Problem Set 4

November 25, 2023 11:04 AM

Problem 4.1

November 25, 2023 11:07 AM

Exercise 6.1 (Least squares and total least squares) Find the least-squares line and the total least-squares[3] line for the data points (x_i, y_i), $i = 1, \dots, 4$, with $x = (-1, 0, 1, 2)$, $y = (0, 0, 1, 1)$. Plot both lines on the same set of axes.

$y = ax + b \longrightarrow$ solve for a & x

$$\min_x \sum (y_i - ax_i - b_i)^2 \longleftrightarrow \min_x \Big\| \underbrace{\begin{bmatrix} \vdots \\ y \\ \vdots \end{bmatrix}}_{b} - \underbrace{\begin{bmatrix} \vdots & \vdots \\ x & \vdots \\ \vdots & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{X} \Big\|_2^2$$

Based on leasts squares:

$$\begin{bmatrix} a^* \\ b^* \end{bmatrix} = (A^T A)^{-1} A^T y$$

$$A = \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \quad X = \begin{bmatrix} a^* \\ b^* \end{bmatrix}$$

$$\begin{bmatrix} a^* \\ b^* \end{bmatrix} = \left(\begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}^T \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}^T \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

$$= \left(\begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

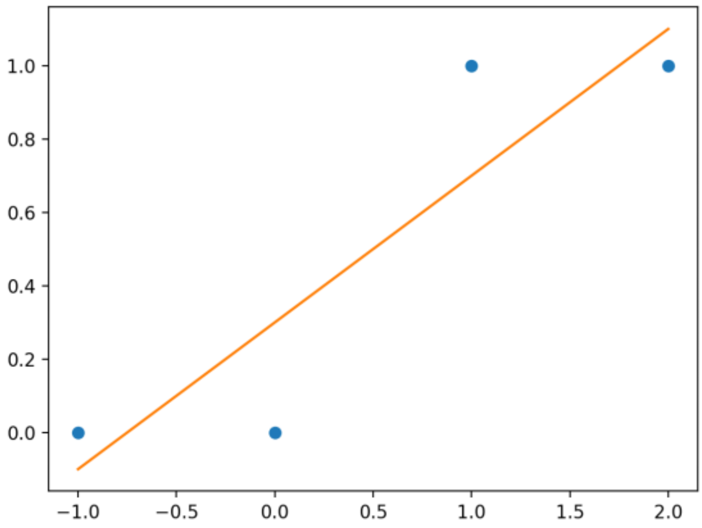
$$= \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$= \frac{1}{10} \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$= \frac{1}{10} \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{4}{10} \\ \frac{3}{10} \end{bmatrix}$$

$\therefore \ a = \frac{4}{10}, \ b = \frac{3}{10}$


1.0
0.8
0.6
0.4
0.2
0.0
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

Problem 4.2

November 25, 2023 11:10 AM

Exercise 6.3 (Lotka's law and least squares) Lotka's law describes the frequency of publication by authors in a given field. It states that $X^a Y = b$, where X is the number of publications, Y the relative frequency of authors with X publications, and a and b are constants (with $b > 0$) that depend on the specific field. Assume that we have data points (X_i, Y_i), $i = 1, \ldots, m$, and seek to estimate the constants a and b.

1. Show how to find the values of a, b according to a linear least-squares criterion. Make sure to define the least-squares problem involved precisely.

