香港中文大學
數學系
正交投影與最小平方問題練習

最小平方法（Least-squares method）是獲取代表數據模式的最佳擬合曲線最直觀的方法。讓我們來看看如何使用最小平方法（曲線擬合）獲得最佳的可能解。

最小平方問題

考慮xu-平面上的4 個點（二維數據）：

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

我們可以在下圖中看到帶有x 和u 的圖。我們可以透過畫一條直線來建立x 和u 之間的關係。這非常直觀。最小平方法可用於找出一條最能代表給定數據的直線。所獲得的直線稱為最小平方線（least-squares line）或最佳擬合線（best fit line）。

數據摘要表

x	u
0	1
1	3
2	4
3	4

讓我們考慮最簡單的情況，即求解最佳可能的直線$u = a + bx$ 來描述給定的數據(x_i, u_i)。也就是尋找最適合(x_i, u_i) 的線性函數。理想情況是找出能使所有數據(x_i, u_i) 滿足$u_i = a + bx_i$ 的u-截距a 和斜率b。如下所示，我們可以使用包含兩個未知數a 和b 的矩陣形式線性方程組來找出解。

構建最小平方問題

數據 (x,u)	線性函數$u = a + bx$	線性方程組	矩陣形式
$(0,1)$	$1 = a + b\cdot 0$		
$(1,3)$	$3 = a + b\cdot 1$	$\begin{cases} a = 1 \\ a + b = 3 \\ a + 2b = 4 \\ a + 3b = 4 \end{cases}$	$A\mathbf{u} = \mathbf{y}$
$(2,4)$	$4 = a + b\cdot 2$		其中 $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix}$
$(3,4)$	$4 = a + b\cdot 3$		

當我們只有兩個數據點時，很容易就能找出$u = a + bx$ 中的a 和b。如果我們有多於兩個數據點，這種建模需要我們使用線性方程組。我們通常會使用大量數據來尋找最佳的曲線擬合。結果，我們最終得到的方程式數量會多於未知數的數量。在這種情況下，我們預期$A\mathbf{u} = \mathbf{y}$ 不會有（唯一的）解$\mathbf{u}$。因此，我們嘗試找出一個近似值$\hat{\mathbf{u}}$，使得$A\mathbf{u}$ 與$\mathbf{y}$ 之間的距離最小化，

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

即使直線未能穿過所有四個點，我們仍會嘗試得出誤差最小的直線。在數學上，這意味著$\|A\mathbf{u} - \mathbf{y}\| = 0$。這個問題被稱為最小平方問題，而$\hat{\mathbf{u}}$ 則被稱為最佳解（或最小平方解）。儘管$\hat{\mathbf{u}} \approx \mathbf{y}$，即使$\hat{\mathbf{u}}$ 可能不滿足$A\mathbf{u} = \mathbf{y}$。

最小平方問題的含義

設$\hat{y}_i$ 為將各個數據點(x_i, y_i) 中的x_i 代入$y = a + bx$ 所得出的值。當y_i 和$\hat{y}_i$ 不相同時，對應的i 就會存在誤差。如果所有的i 對應的y_i 和$\hat{y}_i$ 都相同，那麼直線$y = a + bx$ 就有唯一解。由於存在y_i 和$\hat{y}_i$ 不相同的情況，換言之，即$(y_i - \hat{y}_i)^2$ 並非全部為零，我們將嘗試找出能使誤差最小化的a 和b。將所有給定數據的誤差平方相加，可得出以下的誤差函數$E(\mathbf{u})$。

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

或

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

誤差$E(\mathbf{u})$ 最終等於$A\mathbf{u}$ 和$\mathbf{y}$ 之間距離的平方。我們很容易可以看出範數（norm）與內積（inner product）是如何與誤差產生關聯的。解決最小平方問題，即是解決尋找能使誤差函數最小化$\min E(\mathbf{u})$ 的$\hat{\mathbf{u}}$ 的問題。這個問題的最佳解就是最小平方解$\hat{\mathbf{u}}$。要找出最小平方解，我們需要用到投影（projection）的概念。

