THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics

Exercises on Orthogonal Projection and Least Squares Problem

The least-squares method is the most intuitive way to obtain the best fit curve representing the data patterns.

Let us find out how we can get the best possible solution using the method of least squares (curve fitting).

Least Squares Problem

Consider 4 points in the xu-plane (2-dimensional data):

$$(0,1),\ (1,3),\ (2,4),\ (3,4)$$

We can see the left graph with x and u in the below figure. We can make the relationship between x and u by drawing a straight line. It is quite intuitive.

The least-squares method can be used to find out a straight line that best represents the given data. The straight line obtained is called *the least-squares line* or *the best fit line.*

Data Summary Table

x	u
0	1
1	3
2	4
3	4

Let's consider the simplest case of solving the best possible straight line $u = a + bx$ to describe the given data (x_i, u_i). It is finding the linear function that best fits (x_i, u_i). The ideal situation is to find the u-intercept a and slope b that satisfies $u_i = a + bx_i$ for all data (x_i, u_i).

We can find out the solution using a system of linear equations in matrix form with two unknowns a and b as follows.

Formulating the least squares problem

Data (x,u)	Linear function $u = a + bx$	System of linear equations	Matrix Form
$(0,1)$	$1 = a + b \cdot 0$		
$(1,3)$	$3 = a + b \cdot 1$	$\begin{cases} a = 1 \\ a + b = 3 \\ a + 2b = 4 \\ a + 3b = 4 \end{cases}$	$A\mathbf{u} = \mathbf{y}$
$(2,4)$	$4 = a + b \cdot 2$		where $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix}$
$(3,4)$	$4 = a + b \cdot 3$		

When we have only two data points, then it will be easy to find a and b for $u = a + bx$. If we have more than two data points, this modeling requires us to use a linear system of equations. We usually use a larger amount of data to find the best possible curve fitting.

As a result, we end up with a larger number of equations than the number of unknowns. In this case, we do not expect $A\mathbf{u} = \mathbf{y}$ to have a (unique) solution $\mathbf{u}$.

So, we try to find an approximate $\hat{\mathbf{u}}$ that minimizes the distance between $A\mathbf{u}$ and $\mathbf{y}$,

$$\min_{\mathbf{u}} \|A\mathbf{u} - \mathbf{y}\|$$

We will try to obtain a straight line with the least error even though it does not pass all four points. Mathematically, it means $\|A\mathbf{u} - \mathbf{y}\| = 0$.

This problem is called *least squares problem* and $\hat{\mathbf{u}}$ is called the optimal solution (or the least square solution). Although $\hat{\mathbf{u}} \approx \mathbf{y}$, even though $\hat{\mathbf{u}}$ might not satisfy $A\mathbf{u} = \mathbf{y}$.

Meaning of the least-squares problem

Let $\hat{y}_i$ be the value obtained by inputting x_i into $y = a + bx$ from each data point (x_i, y_i). There exists an error for some i when y_i and $\hat{y}_i$ are not the same.

If y_i and $\hat{y}_i$ are the same for all i, then the line $y = a + bx$ has a unique solution. Since there are cases where y_i and $\hat{y}_i$ are not the same, in other words, $(y_i - \hat{y}_i)^2$ is not all zero, we will try to find a and b that minimize the error. Adding all the squared errors for all the given data gives the following error function $E(\mathbf{u})$.

$$E(\mathbf{u}) = E(a, b) = (a - 1)^2 + (a + b - 3)^2 + (a + 2b - 4)^2 + (a + 3b - 4)^2 = \|A\mathbf{u} - \mathbf{y}\|^2$$

or

$$\min E(\mathbf{u})$$

The error $E(\mathbf{u})$ is eventually equal to the square of the distance between $A\mathbf{u}$ and $\mathbf{y}$. It is easy to see how a norm and an inner product are related to errors.

Solving the least-squares problem is solving a problem of finding $\hat{\mathbf{u}}$ that minimizes the error function $\min E(\mathbf{u})$. The optimal solution to this problem is the least-squares solution $\hat{\mathbf{u}}$. To find the least-squares solution, we need the concept of *projection*.

