SI231b: Matrix Computations

Lecture 9: Least Squares and Orthogonal Projection

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology

ShanghaiTech University

Oct. 14, 2021

Brief Overview

Overdetermined System: $\mathrm{Ax} = \mathrm{b}$, $\mathrm{A} \in \mathbb{R}^{m \times n}$ $(m > n)$, the equation

- often has no solution, since

$$\mathrm{b} \in \mathbb{R}^m, \quad \text{while } \mathcal{R}(\mathrm{A}) \text{ is a subspace (at most of dimensional } n) \text{ of } \mathbb{R}^m$$

- has unique solution when

$$\mathrm{b} \in \mathcal{R}(\mathrm{A}) \text{ and } \mathrm{rank}(\mathrm{A}) = n$$

- has infinite solutions when

$$\mathrm{b} \in \mathcal{R}(\mathrm{A}) \text{ and } \mathrm{rank}(\mathrm{A}) < n$$

In practice, we need to find the full rank least square (LS) solution x_{LS},

$$\mathrm{x}_{LS} = \arg\min \|\mathrm{b} - \mathrm{Ax}\|_2^2,$$

where $\|\cdot\|_2$ represents the vector 2-norm and A is full rank.

Outline

- Motivation Applications
- Geometric Interpretation of Least Square
- Projection onto Subspaces
- Orthogonal Projection

In many applications, we can use the following representation

$$y = Ax,$$

or

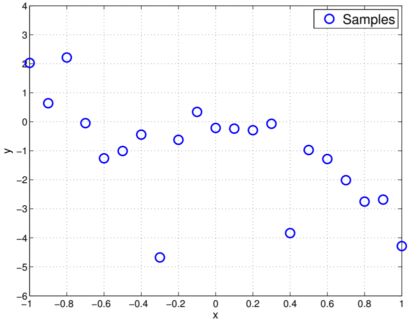
$$y = Ax + v,$$

where

- y is known (given data);
- A is given or stipulated;
- x is to be determined;
- v models the noise or error.

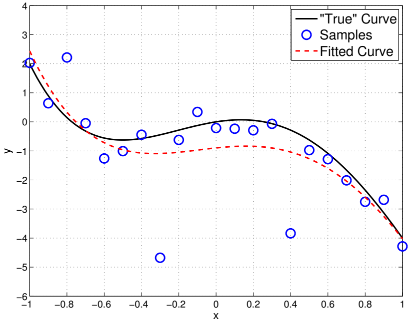
Data Fitting

Given a set of input-output data pairs $(x_i, y_i) \in \mathbb{R}^2$, $i = 1, \ldots, m$, find a function $f(x)$ that fits the data well.



Data Fitting Using Polynomials

Applying a polynomial model $f(x) = \sum_{i=0}^{p} a_i x^i$ and use LS



"True" curve: the true $f(x)$, $p = 5$.

Fitted curve: estimated $f(x)$, a obtained by LS with $p = 5$.

Autoregessive (AR) Model for Time Series

- model current output y_t as being related to its past values in a linear manner

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_q y_{t-q} + v_t, \quad t = 0, 1, \ldots$$

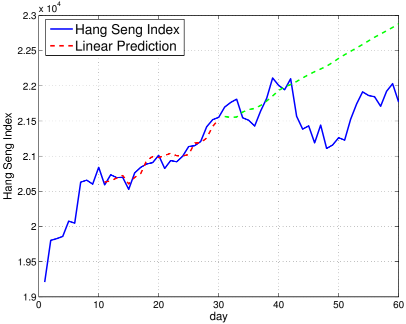
for some coefficient $a \in \mathbb{R}^q$ and some positive integer q (correlation length).

- using AR model predictions can also be made

$$\hat{y}_{t+d} = a_1 \hat{y}_{t+d-1} + a_2 \hat{y}_{t+d-2} + \ldots + a_q \hat{y}_{t+d-q}, \quad d = 1, 2, \ldots$$

where we denote $\hat{y}_{t-i} = y_{t-i}$ for $i = 1, \ldots, q$.