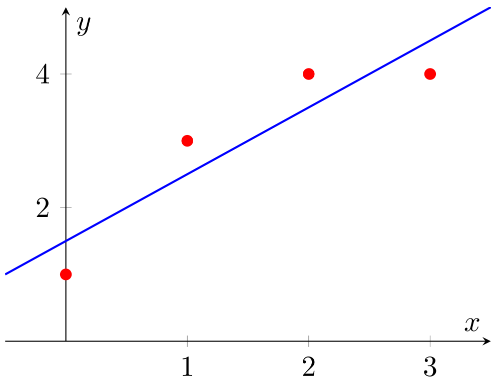

Figure 1: 擬合數據的最小平方線

組件	定義
數據	$(0,1),\ (1,3),\ (2,4),\ (3,4)$
模型	$y = a + bx$
預測值	$\hat{y}_i = a + bx_i$
誤差項	$(y_i - \hat{y}_i)^2$
誤差函數	$E(a,b) = \sum_{i=1}^{4}(y_i - (a + bx_i))^2$
矩陣形式	$E(\mathbf{u}) = \|A\mathbf{u} - \mathbf{y}\|^2$, 其中 $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix}$
最小平方問題	$\min_{\mathbf{u}} E(\mathbf{u})$
最佳解	最小平方解 $\hat{\mathbf{u}}$

投影與最小平方解

要了解最小平方問題，我們需要認識投影。讓我們考慮尋找滿足以下條件的t 的問題：

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

其中t 是實數。

換句話說，這是在尋找一個能夠最小化向量$\mathbf{x}$ 與包含$\mathbf{a}$ 的直線之間距離的純量t 的問題。我們可以看見$\|t\mathbf{a}-\mathbf{x}\|$ 代表$t\mathbf{a}$ 和$\mathbf{x}$ 之間的距離，如圖2 所示。直觀上，當$\mathbf{p}=t\mathbf{a}$ 且$(\mathbf{x}-t\mathbf{a})\perp\mathbf{a}$ 時，可以獲得最短距離。這樣的t 便是$\min\|t\mathbf{a}-\mathbf{x}\|$ 的解，而向量$\mathbf{p}=t\mathbf{a}$ 則稱為$\mathbf{x}$ 在$\mathbf{a}$ 上的投影。由於$(\mathbf{x}-t\mathbf{a})\perp\mathbf{a}$，可以透過以下方式求得純量$t$：

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

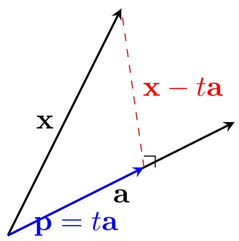

Figure 2: $\mathbf{x}$ 在$\mathbf{a}$ 上的向量投影

例子1

已知$\mathbf{x}=(2,-1,3)$ 及$\mathbf{y}=(4,-1,2)$，求$\mathbf{y}$ 在$\mathbf{x}$ 上的投影。

摘要表

術語	表達式
投影純量	$t=\frac{\mathbf{a}\cdot\mathbf{x}}{\mathbf{a}\cdot\mathbf{a}}$
投影向量	$\mathbf{p}=t\mathbf{a}$
正交性	$\mathbf{a}\cdot(\mathbf{x}-t\mathbf{a})=0$
矩陣形式	$t=(\mathbf{a}^T\mathbf{a})^{-1}\mathbf{a}^T\mathbf{x}$
最小平方目標	$\min\|A\mathbf{u}-\mathbf{y}\|$
行空間組合	$A\mathbf{u}=aA_1+bA_2$

我們可以用類似解決尋找能夠最小化投影（$\mathbf{x}$ 在$\mathbf{a}$ 上）與向量$\mathbf{x}$ 之間差異的t 的問題一樣，來解決最小平方問題$\min\|A\mathbf{u}-\mathbf{y}\|$。為此，我們設$A_1,A_2$ 分別為矩陣A 的第一和第二個行向量。那麼，我們得出$A\mathbf{u}=aA_1+bA_2$。

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

因此，最小平方問題與矩陣A 的行向量的行空間有關。最小平方問題可以被理解為一個尋求投影的問題。平面$A\mathbf{u}=aA_1+bA_2$ 是由$A\mathbf{x}$ 給出的一組圖像。如圖所示，它可以被理解為獲取a 和b 的問題，從而得出包含A_1 和A_2 的平面與向量$\mathbf{y}$ 之間的最小距離。如圖3 所示，$\|(aA_1+bA_2)-\mathbf{y}\|$ 是標記實線的長度（對於a 和b 的值）。我們很容易可以看出，當$(A\mathbf{u}-\mathbf{y})\perp A_1$ 和$(A\mathbf{u}-\mathbf{y})\perp A_2$ 時，可以透過$\mathbf{p}=aA_1+bA_2$ 獲得從$\mathbf{y}$ 到aA_1+bA_2 距離最短的向量。這樣的a 和b 提供了$\min\|A\mathbf{u}-\mathbf{y}\|=\min\|(aA_1+bA_2)-\mathbf{y}\|$ 的解。條件$(A\mathbf{u}-\mathbf{y})\perp A_1$ 和$(A\mathbf{u}-\mathbf{y})\perp A_2$ 意味著$A^T(A\mathbf{u}-\mathbf{y})=\mathbf{0}$，並得出$\hat{\mathbf{u}}=(A^TA)^{-1}A^T\mathbf{y}$。