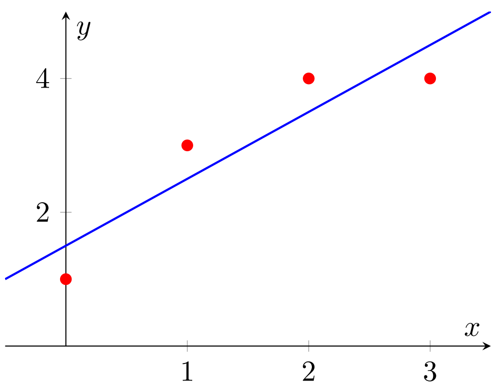

Figure 1: Least-squares line fitting the data

Component	Definition
Data	$(0,1),\ (1,3),\ (2,4),\ (3,4)$
Model	$y = a + bx$
Predicted value	$\hat{y}_i = a + bx_i$
Error term	$(y_i - \hat{y}_i)^2$
Error function	$E(a,b) = \sum_{i=1}^{4}(y_i - (a + bx_i))^2$
Matrix form	$E(\mathbf{u}) = \|A\mathbf{u} - \mathbf{y}\|^2$, where $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix}$
Least-squares problem	$\min_{\mathbf{u}} E(\mathbf{u})$
Optimal solution	Least-squares solution $\hat{\mathbf{u}}$

Projection and least-squares solution

To understand the least-squares problem, we need to know about a projection. Let's consider the problem of finding t satisfying the following:

$$\min \|t\mathbf{a} - \mathbf{x}\|$$

where t is a real number.

In other words, this is the problem of finding t that minimizes the distance between the vector $\mathbf{x}$ and the straight line containing $\mathbf{a}$, where t is the scalar.

We can see that $\|t\mathbf{a} - \mathbf{x}\|$ represents the distance between $t\mathbf{a}$ and $\mathbf{x}$ as shown in Figure 2. Intuitively, the shortest distance can be obtained when $\mathbf{p} = t\mathbf{a}$ and $(\mathbf{x} - t\mathbf{a}) \perp \mathbf{a}$. Such t is a solution for $\min \|t\mathbf{a} - \mathbf{x}\|$, and the vector $\mathbf{p} = t\mathbf{a}$ is called the projection of $\mathbf{x}$ onto $\mathbf{a}$.

Since $(\mathbf{x} - t\mathbf{a}) \perp \mathbf{a}$, the scalar t can be obtained as follows:

$$\mathbf{a} \cdot (\mathbf{x} - t\mathbf{a}) = 0 \implies t(\mathbf{a} \cdot \mathbf{a}) = \mathbf{a} \cdot \mathbf{x} \implies t = \frac{\mathbf{a} \cdot \mathbf{x}}{\mathbf{a} \cdot \mathbf{a}} = (\mathbf{a}^T\mathbf{a})^{-1}\mathbf{a}^T\mathbf{x}$$

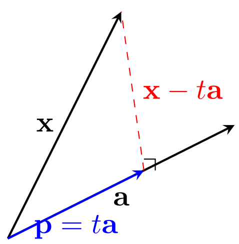

Figure 2: Vector projection of $\mathbf{x}$ onto $\mathbf{a}$

Example 1

Find the projection of $\mathbf{y}$ onto $\mathbf{x}$, for $\mathbf{x} = (2, -1, 3)$, $\mathbf{y} = (4, -1, 2)$.

Summary Table

Term	Expression
Projection Scalar	$t = \dfrac{\mathbf{a} \cdot \mathbf{x}}{\mathbf{a} \cdot \mathbf{a}}$
Projection Vector	$\mathbf{p} = t\mathbf{a}$
Orthogonality	$\mathbf{a} \cdot (\mathbf{x} - t\mathbf{a}) = 0$
Matrix Form	$t = (\mathbf{a}^T\mathbf{a})^{-1}\mathbf{a}^T\mathbf{x}$
Least Squares Objective	$\min \lVert A\mathbf{u} - \mathbf{y} \rVert$
Column Space Combination	$A\mathbf{u} = aA_1 + bA_2$

