbb{R}^2$ vector (given vector)

find the vector $\bar{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ that fullfills $A\bar{x} = \bar{b}$

$$A = \begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix} \qquad \bar{b} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

i.e. find the solution to the following system of equations

$$\left\{ \begin{aligned} 2x_1 + 3x_2 &= -1 \\ 3x_1 + 4x_2 &= 0 \end{aligned} \right. \Rightarrow \text{The Solution is } \boxed{\begin{aligned} x_1 &= 4 \\ x_2 &= -3 \end{aligned}}$$

but what happens if we add another constraint

$$x_1 + 2x_2 = 2$$

we have an over-determined system of equations

$$\begin{pmatrix} 2 & 3 \\ 3 & 4 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}$$

$\Rightarrow$ There is no $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ that fullfills the three equations!!

but we can find some solution that minimizes some kind of an error.

Let write the "error"

$$\|A\bar{x}-\bar{b}\|_2^2$$

then the problem is

$$\min_{\bar{x}\in\mathbb{R}^2} \|A\bar{x}-\bar{b}\|_2^2$$

$$\Downarrow$$

$$\min_{x\in\mathbb{R}^2} \langle A\bar{x}-\bar{b}, A\bar{x}-b\rangle$$

Remember $\|u\|_2=\sqrt{\langle u, u\rangle}=\sqrt{\sum_i^n u_i^2}$

$\|u\|_p=\left(\sum_{i=1}^n |u_i|^p\right)^{1/p}$

in our example $A\bar{x}-b$ is

$$A\bar{x}-b=\begin{pmatrix}2 & 3\\ 3 & 4\\ 1 & 2\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}-\begin{pmatrix}-1\\ 0\\ 2\end{pmatrix}=\begin{pmatrix}2x_1+3x_2+1\\ 3x_1+4x_2-0\\ x_1-2x_2-2\end{pmatrix}$$

$$\langle A\bar{x}-b, A\bar{x}-b\rangle=(2x_1+3x_2+1)^2+(3x_1+4x_2-0)^2+$$
$$(x_1-2x_2-2)^2$$

$$\Downarrow$$

name this as $f(\bar{x})$

then we have to find the $\min_{x_1,x_2} f(\bar{x})$

For doing that we have several options:

1- Compute gradient of $f(\bar{x})$ and set to 0
2- Normal equations
3- Gradient descent

Actually use the same

1. Compute $\nabla f(x)$ and set to zero in our example:

$$f(\bar{x}) = (2x_1 + 3x_2 + 1)^2 + (3x_1 + 4x_2 - 0)^2 + (x_1 - 2x_2 - 2)^2$$

$$\frac{\partial f(\bar{x})}{\partial x_1} = 2(2x_1 + 3x_2 + 1) \cdot 2 + 2(3x_1 + 4x_2 - 0) \cdot 3 + 2(x_1 - 2x_2 - 2)$$

$$\frac{\partial f(\bar{x})}{\partial x_2} = 2(2x_1 + 3x_2 + 1) \cdot 3 + 2(3x_1 + 4x_2 - 0) \cdot 4 + 2(x_1 - 2x_2 - 2) \cdot (-2)$$

$\vdots$

# NORMAL EQUATIONS

Remember $D_{\bar{v}} f(\bar{x}) = \lim_{\varepsilon \to 0} \dfrac{f(\bar{x}+\varepsilon\bar{v}) - f(\bar{x})}{\varepsilon}$

take the $D_{\bar{v}} \langle A\bar{x}-\bar{b}, A\bar{x}-\bar{b} \rangle$

$$\lim_{\varepsilon \to 0} \frac{\langle A(\bar{x}+\varepsilon\bar{v})-\bar{b}, A(\bar{x}+\varepsilon\bar{v})-\bar{b}\rangle - \langle A\bar{x}-\bar{b}, A\bar{x}-\bar{b}\rangle}{\varepsilon}$$

$$\lim_{\varepsilon \to 0} \frac{\langle A\bar{x}+\varepsilon A\bar{v}-b, A\bar{x}+\varepsilon A\bar{v}-b\rangle - \langle A\bar{x}-\bar{b}, A\bar{x}-\bar{b}\rangle}{\varepsilon}$$

