Section 3: Least Squares Modeling

Adhyyan Narang

April 8, 2022

Least Squares Review

Sometimes the system of equations

$$Ax = b$$

has zero solutions. It may be of interest in applications to approximate a solution, which leads to the least squares problem.

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2 \tag{1}$$

In lecture, we saw that the optimal solution to this problem is

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

This was derived by two methods: vector calculus and orthogonal projections.

Practice Problems

VMLS 12.7 Network Tomography

A network consists of n links, labeled $1, \ldots, n$. A path through the network is a subset of the links. (The order of the links on a path does not matter here.) Each link has a (positive) delay, which is the time it takes to traverse it. We let d denote the n-vector that gives the link delays. The total travel time of a path is the sum of the delays of the links on the path. Our goal is to estimate the link delays (i.e., the vector d), from a large number of (noisy) measurements of the travel times along different paths. This data is given to you as an $N \times n$ matrix P, where

$$P_{ij} = \begin{cases} 1 & \text{link } j \text{ is on path } i \\ 0 & \text{otherwise,} \end{cases}$$

and an N-vector t whose entries are the (noisy) travel times along the N paths. You can assume that $N > n$. You will choose your estimate d by minimizing the RMS deviation between the measured travel times (t) and the travel times predicted by the sum of the link delays. Explain how to do this, and give a matrix expression for $\hat{d}$. If your expression requires assumptions about the data P or t, state them explicitly.

VMLS 12.1 Approximating a vector as a multiple of another one

In the special case $n = 1$, the general least squares problem (12.1) reduces to finding a scalar x that minimizes $\|ax - b\|^2$, where a and b are m-vectors. (We write the matrix A here in lower case, since it is an m-vector.) Assuming a and b are nonzero, show that $\|a\hat{x} - b\|^2 = \|b\|^2(\sin\theta)^2$, where $\theta = \angle(a, b)$. This shows that the optimal relative error in approximating one vector by a multiple of another one depends on their angle.

Jupyter Notebook