机器学习入门 课程 4 线性代数

线性回归项目 解题合集

1 (选做) 2.4 算法正确判断了奇异矩阵

1.1 要求

证明下面的命题：

如果方阵 A 可以被分为4个部分:

$A = \begin{bmatrix} I & X \\ Z & Y \end{bmatrix}$，其中 I 为单位矩阵，Z 为全0矩阵，Y 的第一列全0，

那么A为奇异矩阵。

提示：从多种角度都可以完成证明

- 考虑矩阵 Y 和 矩阵 A 的秩
- 考虑矩阵 Y 和 矩阵 A 的行列式
- 考虑矩阵 A 的某一列是其他列的线性组合

1.2 参数假设

设 A 的阶数为 m1+m2，其中 I 阶数为 m1，Y 阶数为 m2。可知 Z 为 m2*m1 阶矩阵，X 为 m1*m2 阶矩阵。

1.3 证明方法一

我们截取方阵 A 的下半部分，得到一个扁长形矩阵，阶数为 m2*(m1+m2)，如下：

A_down = [Z   Y]

由于 Y 的第一列为全 0，因此 r(A) < m2。因此我们能够通过一组初等行变换将 Y 的最后一行变为全 0.

又因为 Z 是全 0，因此我们可以用同样的一组初等行变换将 A_down 的最后一行变为全 0。

综上，我们相当于在未依靠 A 矩阵的上半部分行的情况下，仅利用 A 矩阵的下半部分行进行初等行变换，就能够将 A 矩阵的最后一行变为全 0。这等于说我们能够对 A 作初等行变换，将 A 的最后一行变为全 0。因而 A 必然是奇异矩阵。

得证。

1.4 证明方法二

我们已知一个性质，那就是如果有两个矩阵 A1 和 A2，其秩分别为 r1 和 r2，那么把 A1 和 A2 按行或者按列组合（假设 2 个矩阵维度满足组合条件），新的矩阵 A 一定有：

r(A) = r(A1,A2)≤ r1+r2…………式（0）

将 A 按行分割为：

A1 = [I X] 阶数为 m1*(m1+m2)，设秩为 r1

A2 = [Z Y] 阶数为 m2*(m1+m2)，设秩为 r2

根据秩的基本性质可知

$r1 \leq m1$…………式（1）

由于 Z 为全 0，Y 的第 1 列为 0，又因为 Z 有 m1 列，因此矩阵 A2 的前(m1+1)列都是 0，不为 0 的列最多为(m1+m2)-(m1+1)=m2-1 列，因此可得：

$r2 \leq m2 - 1$…………式（2）

综上得到式 3：

$r1 + r2 \leq m1 + m2 - 1$…………式（3）

代入式（0），得到：

r(A) $\leq m1 + m2 - 1$…………式（4）

而 m1+m2 正是 A 的阶数，因此 A 必然是非奇异矩阵。

得证。

附：式（0）的证明方法大致为：

因为 (A1,A2) 的列向量 可由 A1 的一个极大无关组与 A2 的一个极大无关组合并的向量组线性表示

所以 r(A1,A2) <= r(A1 的一个极大无关组与 A2 的一个极大无关组合并的向量组) <= r(A1)+r(A2)

1.5 证明方法三

设矩阵 X 的第 1 列如下：

$$[x_1, x_2, \dots\dots, x_{m1}]^T$$

于是矩阵 A 大致可以表示为如下样式：

$$\begin{matrix} 1 & \dots & 0 & x_1 & \dots & ? \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & x_{m1} & \dots & ? \\ 0 & \dots & 0 & 0 & \dots & ? \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & \dots & ? \end{matrix}$$

注意到，第 m1+1 列的形式为

$$[x_1, x_2, \dots\dots, x_{m1}, 0, \dots\dots, 0]^T$$

可以表示为:

$$x_1 * [1,0,\dots,0] + x_2 * [0,1,0,\dots,0] + \cdots + x_{m1} * [0,\dots,1,0,\dots,0]$$