2. Is the solution always unique? Formulate a condition on the data points that guarantees unicity.

$$X^a Y = b$$

leasts square: $\min \|\vec{b} - A\vec{x}\|_2^2$

I want $X^a y = b$ in the form $\vec{b} - Ax$ with x relating to a & b

$$\vec{x} = \begin{bmatrix} b \\ x^a \end{bmatrix}$$

$$x^a y = b$$

$$x_2 y = x_1$$

$$0 = x_1 - x_2 y \longrightarrow \begin{bmatrix} 1 & y \end{bmatrix} \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}$$

$$\min \left\| \begin{bmatrix} 1 & y \end{bmatrix} \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix} - \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \right\|_2^2$$

$$x^* = (A^T A)^{-1} A^T b$$

$$x^* = (A^T A)^{-1} A^T b$$

$$= \left(\begin{bmatrix} 1 & y_i \\ \vdots & \vdots \\ 1 & y_m \end{bmatrix}^T \begin{bmatrix} 1 & y_i \\ \vdots & \vdots \\ 1 & y_m \end{bmatrix}^{-1} \right) \begin{bmatrix} 1 & y_i \\ \vdots & \vdots \\ 1 & y_m \end{bmatrix}^T \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

2.

Problem 4.3

November 25, 2023 11:11 AM

Exercise 6.6 The Michaelis–Menten model for enzyme kinetics relates the rate y of an enzymatic reaction to the concentration x of a substrate, as follows:

$$y = \frac{\beta_1 x}{\beta_2 + x},$$

where β_i, $i = 1, 2$, are positive parameters.

1. Show that the model can be expressed as a linear relation between the values $1/y$ and $1/x$.

2. Use this expression to find an estimate $\hat{\beta}$ of the parameter vector β using linear least squares, based on m measurements (x_i, y_i), $i = 1, \ldots, m$.

3. The above approach has been found to be quite sensitive to errors in input data. Can you experimentally confirm this opinion?

1. $$y = \frac{\beta_1 x}{\beta_2 + x}$$

$$\frac{1}{y} = \frac{\beta_2 + x}{\beta_1 x} = \frac{\frac{1}{x}(\beta_2 + x)}{\frac{1}{x}\beta_1 x} = \frac{\beta_2\left(\frac{1}{x}\right) + 1}{\beta_1}$$

$$\frac{1}{y} = \frac{\beta_2}{\beta_1}\left(\frac{1}{x}\right) + \frac{1}{\beta_1}$$

2. $$\min \sum_{i=1}^{m} \left(\frac{1}{y} - \frac{\beta_2}{\beta_1}\frac{1}{x} - \frac{1}{\beta_1} \right)^2$$

$$\min \left\| \underbrace{\begin{bmatrix} \frac{1}{y_1} \\ \vdots \\ \frac{1}{y_m} \end{bmatrix}}_{b} - \underbrace{\begin{bmatrix} \frac{1}{x} & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \frac{\beta_2}{\beta_1} \\ \frac{1}{\beta_1} \end{bmatrix}}_{x} \right\|_2^2$$

$$x^* = \begin{bmatrix} \frac{\beta_2}{\beta_1} \\ \frac{1}{\beta_1} \end{bmatrix} = \left(\begin{bmatrix} \frac{1}{x} & 1 \end{bmatrix}^T \begin{bmatrix} \frac{1}{x} & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} \frac{1}{x} & 1 \end{bmatrix}^T \begin{bmatrix} \frac{1}{y_1} \\ \vdots \\ \frac{1}{y} \end{bmatrix}$$

$$\hat{} = \begin{bmatrix} \vdots \\ \frac{1}{\beta_1} \end{bmatrix}^{-1} \left(\begin{bmatrix} x \end{bmatrix} \begin{bmatrix} x \end{bmatrix} \right) \begin{bmatrix} \bar{x} \end{bmatrix} \begin{bmatrix} \vdots \\ \frac{1}{y_m} \end{bmatrix}$$

3. If y or x are close to zero, then you'll get really weird results since its dividing by $\simeq 0$

Problem 4.4

November 25, 2023 11:12 AM

Problem 4.4 (Solving least squares problems using the Moore-Penrose pseudoinverse) In this problem you derive the result that the Moore-Penrose pseudoinverse can be used to solve least squares problems for overdetermined, underdetermined, and rank-deficient systems. Recall that least square problems consider the system of linear equations $Ax = b$ where $A \in \mathbb{R}^{m,n}$ and $\text{rank}(A) = r$. If the compact SVD of A is $A = U_r \Sigma V_r^T$ where $\Sigma \in \mathbb{R}^{r,r}$ is a positive definite diagonal matrix of (non-zero) singular values, $U_r \in \mathbb{R}^{m,r}$ and $V_r \in \mathbb{R}^{n,r}$ each contain orthonormal columns, then the Moore-Penrose pseudoinverse is $A^\dagger = V_r \Sigma^{-1} U_r^T$.