We can solve the least-squares problem $\min \lVert A\mathbf{u} - \mathbf{y} \rVert$ similar to the problem of determining t that minimizes the difference between the projection (of $\mathbf{x}$ onto $\mathbf{a}$) and the vector $\mathbf{x}$. For this purpose, let A_1, A_2 be the first and second column vector of A, respectively. Then, we have $A\mathbf{u} = aA_1 + bA_2$.

$$\min \lVert A\mathbf{u} - \mathbf{y} \rVert = \min \lVert (aA_1 + bA_2) - \mathbf{y} \rVert$$

Hence, the least-squares problem is related to the column space of column vectors of A. The least-square problem can be interpreted as a problem of finding a projection.

The plane $A\mathbf{u} = aA_1 + bA_2$ is a set of images given by $A\mathbf{x}$. As shown in the figure, it can be understood as the problem of obtaining a and b that gives the minimal distance between the plane containing A_1 and A_2 and the vector $\mathbf{y}$.

As shown in Figure 3, $\lVert (aA_1 + bA_2) - \mathbf{y} \rVert$ is the length of the marked real line for the values of a and b. It is easy to see that the vector with the shortest distance from $\mathbf{y}$ to $aA_1 + bA_2$ can be obtained by $\mathbf{p} = aA_1 + bA_2$ when $(A\mathbf{u} - \mathbf{y}) \perp A_1$ and $(A\mathbf{u} - \mathbf{y}) \perp A_2$.

Such a and b give the solution for $\min \lVert A\mathbf{u} - \mathbf{y} \rVert = \min \lVert (aA_1 + bA_2) - \mathbf{y} \rVert$. The conditions $(A\mathbf{u} - \mathbf{y}) \perp A_1$ and $(A\mathbf{u} - \mathbf{y}) \perp A_2$ imply that $A^T(A\mathbf{u} - \mathbf{y}) = \mathbf{0}$ and give $\hat{\mathbf{u}} = (A^TA)^{-1}A^T\mathbf{y}$.

$$\begin{cases} A_1 \cdot (A\mathbf{u} - \mathbf{y}) = 0 \\ A_2 \cdot (A\mathbf{u} - \mathbf{y}) = 0 \end{cases} \iff \begin{bmatrix} A_1 \cdot (A\mathbf{u} - \mathbf{y}) \\ A_2 \cdot (A\mathbf{u} - \mathbf{y}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\iff \begin{bmatrix} A_1^T \\ A_2^T \end{bmatrix} (A\mathbf{u} - \mathbf{y}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \iff A^T(A\mathbf{u} - \mathbf{y}) = \mathbf{0}$$

$$\iff A^TA\mathbf{u} = A^T\mathbf{y} \iff \hat{\mathbf{u}} = (A^TA)^{-1}A^T\mathbf{y}$$

Note

If $x_i \neq x_j$ for a data set $\{(x_i, y_i)\}$, then A^TA is always invertible (Vandermonde determinant) and $\hat{\mathbf{u}} = (A^TA)^{-1}A^T\mathbf{y}$.

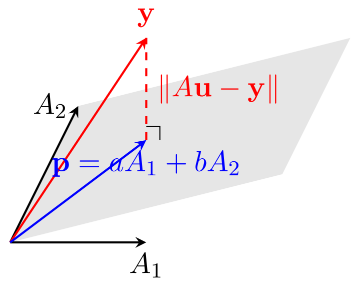

Figure 3: Projection of $\mathbf{y}$ onto the column space of A

Summary of Projection and Least-Squares Concepts

Concept	Definition/Formula
Line projection problem	$\min \lVert t\mathbf{a} - \mathbf{x} \rVert$
Optimal scalar t	$t = \frac{\mathbf{a}\cdot\mathbf{x}}{\mathbf{a}\cdot\mathbf{a}} = (\mathbf{a}^T\mathbf{a})^{-1}\mathbf{a}^T\mathbf{x}$
Least-squares problem	$\min \lVert A\mathbf{u} - \mathbf{y} \rVert = \min \lVert aA_1 + bA_2 - \mathbf{y} \rVert$
Orthogonality conditions	$(A\mathbf{u} - \mathbf{y}) \perp A_1,\ (A\mathbf{u} - \mathbf{y}) \perp A_2$
Normal equations	$A^T(A\mathbf{u} - \mathbf{y}) = \mathbf{0} \implies A^TA\mathbf{u} = A^T\mathbf{y}$
Least-squares solution	$\hat{\mathbf{u}} = (A^TA)^{-1}A^T\mathbf{y}$

Example 2: Finding a least-squares line for data

Find a least-squares solution of $A\mathbf{u} = \mathbf{y}$ from the given data.

