

SI231b: Matrix Computations

Lecture 7: Least Squares and Orthogonal Projection

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology
ShanghaiTech University

Sept. 29, 2020

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

- ▶ often has no solution, since

$$b \in \mathbb{R}^m, \quad \text{while } \mathcal{R}(A) \subset \mathbb{R}^n$$

- ▶ has unique solution when

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

- ▶ has infinite solutions when

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

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

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

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

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

In many applications, we can use the following representation

$$y = Ax,$$

or

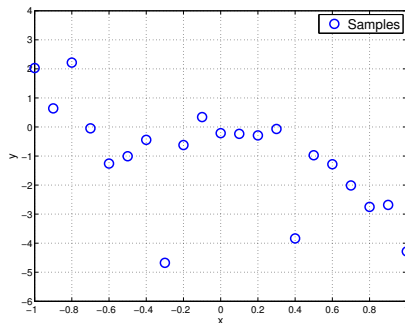
$$y = Ax + v,$$

where

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

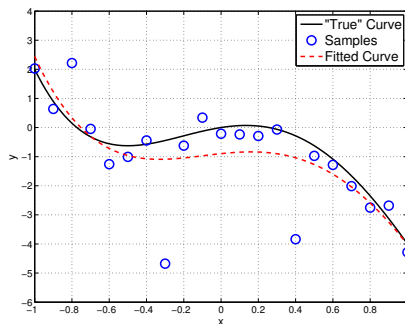
Data Fitting

Given a set of input-output data pairs $(x_i, y_i) \in \mathbb{R}^2$, $i = 1, \dots, m$, 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 $p = 5$.

Autoregressive (AR) Model for Time Series

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

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

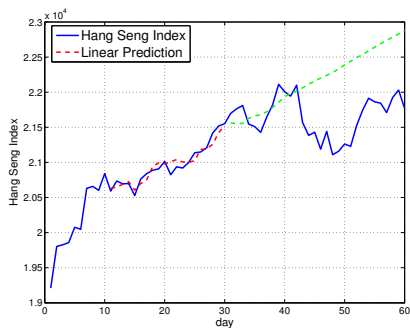
for some coefficient $\mathbf{a} \in \mathbb{R}^q$ and some positive integer q (correlation length).

- ▶ using AR model predictions can also be made

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

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

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, $\hat{y}_{t+d} = \sum_{i=1}^q a_i \hat{y}_{t+d-i}$.

Polynomial Model for Time Series

$$y_t = a_0 + a_1 t + a_2 t^2 + \dots + a_p t^p + v_t, \quad t = 0, 1, \dots,$$

- ▶ once a computed, predictions of time series can be made
- ▶ to compute a, we make use of

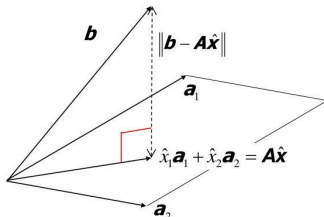
$$\underbrace{\begin{bmatrix} y_0 \\ \vdots \\ y_t \\ \vdots \\ y_{T-1} \end{bmatrix}}_y = \underbrace{\begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & t & \dots & t^p \\ \vdots & \vdots & & \vdots \\ 1 & T-1 & \dots & (T-1)^p \end{bmatrix}}_A \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix}}_x + \underbrace{\begin{bmatrix} v_0 \\ \vdots \\ v_t \\ \vdots \\ v_{T-1} \end{bmatrix}}_v$$

- ▶ A is Vandermonde with distinct roots
- ▶ thus A has full rank (recall the determinant of a Vandermonde matrix)

Geometric Interpretation of Least Square

$$\mathbf{x}_{LS} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2$$

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



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

Projectors

A projector is a **square matrix** that satisfies

$$P^2 = P.$$

- ▶ such a matrix is call 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: $\mathcal{R}(P)$

How to distinguish orthogonal and oblique projection?

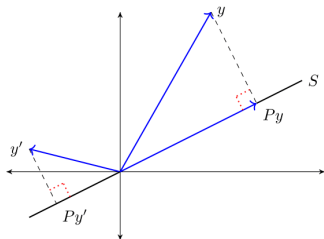


Figure 1: orthogonal projection

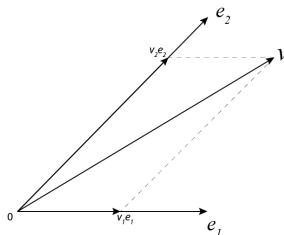


Figure 2: oblique projection

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 $I - P$ is also a projector (why?)

$$(I - P)^2 = I - 2P + P^2 = (I - P)$$

The projector $I - P$ is called the **complementary projector** of P .

Question: onto which subspace does $I - P$ project?

Answer: $\mathcal{N}(P)$

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

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

Then,

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

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

Direct Sum

Let \mathcal{U} , \mathcal{W} be subspaces of \mathcal{V} . Then \mathcal{V} is said to be the direct sum of \mathcal{U} and \mathcal{W} , and we write $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$, if $\mathcal{V} = \mathcal{U} + \mathcal{W}$ and $\mathcal{U} \cap \mathcal{W} = \{0\}$.

Projection onto subspaces

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

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.

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

► $\mathcal{R}(P)$

► $\mathcal{N}(P)$

and P projects onto $\mathcal{R}(P)$ along $\mathcal{N}(P)$.

Theorem

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