数据处理中的矩阵方法-第二次作业

李厚华 202418019427056

2025 年 3 月 4 日

1 病态情形

当 X^TX 不可逆时，可以利用奇异值分解（SVD）和伪逆的方法，求解 $c=(X^TX)^{-1}X^Ty$ 的具体过程如下：

1. 对矩阵 X 进行奇异值分解

设 X 是一个 $m\times n$ 的矩阵（m 为样本数量，n 为特征数量），通过奇异值分解可以将 X 表示为：$X=U\Sigma V^T$

其中：

- U 是一个 $m\times m$ 的正交矩阵，其列向量是 XX^T 的特征向量。
- Σ 是一个 $m\times n$ 的对角矩阵，对角线上的元素 σ_i （$i=1,\cdots,\min(m,n)$）称为奇异值，且 $\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_{\min(m,n)}\geq 0$。
- V 是一个 $n\times n$ 的正交矩阵，其列向量是 X^TX 的特征向量。

2. 求 $(X^TX)^{-1}$

根据矩阵运算规则：

$$X^TX=(U\Sigma V^T)^T(U\Sigma V^T)=V\Sigma^TU^TU\Sigma V^T=V\Sigma^T\Sigma V^T \tag{1}$$

当 X^TX 不可逆时，意味着部分奇异值 $\sigma_i=0$。我们可以通过求 $\Sigma^T\Sigma$ 的伪逆来处理，设 Σ^+ 是 Σ 的伪逆矩阵，它也是对角矩阵，对于 $\sigma_i\neq 0$ 的元素，$(\Sigma^+)_{ii}=\frac{1}{\sigma_i}$；对于 $\sigma_i=0$ 的元素，$(\Sigma^+)_{ii}=0$。

那么 (X^TX) 的伪逆为：

$$(X^TX)^+=V(\Sigma^T\Sigma)^+V^T=V\Sigma^+(\Sigma^+)^TV^T \tag{2}$$

3. 计算 c

由 $c=(X^TX)^{-1}X^Ty$ 可得：

$$\begin{aligned} c&=(X^TX)^+X^Ty=V\Sigma^+(\Sigma^+)^TV^TV\Sigma^TU^Ty\\ &=V\Sigma^+(\Sigma^+)^T\Sigma^TU^Ty \end{aligned} \tag{3}$$

4. 其他方法

岭回归（Ridge Regression）：在原目标函数中加入 L_2 正则项，即

$$J(c) = \|Xc^T - Y\|^2 + \lambda c^T c \tag{4}$$

这样 $X^T X + \lambda I$ 可逆，其中 I 是单位矩阵，λ 是正则化参数。解为：

$$c = (X^T X + \lambda I)^{-1} X^T y \tag{5}$$

2 异常情形

异常点对最小二乘法影响较大，可采取以下方法：

1. 鲁棒回归：改用对异常值不敏感的损失函数，如：Huber 损失（结合 L1 和 L2 损失）。最小绝对偏差（LAD）（L1 损失）。
2. 数据清洗：检测并删除异常点（如基于残差）。
3. RANSAC 算法：迭代拟合模型并排除离群点。
4. 加权最小二乘法：降低异常点的权重（如使用 Tukey 双权重函数）。

3 等式约束下的最小二乘解

本题可使用拉格朗日乘数法求解等式约束的最小二乘问题。

1. 构建拉格朗日函数

设拉格朗日乘数为 λ, 则拉格朗日函数为：

$$\begin{aligned} L(\mathbf{c}, \lambda) &= \|\mathbf{y} - \mathbf{X}\mathbf{c}\|^2 + \lambda(\mathbf{c}^T \mathbf{1} - 1) \\ &= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\mathbf{c} - \mathbf{c}^T\mathbf{X}^T\mathbf{y} + \mathbf{c}^T\mathbf{X}^T\mathbf{X}\mathbf{c} + \lambda(\mathbf{c}^T\mathbf{1} - 1) \end{aligned} \tag{6}$$

2. 求偏导数并令其为 0

分别对 $\mathbf{c}$ 和 λ 求偏导数：

- 对 $\mathbf{c}$ 求偏导数：

$$\frac{\partial L}{\partial \mathbf{c}} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\mathbf{c} + \lambda\mathbf{1} = 0 \tag{7}$$

整理可得：

$$\mathbf{X}^T\mathbf{X}\mathbf{c} = \mathbf{X}^T\mathbf{y} - \frac{\lambda}{2}\mathbf{1} \tag{8}$$

- 对 λ 求偏导数：

$$\frac{\partial L}{\partial \lambda} = \mathbf{c}^T\mathbf{1} - 1 = 0 \tag{9}$$

3. 求解 $\mathbf{c}$

由 $\mathbf{X}^T\mathbf{X}\mathbf{c} = \mathbf{X}^T\mathbf{y} - \frac{\lambda}{2}\mathbf{1}$ 可得 $\mathbf{c} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{y} - \frac{\lambda}{2}\mathbf{1})$，将其代入 $\mathbf{c}^T\mathbf{1} = 1$ 中:

$$
\begin{aligned}
\left[(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{y} - \frac{\lambda}{2}\mathbf{1})\right]^T \mathbf{1} = 1 \\
\mathbf{1}^T(\mathbf{X}^T\mathbf{X})^{-T}(\mathbf{X}^T\mathbf{y} - \frac{\lambda}{2}\mathbf{1}) = 1 \\
\mathbf{1}^T(\mathbf{X}^T\mathbf{X})^{-T}\mathbf{X}^T\mathbf{y} - \frac{\lambda}{2}\mathbf{1}^T(\mathbf{X}^T\mathbf{X})^{-T}\mathbf{1} = 1
\end{aligned}
$$

可以解得:

$$
\lambda = \frac{2(\mathbf{1}^T(\mathbf{X}^T\mathbf{X})^{-T}\mathbf{X}^T\mathbf{y} - 1)}{\mathbf{1}^T(\mathbf{X}^T\mathbf{X})^{-T}\mathbf{1}} \tag{10}
$$

将 λ 代回 $\mathbf{c} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{y} - \frac{\lambda}{2}\mathbf{1})$, 可得最小二乘解:

$$
\mathbf{c} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{y} - \frac{(\mathbf{1}^T(\mathbf{X}^T\mathbf{X})^{-T}\mathbf{X}^T\mathbf{y} - 1)\mathbf{1}}{\mathbf{1}^T(\mathbf{X}^T\mathbf{X})^{-T}\mathbf{1}}) \tag{11}
$$