$$\lim_{\epsilon \to 0} \frac{\cancel{\langle A\bar{x}-b, Ax-b\rangle} + \langle A\bar{x}-b, \epsilon A\bar{v}\rangle + \langle \epsilon A\bar{v}, A\bar{x}-b\rangle + \langle \epsilon A\bar{v}, \epsilon A\bar{v}\rangle - \cancel{\langle A\bar{x}-\bar{b}, A\bar{x}-\bar{b}\rangle}}{\epsilon}$$

$$\lim_{\epsilon \to 0} \frac{2\langle A\bar{x}-b, \epsilon A\bar{v}\rangle + \langle \epsilon A\bar{v}, \epsilon A\bar{v}\rangle}{\epsilon}$$

$$\lim_{\epsilon \to 0} \frac{2\cancel{\epsilon}\langle A\bar{x}-\bar{b}, A\bar{v}\rangle + \epsilon^{\cancel{2}}\langle A\bar{v}, A\bar{v}\rangle}{\cancel{\epsilon}}$$

Remember:

$D_{\bar{v}} f(\bar{x}) = \langle \nabla f(\bar{x}), \bar{v}\rangle$

$$\lim_{\epsilon \to 0} 2\langle A\bar{x}-\bar{b}, A\bar{v}\rangle + \epsilon\langle A\bar{v}, A\bar{v}\rangle = 2\langle A\bar{x}-b, A\bar{v}\rangle = \langle 2A^{T}A\bar{x}-b, \bar{v}\rangle$$

$$\boxed{\nabla f(\bar{x}) = 2A^{T}A\bar{x} - A^{T}\bar{b}}$$

Remember the necessary condition for minimum: $\nabla f(\bar{x}) = 0$, Then

$2A^{T}A\bar{x}-b=0 \Rightarrow A^{T}A\bar{x}-A^{T}\bar{b}=0 \Rightarrow \boxed{\bar{x} = (A^{T}A)^{-1}A^{T}\bar{b}}$

$\hookrightarrow$ Normal equations

In our example:

$$A^T = \begin{pmatrix} 2 & 3 & 1 \\ 3 & 4 & 2 \end{pmatrix} \quad ; \quad A^TA = \begin{pmatrix} 14 & 20 \\ 20 & 29 \end{pmatrix}$$

$$(A^TA)^{-1} = \begin{pmatrix} 4.83 & -3.3 \\ -3.3 & 2.3 \end{pmatrix} \quad ; \quad (A^TA)^{-1} \cdot A^T = \begin{pmatrix} -0.33 & 1.16 & -1.83 \\ 0.33 & -0.66 & 1.33 \end{pmatrix}$$

$$(A^TA)^{-1} A^T \bar{b} = \boxed{\begin{pmatrix} -3.33 \\ 2.33 \end{pmatrix} = \bar{x}}$$

We can also solve the problem using the SVD Decomposition for computing the pseudo-Inverse

$$x = V \Sigma^+ U^T \bar{b}$$

$A = U \Sigma V^T$

$A^+ = V \Sigma^+ U^T \Rightarrow$ PSEUDO INVERSE

Notice that this is equivalent to solve the normal equations

Proof:

Let's start with the normal equations

$$\bar{x} = (A^TA)^{-1} A^T \bar{b} \qquad A^TA\bar{x} = A^T\bar{b}$$

Decompose $A$ using SVD $\Rightarrow A = U \Sigma V^T$

$$A^TA = (U\Sigma V^T)^T (U \Sigma V^T) = \overbrace{(V \Sigma^T \cancel{U^T})}^{\rightarrow A^T} (\cancel{U} \Sigma V^T) =$$

$\Sigma$

Remember $U^T = U^{-1}$

$$A^TA = V \Sigma^2 V^T$$

$$A^TA\bar{x} = A^T\bar{b} \Rightarrow \cancel{V} \Sigma^2 V^T \bar{x} = \cancel{V} \cancel{\Sigma} U^T b \Rightarrow$$

$$\Rightarrow \Sigma V^T \bar{x} = U^T b \Rightarrow V^T \bar{x} = \Sigma^{-1} U^T b \Rightarrow$$

$\hookrightarrow \Sigma^+$

$$\Rightarrow \bar{x} = V \Sigma^+ U^T b$$

$\rightarrow A^+$ (pseudo-inversa)