5.3 Least Squares Problems

In this section, we are going to take a look at an important application of orthogonal projections to inconsistent systems of equations (no solution to the system).

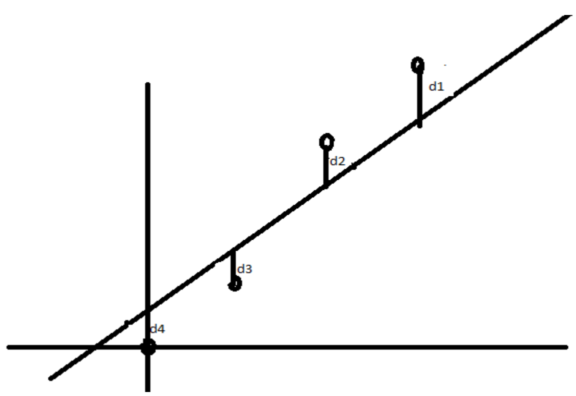
Exercise 1: Please find the equation of the line that passes through four points: $(0,0),(1,1),(2,3),(3,4)$ if possible?

Suppose the line equation is $y = kx + b$.

If we plug these points into the line, we have the following system of equations and matrix form of the system:

$$\begin{cases} b = 0 \\ k + b = 1 \\ 2k + b = 3 \\ 3k + b = 4 \end{cases} \Leftrightarrow \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} k \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 4 \end{pmatrix}$$

You can find that there is no solution to this system, so the system is not consistent. It's not possible to find a line to pass through the given four points, but we wish to find a line that will as closely as possible (in terms of $\sqrt{d_1^2 + d_2^2 + d_3^2 + d_4^2}$, see the following figure) approximate each of the points. More generally, given an inconsistent system of equations, $Ax = b$, can we find a vector, $\bar{x}$, so that $A\bar{x}$ will be as close to b as possible.



To answer this question, take a look at the general situation. So, let's suppose we have an inconsistent system of m equations and n unknowns, $Ax = b$, where A is an $m \times n$ matrix. Let's define

$$d = b - Ax$$

We call d the error vector, $\| d \| = \| b - Ax \|$ the error, since it measures the distance between Ax and b for any $x \in R^n$. We wish to find $\bar{x}$, such that $\| \bar{d} \| = \| b - A\bar{x} \|$ is the smallest possible error, or $\| \bar{d} \| = \| b - A\bar{x} \| = \{\min \| b - Ax \| \text{ for } x \in R^n\}$. The vector $\bar{x}$ is the **least squares solution** to inconsistent system: $Ax = b$.

Theorem 1: Let S be a subspace of R^m. For each $b \in R^m$, there is a unique element $proj_S b$ in S that is closest to b, that is,

$$\| b - s \| > \| b - proj_S b \|$$

for any $s \in S$ and $s \neq proj_S b$.

Proof. Since $R^m = S \oplus S^{\perp}$, each element $b \in R^m$ can be written uniquely as a sum

$$b = proj_S b + z$$

where $proj_S b \in S, z \in S^{\perp}$. For any $s \in S$ and $s \neq proj_S b$, we have $\| b - s \|^2 = \| (b - proj_S b) + (proj_S b - s) \|^2 = \| b - proj_S b \|^2 + \| proj_S b - s \|^2$ since $b - proj_S b = z \in S^{\perp}, proj_S b - s \in S$ by Pythagorean Law. Thus, $\| b - s \| > \| b - proj_S b \|$ since $\| proj_S b - s \|^2 > 0$.

Theorem 2: If A is an $m \times n$ matrix with linearly independent columns, then $A^T A$ is nonsingular.

Proof: Let $z = (z_1, z_2, \cdots, z_n)^T$ is a solution to $A^T Ax = 0$. So,

$$\left.\begin{array}{l} A^T Az = 0 \rightarrow Az \in N(A^T) \\ Az \in R(A) = N(A^T)^{\perp} \end{array}\right\} \rightarrow Az \in N(A^T) \bigcap N(A^T)^{\perp} \rightarrow Az = 0 \text{ since } S \bigcap S^{\perp} = 0.$$

$Az = 0 \Leftrightarrow a_1 z_1 + a_2 z_2 + \cdots + a_n z_n = 0 \rightarrow z_1 = z_2 = \cdots = z_n = 0 (z = 0)$ since the column vectors $a_1, a_2, \cdots, a_n$ of A are linearly independent . So $A^T Ax = 0$ has only a trivial solution $z = 0$.
Therefore $A^T A$ is nonsingular since $A^T Ax = 0$ has only a trivial solution $x = 0 \Leftrightarrow A^T A$ is nonsingular.

Theorem 3: Given the system of equations $Ax = b$, **a least square solution** to the system denoted by $\bar{x}$, will also be a solution to the associated normal system,

$$A^T Ax = A^T b$$

Moreover, if A has linearly independent columns, then there is a **unique least squares solution** given by,

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

Proof. Let $\bar{x}$ be a least squares solution, so $A\bar{x} = proj_S b$, where $S = R(A)$ is the column space of A.
From the proof of Theorem 1, we have $b - A\bar{x} = b - proj_S b \in S^{\perp} = R(A)^{\perp}$
$b - A\bar{x} = b - proj_S b \in N(A^T)$ (since $R(A)^{\perp} = N(A^T)$). Thus, we have

$$A^T (b - A\bar{x}) = 0 \rightarrow A^T A\bar{x} = A^T b.$$

If A has linearly independent columns, from Theorem 2, we have $A^T A$ is nonsingular. So we have $A^T Ax = A^T b$ has a unique solution: $x = (A^T A)^{-1} A^T b$.

Example 1: Find the least squares solution to the system

$$\begin{cases} x_1 + x_2 = 3 \\ -2x_1 + 3x_2 = 1 \\ 2x_1 - x_2 = 2 \end{cases}$$

Solution: The normal system for this system is

$$\begin{pmatrix} 1 & -2 & 2 \\ 1 & 3 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 2 \\ 1 & 3 & -1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} \to \begin{pmatrix} 9 & -7 \\ -7 & 11 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix} \to \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{83}{50} \\ \frac{71}{50} \end{pmatrix}$$

Example 2: Use least squares to find the equation of the line that will best approximate the four points: $(0,0),(1,1),(2,3),(3,4)$ (from Exercise 1).

$$\begin{cases} b = 0 \\ k + b = 1 \\ 2k + b = 3 \\ 3k + b = 4 \end{cases} \Leftrightarrow \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} k \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 4 \end{pmatrix}$$

Consider the corresponding normal system:

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} k \\ b \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 3 \\ 4 \end{pmatrix} \to \begin{pmatrix} 14 & 6 \\ 6 & 4 \end{pmatrix}\begin{pmatrix} k \\ b \end{pmatrix} = \begin{pmatrix} 19 \\ 8 \end{pmatrix} \to \begin{pmatrix} k \\ b \end{pmatrix} = \frac{1}{20}\begin{pmatrix} 4 & -6 \\ -6 & 14 \end{pmatrix}\begin{pmatrix} 19 \\ 8 \end{pmatrix} = \begin{pmatrix} \frac{6}{5} \\ -\frac{1}{10} \end{pmatrix}$$

Thus, the best fitting line is $y = \frac{6}{5}x - \frac{1}{10}$.

HW: 1,2,3,5,7