即第 m1+1 列可以表示为前 m1 列的线性组合，组合系数就是 $(x_1, x_2, \dots\dots, x_{m1})$。

因此，A 的列向量组不是线性无关的，故 A 必然是奇异矩阵。命题得证。

2 (选做) 3.3 找到参数 m,b 使得平方平均误差最小

2.1 要求

假定我们得到了 n 个样本点：

$$(x_1, y_1), (x_2, y_2), \dots\dots, (x_n, y_n)$$

我们假定 x 与 y 的关系为：

$$y = m * x + b$$

现在要求：找到一组参数 m,b，使得该关系对于样本点的拟合效果最好。拟合效果最好的标准是拟合直线与样本点间的“平均平方误差”（MSE）最小。

MSE 的计算方法为：

$$\text{MSE} = \frac{1}{n} * \sum_{\mathrm{i}=1}^{n} (y_i - m * x_i - b)^2$$

以下按提示步骤进行解答

2.2 3.3.1 计算目标函数相对于参数的导数

2.2.1 要求

已知：

$$E = \frac{1}{2} \sum_{i=1}^{n} (y_i - mx_i - b)^2$$

证明以下公式：

$$\frac{\partial E}{\partial m} = \sum_{i=1}^{n} -x_i(y_i - mx_i - b)$$

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{n} -(y_i - mx_i - b)$$

2.2.2 证明过程：

首先，根据求导的线性特征：

$$\frac{\partial(y_1 + y_2 + \cdots + y_n)}{\partial x} = \frac{\partial y_1}{\partial x} + \frac{\partial y_2}{\partial x} + \cdots + \frac{\partial y_n}{\partial x} \ldots\ldots\ldots\ldots 式(1)$$

故原式可以写为：

$$\frac{\partial E}{\partial m} = \frac{1}{2} * \sum_{i=1}^{n} \frac{\partial((y_i - m * x_i - b)^2)}{\partial m} \ldots\ldots\ldots\ldots 式(2)$$

$$\frac{\partial E}{\partial b} = \frac{1}{2} * \sum_{i=1}^{n} \frac{\partial((y_i - m * x_i - b)^2)}{\partial b} \ldots\ldots\ldots\ldots 式(3)$$

于是，问题转化为已知函数 f：

$$\mathrm{f(m, b)} = (y_i - m * x_i - b)^2 \ldots\ldots\ldots\ldots 式(4)$$

求:

$$\frac{\partial f}{\partial m}\text{和}\frac{\partial f}{\partial b}$$

函数 f 本身又可以看作一个复合函数:

$$\mathrm{f(g)} = g^2$$

$$\mathrm{g(m,b)} = y_i - m * x_i - b$$

利用复合函数求导公式:

$$\frac{\partial f}{\partial m} = \frac{\partial f}{\partial g} * \frac{\partial g}{\partial m} = \left(\frac{\partial g^2}{\partial g}\right) * \left(\frac{\partial(\,-x_i * m\ +\ (y_i - b)\,)}{\partial m}\right)$$

$$= (2 * g) * (-x_i) = -2x_i * (y_i - m * x_i - b) \dots\dots\dots\dots \text{式}(5)$$

同理可得

$$\frac{\partial f}{\partial m} = \frac{\partial f}{\partial g} * \frac{\partial g}{\partial b} = \left(\frac{\partial g^2}{\partial g}\right) * \left(\frac{\partial(\,-b\ +\ (y_i - x_i * m)\,)}{\partial b}\right)$$

$$= (2 * g) * (-x_i) = -2 * (y_i - m * x_i - b)\ \dots\dots\dots\dots \text{式}(6)$$

将式(5)和式(6)代入式(2)和式(3),得到式(7) 式(8):

$$\frac{\partial E}{\partial m} = \sum_{i=1}^{n} -x_i * (y_i - m * x_i - b) \dots\dots\dots\dots \text{式}(7)$$

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{n} -(y_i - m * x_i - b) \dots\dots\dots\dots \text{式}(8)$$

