Back to Over-Determined System

- Let's start with the original problem:

Person ID	Weight	Height	Is_smoking	Life-span
1	60kg	5.5ft	Yes (=1)	66
2	65kg	5.0ft	No (=0)	74
3	55kg	6.0ft	Yes (=1)	78

$$\underbrace{\begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}^{\mathbf{x}} \overset{=}{=} \overbrace{\begin{bmatrix} 66 \\ 74 \\ 78 \end{bmatrix}}^{\mathbf{b}}$$

- Using the inverse matrix, the solution is $\mathbf{x} = \begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix}$.

Back to Over-Determined System

- Let's add an additional example:

Person ID	Weight	Height	Is_smoking	Life-span
1	60kg	5.5ft	Yes (=1)	66
2	65kg	5.0ft	No (=0)	74
3	55kg	6.0ft	Yes (=1)	78
4	50kg	5.0ft	Yes (=1)	72

$$\rightarrow \underbrace{\begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 0 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix}}_{\mathbf{b}}$$

- Now, let's use the previous solution $\mathbf{x} = \begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix}$

Errors

$$\underbrace{\begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix}}_{\mathbf{x}} = \begin{bmatrix} 66 \\ 74 \\ 78 \\ 60 \end{bmatrix} \neq \underbrace{\begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix}}_{\mathbf{b}} \quad \Bigg| \quad \begin{array}{c} (\mathbf{b} - A\mathbf{x}) \\ 0 \\ 0 \\ 0 \\ 12 \end{array}$$

Back to Over-Determined System

- How about using slightly different solution $\mathbf{x} = \begin{bmatrix} -0.12 \\ 16 \\ -9.5 \end{bmatrix}$?

$$
\begin{array}{ccccccc}
A & \mathbf{x} & & & \neq & \mathbf{b} & (\mathbf{b} - A\mathbf{x}) \\
\begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 1 \end{bmatrix} & \begin{bmatrix} -0.12 \\ 16 \\ -9.5 \end{bmatrix} & = & \begin{bmatrix} 71.3 \\ 72.2 \\ 79.9 \\ 64.5 \end{bmatrix} & \neq & \begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix} & \begin{matrix} -5.3 \\ 1.8 \\ -1.9 \\ 7.5 \end{matrix}
\end{array}
$$

Errors

Which One is Better Solution?

$$\begin{array}{cccccc|c}
A & \mathbf{x} & & & \neq & \mathbf{b} & \begin{matrix}\textbf{Errors}\\(\mathbf{b}-A\mathbf{x})\end{matrix} \\
\begin{bmatrix}60 & 5.5 & 1\\65 & 5.0 & 0\\55 & 6.0 & 1\\50 & 5.0 & 1\end{bmatrix} & \begin{bmatrix}-0.12\\16\\-9.5\end{bmatrix} & = & \begin{bmatrix}71.3\\72.2\\79.9\\64.5\end{bmatrix} & \neq & \begin{bmatrix}66\\74\\78\\72\end{bmatrix} & \begin{matrix}-5.3\\1.8\\-1.9\\7.5\end{matrix} \\
\\
\begin{bmatrix}60 & 5.5 & 1\\65 & 5.0 & 0\\55 & 6.0 & 1\\50 & 5.0 & 1\end{bmatrix} & \begin{bmatrix}-0.4\\20\\-20\end{bmatrix} & = & \begin{bmatrix}66\\74\\78\\60\end{bmatrix} & \neq & \begin{bmatrix}66\\74\\78\\72\end{bmatrix} & \begin{matrix}0\\0\\0\\12\end{matrix}
\end{array}$$

Least Squares: Best Approximation Criterion

- Let's use the squared sum of errors:

Errors | **Sum of squared errors**

$$\underset{A}{\begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 1 \end{bmatrix}} \underset{\mathbf{x}}{\begin{bmatrix} -0.12 \\ 16 \\ -9.5 \end{bmatrix}} = \begin{bmatrix} 71.3 \\ 69 \\ 79.9 \\ 64.5 \end{bmatrix} \neq \underset{\mathbf{b}}{\begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix}} \quad \underset{(\mathbf{b} - A\mathbf{x})}{\begin{matrix} -5.3 \\ 1.8 \\ -1.9 \\ 7.5 \end{matrix}} \quad \left((-5.3)^2 + 1.8^2 + (-1.9)^2 + 7.5^2\right)^{0.5} = 9.55$$

Better solution

$$\begin{bmatrix} 60 & 5.5 & 1 \\ 65 & 5.0 & 0 \\ 55 & 6.0 & 1 \\ 50 & 5.0 & 1 \end{bmatrix} \begin{bmatrix} -0.4 \\ 20 \\ -20 \end{bmatrix} = \begin{bmatrix} 66 \\ 74 \\ 78 \\ 60 \end{bmatrix} \neq \begin{bmatrix} 66 \\ 74 \\ 78 \\ 72 \end{bmatrix} \quad \begin{matrix} 0 \\ 0 \\ 0 \\ 12 \end{matrix} \quad (0^2 + 0^2 + 0^2 + 12^2)^{0.5} = 12$$

Least Squares Problem

- Now, the sum of squared errors can be represented as $\|\mathbf{b} - A\mathbf{x}\|$.
- **Definition**: Given an overdetermined system $A\mathbf{x} \simeq \mathbf{b}$ where $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^{n}$, and $m \gg n$, a least squares solution $\hat{\mathbf{x}}$ is defined a s

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|$$

- The most important aspect of the least-squares problem is that no matter what $\mathbf{x}$ we select, the vector $A\mathbf{x}$ will necessarily be in the column space Col A.
- Thus, we seek for $\mathbf{x}$ that makes $A\mathbf{x}$ as the closest point in Col A to $\mathbf{b}$.