

Линейная регрессия

Метод наименьших квадратов как решение матричного уравнения. Часть 2

Skillbox

образовательная платформа

Для решения задачи столбцы матрицы X должны быть линейно независимы.

Если признаки модели связаны строгой функциональной зависимостью, то определитель матрицы $X^T X$ будет равен нулю.

Пример

Пусть матрица $X =$

1	2	1
2	4	1
3	6	1

Тогда $X^T =$

1	2	3
2	4	6
1	1	1

$X^T X =$

6	11	16
11	21	31
16	31	46

И определитель $|X^T X| = 0$

Так как определитель матрицы равен 0, то мы не сможем найти обратную матрицу $(X^T X)^{-1}$. И тогда невозможно будет найти вектор w .

Мультиколлинеарность

Мультиколлинеарность — тесная корреляционная взаимосвязь между выбираемыми для анализа признаками.

Абсолютную — когда определитель матрицы $X^T X$ равен нулю.

Частичную — определитель матрицы $X^T X$ в точности не равен 0, но мало от него отличается.

Подходы, чтобы устранить эффекты мультиколлинеарности

- 1 Произвести отбор признаков
- 2 Преобразовать сами признаки
- 3 Получение смещённых оценок

Получения смещённых оценок

Было $w = (X^T X)^{-1} X^T y$

Стало

$w = (X^T X + \alpha I)^{-1} X^T y,$

где I — единичная матрица, а коэффициент α , как правило, от 0,1 до 0,4

Пример

Пусть у нас имеется выборка из 5 объектов, описанных двумя признаками. То есть $n = 5$, $k = 2$.
И есть значения целевой функции:

X_1	X_2	Y
1	2	6
2	4	11
3	6	16
4	8	21
5	10	26

Искомая функция выглядит следующим образом:

$$\hat{y} = w_0 + w_1 \times X_1 + w_2 \times X_2$$

Ваша задача — определить значения вектора весов: w_0 и w_1 и w_2

Пример

Дополним фиктивную переменную x_0

x_0	x_1	x_2
1	1	2
1	2	4
1	3	6
1	4	8
1	5	10

Это и есть матрица X .

Тогда транспонированная матрица X^T

1	1	1	1	1
1	2	3	4	5
2	4	6	8	10

Тогда матрица $X^T X$

5	15	30
15	55	110
30	110	220

Но определитель этой матрицы равен 0.

Пример

Дополним матрицу $X^T X + \alpha I$, пусть $\alpha = 0,1$

Тогда матрица $X^T X$

5	15	30
15	55	110
30	110	220

Матрица I

1	0	0
0	1	0
0	0	1

Матрица αI

0,1	0	0
0	0,1	0
0	0	0,1

Тогда матрица $X^T X + \alpha I$

5,1	15	30
15	55,1	110
30	110	220,1

Определитель этой матрицы равен 27,8.

Пример

Матрица, обратная к $X^T X + \alpha I$

0,98953275	-0,0539549	-0,1079098
-0,0539549	8,00366893	-3,9926621
-0,1079098	-3,9926621	2,01467573

Значение $X^T y$

80
290
580

И тогда $w = (X^T X + \alpha I)^{-1} X^T y$

0,92802417
1,00356102
2,00712205

Таким образом уравнение регрессии имеет вид

$$\hat{y} = 0,9 \times 1 + 1 \times x_1 + 2 \times x_2 = 0,9 + x_1 + 2 \times x_2$$

Однако заметив, что $x_2 = 2 \times x_1$, можете сделать следующее преобразование

$$\hat{y} = 0,9 \times x_1 + 2 \times x_2 = 0,9 + x_1 + 2 \times 2 \times x_1 = 0,9 + 5 x_1$$