命题得证

2.3 3.3.2 实例推演

2.3.1 要求

现在我们有了一个二元二次方程组

$$\begin{cases} \sum_{i=1}^{n} -x_i(y_i - mx_i - b) = 0 \\ \sum_{i=1}^{n} -(y_i - mx_i - b) = 0 \end{cases}$$

为了加强理解，我们用一个实际例子演练。

我们要用三个点 $(1, 1), (2, 2), (3, 2)$ 来拟合一条直线 y = m*x + b, 请写出

- 目标函数 E,
- 二元二次方程组，
- 并求解最优参数 m, b

2.3.2 求解过程

2.3.2.1 求目标函数

根据下列公式:

$$E = \frac{1}{2}\sum_{i=1}^{n}(y_i - mx_i - b)^2 \quad \cdots\cdots式（9）$$

根据题意，代入以下数值：
n = 3，$(x_1, y_1) = (1,1)$ $(x_2, y_2) = (2,2)$ $(x_3, y_3) = (3,2)$，得到：

得到：

$$\mathrm{E} = (1 - m - b)^2 + (2 - 2m - b)^2 + (2 - 3m - b)^2 \ldots\ldots 式（10）$$

目标方程得解，如式（2）所示。

2.3.2.2 求二元二次方程组

将以下数值代入步骤一的式（7）和式（8），得：

$$\frac{\partial E}{\partial m} = -1 * [1 * (1 - m - b) + 2 * (2 - 2m - b) + 3 * (2 - 3m - b)]$$

$$\dots\dots\dots\dots 式(11)$$

$$\frac{\partial E}{\partial b} = -1 * [(1 - m - b) + (2 - 2m - b) + (2 - 3m - b)] \dots\dots\dots\dots 式(12)$$

化简得到

$$\frac{\partial E}{\partial m} = 14m + 6b - 11 \dots\dots\dots\dots 式(13)$$

$$\frac{\partial E}{\partial b} = 6m + 3b - 5 \dots\dots\dots\dots 式(14)$$

因此最优方程组为：

$$14m + 6b = 11 \dots\dots\dots\dots 式(15)$$

$$6m + 3b = 5 \dots\dots\dots\dots 式(16)$$

二元二次方程组得解，如式（15）（16）所示。

2.3.2.3 求解目标参数

对上一步得到的二元二次方程进行求解，首先消元：

$$式(7) - 式(8) * \frac{7}{3}\text{：} 6\mathrm{b} - 3\mathrm{b} * \frac{7}{3} = 11 - 5 * \frac{7}{3} \dots\dots\dots\dots 式(17)$$

式（17）化简得到

$$b = \frac{2}{3} \dots\dots\dots\dots 式(18)$$

将式（18）代入式（17），得到：

$$\mathrm{m} = \frac{1}{6} * (5 - 3\mathrm{b}) = \frac{1}{6} * (5 - 2) = \frac{1}{2} \dots\dots\dots\dots 式(19)$$

得解：

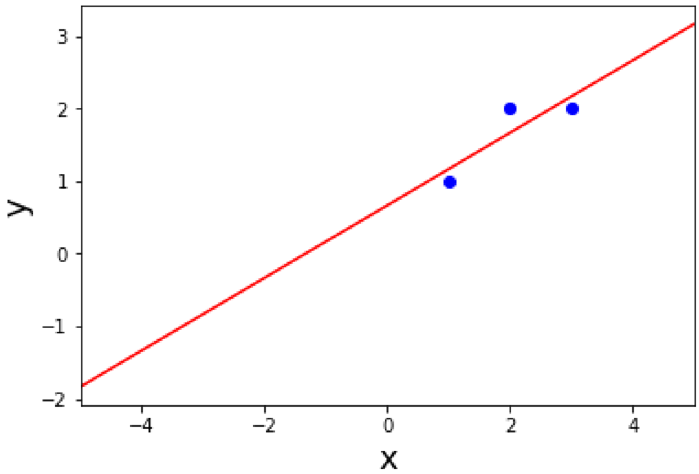
$$\mathrm{m} = \frac{1}{2}, \mathrm{b} = \frac{2}{3} \dots\dots\dots\dots 式(20)$$