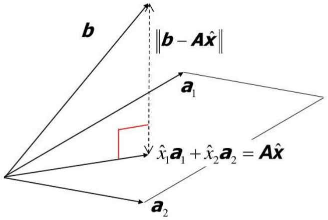
AR Model for Stock Market



- blue: Hang Seng Index during a certain time period.
- red: training phase, $\hat{y}_t = \sum_{i=1}^{q} a_i y_{t-i}$, a is obtained by LS, and $q = 10$.
- green: prediction phase, $\hat{y}_{t+d} = \sum_{i=1}^{q} a_i \hat{y}_{t+d-i}$.

$$x_{LS} = \arg\min_{x \in \mathbb{R}^n} \|b - Ax\|_2^2$$

1. find $\tilde{b} \in \mathcal{R}(A)$ such that $\|b - \tilde{b}\|_2$ is minimized
   - recall the distance between two vectors using vector norms
2. solve $Ax_{LS} = \tilde{b}$ to obtain x_{LS}



Question: how to obtain $\tilde{b} \in \mathcal{R}(A)$?

Projectors

A projector is a square matrix that satisfies

$$\mathrm{P}^2 = \mathrm{P}.$$

- such a matrix is called idempotent
- geometric interpretation?

Note: this definition of projectors include both

- orthogonal projectors (key in our course)
- oblique projectors (will not be addressed)

Question: onto which subspace does P project?

Answer: $\mathcal{R}(\mathrm{P})$

How to distinguish orthogonal and oblique projection?

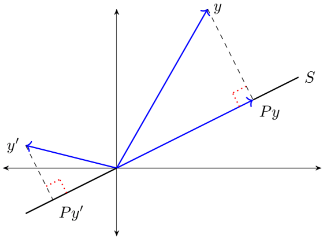


Figure 1: orthogonal projection

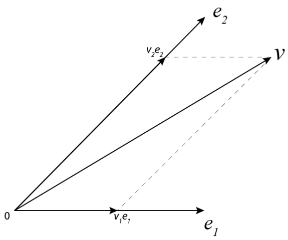


Figure 2: oblique projection

Answer: the projection direction $Pv - v$

Note: $P(Pv - v) = 0$, which means $(Pv - v) \in \mathcal{N}(P)$.

If P is a projector, then $\mathsf{I}-\mathsf{P}$ is also a projector (why?)

$$(\mathsf{I}-\mathsf{P})^2=\mathsf{I}-2\mathsf{P}+\mathsf{P}^2=(\mathsf{I}-\mathsf{P})$$

The projector $\mathsf{I}-\mathsf{P}$ is called the complementary projector of P.

Question: onto which subspace does $\mathsf{I}-\mathsf{P}$ project?

Answer: $\mathcal{N}(\mathsf{P})=\mathcal{R}(\mathsf{I}-\mathsf{P})$

First, $\mathcal{N}(\mathsf{P})\subset\mathcal{R}(\mathsf{I}-\mathsf{P})$ (give your explanation here)

Second, $\mathcal{R}(\mathsf{I}-\mathsf{P})\subset\mathcal{N}(\mathsf{P})$ (you are supposed to work it out indepently)

Then,

$$\mathcal{R}(\mathsf{I}-\mathsf{P})=\mathcal{N}(\mathsf{P}) \text{ and } \mathcal{R}(\mathsf{P})=\mathcal{N}(\mathsf{I}-\mathsf{P})$$

$$\mathcal{R}(\mathsf{P})\cap\mathcal{N}(\mathsf{P})=\{0\}$$

Projection onto Subspaces

Suppose $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$, then there is a projector P such that $\mathcal{R}(\mathrm{P}) = \mathcal{U}$ and $\mathcal{N}(\mathrm{P}) = \mathcal{W}$, we say that P is a projector onto $\mathcal{U}$ along $\mathcal{W}$.

Previous analysis show that the projector $\mathrm{P} \in \mathbb{R}^{m\times m}$ separates $\mathbb{R}^m$ into two subspaces

- $\mathcal{R}(\mathrm{P})$
- $\mathcal{N}(\mathrm{P})$

and

$$\mathbb{R}^m = \mathcal{R}(\mathrm{P}) \oplus \mathcal{N}(\mathrm{P}) \quad \text{can you prove this?}$$

P projects $\mathbb{R}^m$ onto $\mathcal{R}(\mathrm{P})$ along $\mathcal{N}(\mathrm{P})$.

Orthogonal projector

An orthogonal projector P is the one that projects onto a subspace $\mathcal{U}$ along a subspace $\mathcal{W}$ when $\mathcal{U}$ and $\mathcal{W}$ are orthogonal.

Warning: orthogonal projectors are not orthogonal matrices.

Theorem

A projector P is orthogonal if and only if $P = P^T$.

Proof ?