什么是正规方程

简单来说，**正规方程 (Normal Equations)** 是将一个可能无解的线性方程组 **$Ax = y$**，转化为一个可解的新方程组，用来寻找原问题的**最优近似解**（即最小二乘解）。

这个新方程组就是：

$A^TAx = A^Ty$

它的解 **x** 能让 **Ax** 成为 **y** 在 **A** 的列空间上的投影，从而使误差 **$||y - Ax||$** 最小。

从几何角度推导正规方程

在一个三维空间中进行讨论。

1. 问题设定

我们有一个线性方程组 **$Ax = y$**。

- **A** 是一个 m x 3 的矩阵，我们可以将其列向量看作三维空间中的三个向量 **$A = [\alpha_1, \alpha_2, \alpha_3]$**。
- **x** 是一个 3 x 1 的列向量，包含我们想要求的未知数 **$x = [x_1, x_2, x_3]^T$**。
- **y** 是一个 m x 1 的列向量，代表目标向量。

根据矩阵向量乘法的定义，**Ax** 的结果是 **A** 的列向量的线性组合：
$Ax = x_1\alpha_1 + x_2\alpha_2 + x_3\alpha_3$

这个线性组合所能构成的所有向量的集合，被称为矩阵 **A** 的**列空间** (Column Space)。在假设中，**$\alpha_1, \alpha_2, \alpha_3$** 是共面的，所以 **A** 的列空间是一个平面。

2. 几何直觉：投影

假设向量 **y** 不在该平面上。这意味着 **y** 不在 **A** 的列空间中。因此，方程 **$Ax = y$** 无解。我们无法找到一组系数 **(x_1, x_2, x_3)** 使得 **$\alpha_1, \alpha_2, \alpha_3$** 的线性组合恰好等于 **y**。

既然无法找到精确解，我们的目标就变为寻找一个**最优近似解**。从几何上看，这个最优解就是让向量 **Ax** 成为 **y** 在 **A** 的列空间（那个平面）上的**正交投影** (Orthogonal Projection)。我们将这个投影向量表示为 **$\hat{y}$**。

为什么正交投影是最近的？因为点到平面的最短距离是垂线段。所以，误差向量 **$e = y - Ax$** 的长度 (即 **$||y - Ax||$**) 在 **Ax** 是 **y** 的正交投影时达到最小。这就是“最小二乘法”的几何意义。

3. 正交性条件

当 **Ax** 是 **y** 在由 **$\alpha_1, \alpha_2, \alpha_3$** 张成的平面上的正交投影时，误差向量 **$e = y - Ax$** 必须与这个平面**垂直**。

一个向量垂直于一个平面，意味着它垂直于这个平面上的**所有**向量。由于 **$\alpha_1, \alpha_2, \alpha_3$** 在这个平面内，所以误差向量 **$(y - Ax)$** 必然与 **$\alpha_1, \alpha_2, \alpha_3$** 都正交。

4. 转换为代数方程

两个向量正交，意味着它们的点积（内积）为零。因此，我们可以写出以下方程：

- **$\alpha_1 \cdot (y - Ax) = 0 \Rightarrow \alpha_1^T (y - Ax) = 0$**
- **$\alpha_2 \cdot (y - Ax) = 0 \Rightarrow \alpha_2^T (y - Ax) = 0$**
- **$\alpha_3 \cdot (y - Ax) = 0 \Rightarrow \alpha_3^T (y - Ax) = 0$**

这三个独立的方程可以被优雅地合并成一个矩阵方程。我们知道矩阵 **A** 的转置 **A^T** 是由 **$\alpha_1, \alpha_2, \alpha_3$** 的转置作为行向量构成的：
$A^T = [\ \alpha_1^T; \alpha_2^T; \alpha_3^T\ ]$

因此，将这三个方程组合起来，就得到了：
$A^T (y - Ax) = 0$

这个方程就是**正规方程**。

5. 求解 x

现在我们得到了一个可以求解的方程：
$A^T (y - Ax) = 0$
展开后得到：
$A^Ty - A^TAx = 0$
移项后：
$A^TAx = A^Ty$

如果矩阵 **A^TA** 是可逆的，我们就可以在等式两边同时左乘 **$(A^TA)^{-1}$** 来解出 **x**：
$x = (A^TA)^{-1} A^Ty$

至此，我们从几何投影的角度完整地推导出了正规方程的解。

关于 "左边同时乘以 A^T"

为什么很有意思的事情是，如果左边同时乘以 $\mathbf{A}^T$，等式自然成立？

这句话可能有一点误解。我们并不是从 $\mathbf{y} - \mathbf{A}\mathbf{x}$ 出发，然后随意地在左边乘以 $\mathbf{A}^T$。

正确地理解是：**乘以 $\mathbf{A}^T$ 是将几何上的“正交”条件转化为代数方程的核心步骤。**

- **几何意义**: 我们要找到的解 $\mathbf{x}$ 必须满足误差向量 $(\mathbf{y} - \mathbf{A}\mathbf{x})$ 与 $\mathbf{A}$ 的列空间正交。
- **代数翻译**: "与 $\mathbf{A}$ 的列空间正交" 这个条件，在代数上就等价于 "与 $\mathbf{A}$ 的所有列向量 $\boldsymbol{\alpha}_i$ 正交"。
- **矩阵运算**: 将误差向量 $(\mathbf{y} - \mathbf{A}\mathbf{x})$ 与**所有**列向量 $\boldsymbol{\alpha}_i$ 做点积并使其等于零，这个运算恰好可以用矩阵乘法 $\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}) = \mathbf{0}$ 来简洁地表示。

所以，左乘 $\mathbf{A}^T$ 并不是一个巧合的代数技巧，而是我们建立在几何直觉上的“正交性”要求的直接数学体现。 它是一个将几何问题转化为代数问题的桥梁。