(a) Recall that the *overdetermined* least squares problem considers an $A \in \mathbb{R}^{m,n}$ when $m > n$ (more constraints in the y vector than parameters in the x vector). Here the objective is to find an x that minimizes $\|Ax - y\|_2$. Show that an optimal solution is $x^* = A^\dagger y$. To show this recall that an optimal solution x^* must satisfy the "normal" equations $A^T A x^* = A^T y$. Verify that $x^* = A^\dagger y$ satisfies the normal equations.

(b) Recall that the *underdetermined* least squares problem considers an $A \in \mathbb{R}^{m,n}$ when $m < n$ (fewer constraints in the y vector than parameters in the x vector). Here the objective is to find the x that minimizes $\|x\|_2$ while satisfying $Ax = y$ (equivalently, satisfying $\|Ax - y\|_2 = 0$). Show that the optimal solution $x^* = A^\dagger y$. To show this recall that the optimal solution x^* must satisfy two conditions: (i) $x^* \in \mathcal{R}(A^T)$ and (ii) $Ax^* = y$. Verify that $x^* = A^\dagger y$ satisfies (i) all the time. Under what conditions does x^* satisfy condition (ii)? (Hint, think about the rank of A.)

In the above two parts we haven't explicitly considered the role of the rank r of the A matrix. But we also note that the only place where the rank of A comes into the discussion of parts (a) and (b) is in the discussion of condition (ii) of part (b).

Recall that $r \leq \min\{m, n\}$. When $r = \min\{m, n\}$ the A matrix is full column rank in the overdetermined problem and is full row rank in the underdetermined problem. In both these full-rank cases x^* has the simple expression presented in class. Now we consider what happens when $r < \min\{m, n\}$.

(c) In this part consider the situation where $\text{rank}(A) = r < m < n$. This is a "rank-deficient" underdetermined least squares problem. If we set $x^* = A^\dagger y$, what characteristics do Ax^* and $\|x^*\|_2$ satisfy? (Hint: this is a type of hybrid problem that at the same time can have characteristics of both overdetermined and underdetermined least squares.)

a) $x^* = A^+ y = V_r \Sigma^{-1} U_r^T = (A^T A)^{-1} A^T$

must satisfy $A^T A x^* = A^T y$

$A^+ = (A^T A)^{-1} A^T$

$A^T A (A^T A)^{-1} A^T y$

$= I A^T y$

$= A^T y$

$= A^T y$

$(U_r \Sigma V_r^T)^T \cancel{(U_r \Sigma V_r^T)(V_r \Sigma^{-1} U_r^T)} y$

$= (U_r \Sigma V_r^T)^T y$

$= A^T y$

b) i) $x^* = A^+ y = (V_r \Sigma^{-1} U_r^T) y = (AA^T) A^T$

Since A^+ is based on A

$A^+ \subset R(A)$

Since $A^+ \subset R(A)$,

$A^+ y \subset R(A)$

$x^* \subset R(A)$ always

ii) Does x^* satisfy $Ax^* = y$

$Ax^* = A(AA^T)^{-1} A^T y$

$$= AA^T (AA^T)^{-1} y$$

$$= y$$

this depends on if the Matrix A is full rank, or else you cannot simply commute A^T to the other side

c) $x^* = A^+ y$ where $\text{Rank}(A) = r < m < n$