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix}$$

Solution

The least-squares solution of $A\mathbf{u} = \mathbf{y}$ is given by

$$\hat{\mathbf{u}} = (A^TA)^{-1}A^T\mathbf{y}$$

which minimizes $\lVert A\mathbf{u} - \mathbf{y} \rVert$.

Calculating:

$$\hat{\mathbf{u}} = \begin{bmatrix} a \\ b \end{bmatrix} = (A^TA)^{-1}A^T\mathbf{y} = \begin{bmatrix} \frac{7}{10} & -\frac{3}{10} \\ -\frac{3}{10} & \frac{2}{10} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ 1 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1 \end{bmatrix}$$

The least-squares line is

$$y = a + bx = \frac{3}{2} + x$$

where

$$\hat{\mathbf{u}} = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1 \end{bmatrix}.$$

We will study the process of finding a least-squares curve (line), called *linear regression.*

Example 3: Finding a suitable curve for data (Curve Fitting)

Consider 4 points in the xu-plane (2-dimensional data):

$$(0,1),\ (1,3),\ (2,4),\ (3,4)$$

Let us find the best fit curve (a quadratic approximation of $u = a + bx + cx^2$) to describe (x_i, u_i). Since all 4 points of data should satisfy the quadratic equation $u = a + bx + cx^2$, they can be expressed in matrix form as follows.

Formulating the quadratic least-squares problem

Data (x, u)	Quadratic function $u = a + bx + cx^2$	System of linear equations	Matrix representation
$(0, 1)$	$1 = a + b \cdot 0 + c \cdot 0^2$	$\begin{cases} a = 1 \\ a + b + c = 3 \\ a + 2b + 4c = 4 \\ a + 3b + 9c = 4 \end{cases}$	$A\mathbf{u} = \mathbf{y}$ where $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix}$
$(1, 3)$	$3 = a + b \cdot 1 + c \cdot 1^2$		
$(2, 4)$	$4 = a + b \cdot 2 + c \cdot 2^2$		
$(3, 4)$	$4 = a + b \cdot 3 + c \cdot 3^2$		

Simply, we only need to find a vector $\hat{\mathbf{u}} = (a, b, c)$ for $u = a + bx + cx^2$ that minimizes $\|A\mathbf{u} - \mathbf{y}\|$. To obtain this quadratic function for a data set, we will have a system of equations with no unique solution. So we will have a least-squares problem $\min \|A\mathbf{u} - \mathbf{y}\|$ to solve.

The following function $E(\mathbf{u})$ represents an error for this problem:

$$E(\mathbf{u}) = E(a, b, c) = (a-1)^2 + (a+b+c-3)^2 + (a+2b+4c-4)^2 + (a+3b+9c-4)^2 = \|A\mathbf{u} - \mathbf{y}\|^2$$

We can find

$$\hat{\mathbf{u}} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = (A^T A)^{-1} A^T \mathbf{y} = \begin{bmatrix} 1 \\ \frac{3}{2} \\ -\frac{1}{2} \end{bmatrix}$$

So the least-squares curve is

$$u = 1 + \frac{3}{2}x - \frac{1}{2}x^2.$$

Component	**Value**
Data Points	$(0, 1)$, $(1, 3)$, $(2, 4)$, $(3, 4)$
Quadratic Model	$u = a + bx + cx^2$
Design Matrix A	$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$
Least-Squares Solution	$\hat{\mathbf{u}} = \begin{bmatrix} 1 \\ 1.5 \\ -0.5 \end{bmatrix}$
Fitted Quadratic Curve	$u = 1 + \frac{3}{2}x - \frac{1}{2}x^2$