CS 127/227A, Fall 2022

Optimization Models in Engineering

Gireeja Ranade, UC Berkeley
Zhiyu An

Contents

1. Linear Algebra

1.a. Least-Squares Problem Statement

Definition 1.1 (Least Squares)
Assume matrix A and vectors $\vec{x}$ and $\vec{b}$. The problem defined by

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|^2$$

is a Least Squares Problem (LSP).

Example 1.2
Assume we have two dimensional data set $\vec{x}$ and $\vec{y}$ and we want to formalize a LSP to find a linear correlation between x and y. We first formalize the goal linear correlation as

$$y = mx + c$$

where we want to find the optimal values for m and c to minimize the squared loss across all data points. Summarizing the above equation for all data points gives us

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Where

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_n & 1 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} m \\ c \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

And therefore

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|^2 = \min_{m,c} \sum_{i=1}^{n} (y_i - (mx_i + c))^2$$

Theorem 1.3 (Ordinary Least Squares)

Given the column space of the matrix A, for vector $\vec{b}$ not in the said column space, $A\vec{x} - \vec{b} = \vec{e}$ must be orthogonal to the columns of A. (Pythagora's theorem)

Therefore, the dot products of every column of A and $\vec{e}$ must be zero, i.e.

$$\begin{aligned} A^\top(A\vec{x} - \vec{b}) &= 0 \\ A^\top A\vec{x} - A^\top\vec{b} &= 0 \\ A^\top A\vec{x} &= A^\top\vec{b} \\ \vec{x} &= (A^\top A)^{-1}A^\top\vec{b} \end{aligned}$$

We conclude that the solution for Ordinary Least Squares (OLS) is

$$\vec{x}^* = \underset{\vec{x}}{\operatorname{argmin}} \|A\vec{x} - \vec{b}\|^2 = (A^\top A)^{-1}A^\top\vec{b}$$

1.b. Norm

Definition 1.4 (Norm)

A Norm is defined as

$$f := \mathbf{X} \to \mathbb{R}$$

For vector space $\mathbf{X}$.

The norm of x is denoted as $\|x\|$.

For any vector x and y, we have

- $\|x\| \geq 0$ and $\|x\| = 0$ iff $x = \vec{0}$
- $\|x + y\| \leq \|x\| + \|y\|$
- $\|\alpha x\| = |\alpha| * \|x\|$

Definition 1.5 (l-p Norm)

Generally, l-p norm is defined as

$$\|\vec{x}\|_p := \left(\sum |x_i|^p\right)^{\frac{1}{p}}; \quad 1 \leq p < \infty$$

Commonly used norms:

- $\|\vec{x}\|_1 := \sum |x_i|$
- $\|\vec{x}\|_2 := \sqrt{\sum |x_i|^2}$
- $\|\vec{x}\|_\infty := \max |x_i|$

Theorem 1.6 (Cauchy-Schwartz Inequality)

$$< \vec{x}, \vec{y} >= \vec{x}^\top \vec{y} = \|\vec{x}\|_2 \|\vec{y}\|_2 \cos\theta$$

Since $-1 \le \cos\theta \le 1$,

$$< \vec{x}, \vec{y} >= \vec{x}^\top \vec{y} \le \|\vec{x}\|_2 \|\vec{y}\|_2$$

Theorem 1.7 (Holder's Inequality)

For $p, q \ge 1$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$,

$$|\vec{x}^\top \vec{y}| \le \sum_{i=1}^{n} |x_i y_i| \le \|\vec{x}\|_p \|\vec{y}\|_p$$

i.e., Cauchy-Schwartz is a narrowed case of Holder's Inequality.

1.c. Gram-Schimdt

Theorem 1.8 (Gram-Schimdt/QR-decomposition)

Let X be a vector space with basis $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n\}$, which is orthonormal. For any matrix A,

$$A = QR$$

$$\left[\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n\right] = \left[\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n\right] \begin{bmatrix} \vec{r}_{11} & \vec{r}_{12} & \cdots & \vec{r}_{1n} \\ 0 & \vec{r}_{22} & \cdots & \vec{r}_{2n} \\ 0 & 0 & \ddots & \vec{r}_{3n} \\ 0 & 0 & 0 & \vec{r}_{nn} \end{bmatrix}$$

Where Q is orthonormal andR is upper-triangular.

Theorem 1.9 (Fundamental Theorem of Linear Algebra)

For matrix $A \in \mathbb{R}^{m*n}$,

$$Null(A) \bigoplus Range(A^\top) = \mathbb{R}^n$$

Where $\oplus$ denotes "direct sum" and $Range(A^\top)$ is the column space of $A^\top$. With the said equation we can also conclude that

$$Range(A) \bigoplus Null(A^\top) = \mathbb{R}^m$$

Theorem 1.10 (orthogonal decomposition theorem)
X a vector space and S a subspace of X. Then for any $\vec{x}$ in X,

$$\vec{x} = \vec{s} + \vec{r}, \quad \vec{s} \in S, \quad \vec{r} \in S^{\perp}$$

Such that

$$S^{\perp} = \{\vec{r} \mid < \vec{r}, \vec{s} > = 0, \quad \forall \vec{s} \in S\}$$

Therefore,

$$\mathbf{X} = S \bigoplus S^{\perp}$$

Example 1.11 (Minimum Norm Problem)
We want to find

$$\min \|\vec{x}\|_2^2$$

subject to $A\vec{x} = \vec{b}$. From FTLA we know that

$$\vec{x} = \vec{y} + \vec{z} \;\; s.t. \;\; \vec{y} \in N(A; \;\; \vec{z} \in R(A^{\top}).$$

And

$$A(\vec{y} + \vec{z}) = 0 + A\vec{z} = \vec{b}$$

Since $\vec{y} \perp \vec{z}$,

$$\|\vec{x}\|_2^2 = \|y\|_2^2 + \|z\|_2^2$$

Consider $\vec{z} = A^{\top}\vec{w}$,

$$\begin{aligned} A\vec{z} &= \vec{b} \\ AA^{\top}\vec{w} &= \vec{b} \\ \vec{w} &= (AA^{\top})^{-1}\vec{b} \end{aligned}$$

Therefore

$$\vec{z} = \min \|\vec{x}\|_2^2 = A^{\top}(AA^{\top})^{-1}\vec{b}$$

1.d. Symmetric Matrices

Definition 1.12
Matrix A is symmetric if $A = A^{\top}$, i.e. $A_{ij} = A_{ji}$.
Set $\mathbb{S}^n$ means the set of symmetric matrices of dimension n.