因此对所给点的拟合曲线为：

$$\mathrm{y} = \frac{1}{2} * x + \frac{2}{3} \dots\dots\dots\dots 式(21)$$

2.3.2.4 验证

使用 python 绘出散点图及 m-b 似合直线，如下：



目测该拟合直线基本正确。

2.4 3.3.3 将方程组写成矩阵形式

2.4.1 要求

我们的二元二次方程组可以用更简洁的矩阵形式表达，将方程组写成矩阵形式更有利于我们使用 Gaussian Jordan 消元法求解。

请证明

$$\begin{bmatrix} \frac{\partial E}{\partial m} \\ \frac{\partial E}{\partial b} \end{bmatrix} = X^T X h - X^T Y$$

其中向量 Y, 矩阵 X 和 向量 h 分别为：

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix}, X = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \ldots & \ldots \\ x_n & 1 \end{bmatrix}, h = \begin{bmatrix} m \\ b \end{bmatrix}$$

TODO 证明:

2.4.2 证明过程

证明方法为从头计算下述算式：

$$X^T \mathrm{Xh} - X^T \mathrm{Y} \ldots \ldots \ldots \ldots 式(22)$$

将变量

$$\mathrm{X} = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, h = \begin{bmatrix} m \\ b \end{bmatrix}, \mathrm{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

代入式$X^T\mathrm{Xh}$，得到：

$$X^T\mathrm{Xh} = \begin{bmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & 1 \end{bmatrix} * \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} * \begin{bmatrix} m \\ b \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{bmatrix} * \begin{bmatrix} m \\ b \end{bmatrix}$$

$$= \begin{bmatrix} m * \sum_{i=1}^{n} x_i^2 + b * \sum_{i=1}^{n} x_i \\ m * \sum_{i=1}^{n} x_i + b * n \end{bmatrix} \ldots 式（23）$$

代入式$X^T\mathrm{Y}$，得到：

$$X^T\mathrm{Y} = \begin{bmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & 1 \end{bmatrix} * \begin{bmatrix} y_1 \\ \cdots \\ y_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} x_i * y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix} \ldots 式（24）$$

两式相减，得到：

$$X^T\mathrm{Xh} - X^T\mathrm{Y} = \begin{bmatrix} m * \sum_{i=1}^{n} x_i^2 + b * \sum_{i=1}^{n} x_i \\ m * \sum_{i=1}^{n} x_i + b * n \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{n} x_i * y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}$$

$$=\begin{bmatrix} m*\sum_{i=1}^{n}x_i^2+b*\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}x_i*y_i \\ m*\sum_{i=1}^{n}x_i+b*n-\sum_{i=1}^{n}y_i \end{bmatrix}$$

$$=\begin{bmatrix} m*\sum_{i=1}^{n}x_i^2+b*\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}x_i*y_i \\ m*\sum_{i=1}^{n}x_i+b*\sum_{i=1}^{n}1-\sum_{i=1}^{n}y_i \end{bmatrix}$$

$$=\begin{bmatrix} \sum_{i=1}^{n}(m*x_i^2+b*x_i-x_i*y_i) \\ \sum_{i=1}^{n}(m*x_i+b-y_i) \end{bmatrix}$$

$$=\begin{bmatrix} \sum_{i=1}^{n}-x_i*(y_i-m*x_i-b) \\ \sum_{i=1}^{n}-(y_i-m*x_i-b) \end{bmatrix}\text{... 式（25）}$$

根据式（7）和式（8）:

$$\frac{\partial E}{\partial m}=\sum_{i=1}^{n}-x_i*(y_i-m*x_i-b)\text{ ... ... 式(7)}$$

$$\frac{\partial E}{\partial b}=\sum_{i=1}^{n}-(y_i-m*x_i-b)\text{ ... ... 式(8)}$$

对比可知:

$$X^T\mathrm{X}h-X^T\mathrm{Y}=\begin{bmatrix} \frac{\partial E}{\partial m} \\ \frac{\partial E}{\partial b} \end{bmatrix}\text{... 式（26）}$$

得证。