Solution to homework 2

February 6, 2023

1 Problem 1

We would like to minimize this function:

$$L(m,b)=\sum_{i=1}^{n}(mx_i+b-y_i)^2. \tag{1.1}$$

Compute the first-order derivative:

$$\frac{\partial L}{\partial m}=2\sum_{i=1}^{n}(mx_i+b-y_i)x_i, \quad \frac{\partial L}{\partial b}=2\sum_{i=1}^{n}(mx_i+b-y_i). \tag{1.2}$$

Compute the second-order derivative:

$$\frac{\partial^2 L}{\partial m^2}=2\sum_{i=1}^{n}x_i^2, \quad \frac{\partial^2 L}{\partial b^2}=2\sum_{i=1}^{n}1, \quad \frac{\partial^2 L}{\partial m\partial b}=2\sum_{i=1}^{n}x_i. \tag{1.3}$$

Then the Hessian matrix is:

$$H=\begin{pmatrix}\frac{\partial^2 L}{\partial m^2} & \frac{\partial^2 L}{\partial m\partial b}\\ \frac{\partial^2 L}{\partial m\partial b} & \frac{\partial^2 L}{\partial b^2}\end{pmatrix}=\begin{pmatrix}2\sum_{i=1}^{n}x_i^2 & 2\sum_{i=1}^{n}x_i\\ 2\sum_{i=1}^{n}x_i & 2\sum_{i=1}^{n}1\end{pmatrix} \tag{1.4}$$

We introduce the matrix $X\in\mathbb{R}^{n\times 2}$:

$$X=\begin{pmatrix}x_1 & 1\\ x_2 & 1\\ \cdots & \cdots\\ x_n & 1\end{pmatrix} \tag{1.5}$$

Then

$$X^TX=\begin{pmatrix}x_1 & x_2 & \cdots & x_n\\ 1 & 1 & \cdots & 1\end{pmatrix}\begin{pmatrix}x_1 & 1\\ x_2 & 1\\ \cdots & \cdots\\ x_n & 1\end{pmatrix}=\begin{pmatrix}\sum_{i=1}^{n}x_i^2 & \sum_{i=1}^{n}x_i\\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}1\end{pmatrix} \tag{1.6}$$

Therefore, we have

$$H=2X^TX \tag{1.7}$$

For any vector $z\in\mathbb{R}^2$, we compute

$$z^THz=2z^TX^TXz=2(Xz)^TXz=2\left\|Xz\right\|^2\geq 0. \tag{1.8}$$

Moreover, it is easy to see that $Xz=0$ indicates $z=0$, under the assumption that x_i are not identical. This proves that H is positive-definite.

2 Problem 2

We would like to minimize the function

$$L(m_1, m_2, b) = \sum_{i=1}^{n} (m_1 x_i^{(1)} + m_2 x_i^{(2)} + b - y_i)^2 \tag{2.1}$$

1. Compute the derivatives:

$$\frac{\partial L}{\partial m_1} = 2 \sum_{i=1}^{n} (m_1 x_i^{(1)} + m_2 x_i^{(2)} + b - y_i) x_i^{(1)} \tag{2.2}$$

$$\frac{\partial L}{\partial m_2} = 2 \sum_{i=1}^{n} (m_1 x_i^{(1)} + m_2 x_i^{(2)} + b - y_i) x_i^{(2)} \tag{2.3}$$

$$\frac{\partial L}{\partial m_2} = 2 \sum_{i=1}^{n} (m_1 x_i^{(1)} + m_2 x_i^{(2)} + b - y_i) \tag{2.4}$$

2. Setting

$$\frac{\partial L}{\partial m_1} = 0, \quad \frac{\partial L}{\partial m_2} = 0, \quad \frac{\partial L}{\partial b} = 0. \tag{2.5}$$

We have

$$\begin{aligned} \sum_{i=1}^{n} (m_1 x_i^{(1)} + m_2 x_i^{(2)} + b - y_i) x_i^{(1)} = 0, \\ \sum_{i=1}^{n} (m_1 x_i^{(1)} + m_2 x_i^{(2)} + b - y_i) x_i^{(2)} = 0, \\ \sum_{i=1}^{n} (m_1 x_i^{(1)} + m_2 x_i^{(2)} + b - y_i) = 0. \end{aligned} \tag{2.6}$$

We can rewrite it as a linear system of (m_1, m_2, b):

$$A \begin{pmatrix} m_1 \\ m_2 \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} x_i^{(1)} y_i \\ \sum_{i=1}^{n} x_i^{(2)} y_i \\ \sum_{i=1}^{n} y_i \end{pmatrix} \tag{2.7}$$

where

$$A = \begin{pmatrix} \sum_{i=1}^{n} (x_i^{(1)})^2 & \sum_{i=1}^{n} x_i^{(1)} x_i^{(2)} & \sum_{i=1}^{n} x_i^{(1)} \\ \sum_{i=1}^{n} x_i^{(1)} x_i^{(2)} & \sum_{i=1}^{n} (x_i^{(2)})^2 & \sum_{i=1}^{n} x_i^{(2)} \\ \sum_{i=1}^{n} x_i^{(1)} & \sum_{i=1}^{n} x_i^{(2)} & \sum_{i=1}^{n} 1 \end{pmatrix} \tag{2.8}$$

We introduce the matrix $X \in \mathbb{R}^{n \times 3}$:

$$X = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & 1 \\ x_2^{(1)} & x_2^{(2)} & 1 \\ \cdots & \cdots & \cdots \\ x_n^{(1)} & x_n^{(2)} & 1 \end{pmatrix} \tag{2.9}$$

and

$$Y = (y_1, y_2, \cdots, y_n)^T. \tag{2.10}$$

Then it is easy to verify that $A = X^T X$ and the linear system can be rewritten as

$$X^T X \begin{pmatrix} m_1 \\ m_2 \\ b \end{pmatrix} = X^T Y \tag{2.11}$$

Therefore, we can solve out

$$\begin{pmatrix} m_1 \\ m_2 \\ b \end{pmatrix} = (X^T X)^{-1} X^T Y. \tag{2.12}$$

3. By computing second order derivatives, we can prove that the Hessian matrix can be rewritten as:

$$H = 2X^T X \tag{2.13}$$

Therefore, for any vector $z \in \mathbb{R}^3$, we compute

$$z^T H z = 2z^T X^T X z = 2(Xz)^T X z = 2 \|Xz\|^2 \geq 0. \tag{2.14}$$

Moreover, it is easy to see that $Xz = 0$ indicates $z = 0$, under the assumption that $(x_i^{(1)}, x_i^{(2)})$ are not identical. This proves that H is positive-definite.