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

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

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

注意

對於數據集$\{(x_i, y_i)\}$，如果$x_i \neq x_j$，那麼$A^T A$ 總是可逆的（范德蒙行列式Vandermonde determinant），且$\hat{\mathbf{u}} = (A^T A)^{-1} A^T\mathbf{y}$。

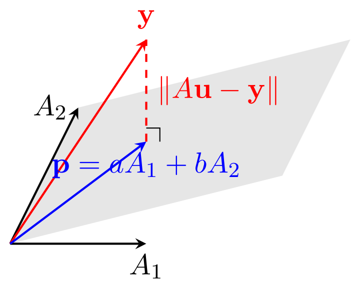

Figure 3: $\mathbf{y}$ 在矩陣A 的行空間上的投影

投影與最小平方概念摘要

概念	定義/ 公式
直線投影問題	$\min \|t\mathbf{a} - \mathbf{x}\|$
最佳純量 t	$t = \frac{\mathbf{a}\cdot\mathbf{x}}{\mathbf{a}\cdot\mathbf{a}} = (\mathbf{a}^T\mathbf{a})^{-1}\mathbf{a}^T\mathbf{x}$
最小平方問題	$\min \|A\mathbf{u} - \mathbf{y}\| = \min \|aA_1 + bA_2 - \mathbf{y}\|$
正交性條件	$(A\mathbf{u} - \mathbf{y}) \perp A_1,\ (A\mathbf{u} - \mathbf{y}) \perp A_2$
正規方程式	$A^T(A\mathbf{u} - \mathbf{y}) = \mathbf{0} \implies A^T A\mathbf{u} = A^T\mathbf{y}$
最小平方解	$\hat{\mathbf{u}} = (A^T A)^{-1} A^T\mathbf{y}$

例子2：為數據尋找最小平方線

從給定的數據中尋找$A\mathbf{u} = \mathbf{y}$ 的最小平方解。

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

解答

$A\mathbf{u} = \mathbf{y}$ 的最小平方解由以下公式給出：

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

此解能使$\|A\mathbf{u} - \mathbf{y}\|$ 最小化。計算過程：

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

最小平方線為

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

其中

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

我們將探討尋找最小平方曲線（直線）的過程，這稱為線性回歸（linear regression）。**例子3：為數據尋找合適的曲線（曲線擬合）**

考慮xu-平面上的4 個點（二維數據）：

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

讓我們找出最佳擬合曲線（即$u = a + bx + cx^2$ 的二次近似值）來描述(x_i, u_i)。由於這4 個數據點都應滿足二次方程式$u = a + bx + cx^2$，因此它們可以用以下的矩陣形式表示。**構建二次最小平方問題**

數據 (x,u)	二次函數$u = a + bx + cx^2$	線性方程組	矩陣表示
$(0,1)$	$1 = a + b \cdot 0 + c \cdot 0^2$		
$(1,3)$	$3 = a + b \cdot 1 + c \cdot 1^2$	$\begin{cases} a = 1 \\ a + b + c = 3 \\ a + 2b + 4c = 4 \\ a + 3b + 9c = 4 \end{cases}$	$A\mathbf{u} = \mathbf{y}$
$(2,4)$	$4 = a + b \cdot 2 + c \cdot 2^2$		其中$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$，$\mathbf{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$，$\mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix}$
$(3,4)$	$4 = a + b \cdot 3 + c \cdot 3^2$		

簡單來說，我們只需要為$u = a + bx + cx^2$ 尋找一個能最小化$\|A\mathbf{u} - \mathbf{y}\|$ 的向量$\hat{\mathbf{u}} = (a,b,c)$。要為數據集獲得這個二次函數，我們將得到一個沒有唯一解的方程組。因此，我們會有一個需要求解的最小平方問題$\min \|A\mathbf{u} - \mathbf{y}\|$。以下的函數$E(\mathbf{u})$ 代表了這個問題的誤差：

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

我們可以求得

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

所以最小平方曲線為

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

組件	數值
數據點	$(0,1)$, $(1,3)$, $(2,4)$, $(3,4)$
二次模型	$u = a + bx + cx^2$
設計矩陣A	$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$
最小平方解	$\hat{\mathbf{u}} = \begin{bmatrix} 1 \\ 1.5 \\ -0.5 \end{bmatrix}$
擬合二次曲線	$u = 1 + \frac{3}{2}x - \frac{1}{2}x^2$