

Практическое задание к видео 4

Закрепите знания по алгоритму поиска коэффициентов регрессии в матричном виде с помощью метода наименьших квадратов для набора данных с имеющейся линейной зависимостью. Обратите внимание: это задание для самопроверки, его не нужно сдавать куратору.

После выполнения задания рекомендуем свериться с ответами ниже.

Задача

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

X_1	X_2	Y
1	2	13
3	6	19
5	10	25
7	14	31
9	18	37

Искомая функция выглядит следующим образом:

$$y^- = w_0 + w_1 * x_1 + w_2 * x_2$$

Найдите коэффициенты искомой функции: w_0 и w_1 и w_2 .

Для этого:

- определите матрицу X ,
- транспонируйте матрицу X ,
- перемножьте матрицы X^T и X ,
- найдите определитель матрицы $X^T X$,
- дополните матрицу $X^T X + \alpha I$,
- найдите определитель матрицы $X^T X + \alpha I$,
- найдите обратную матрицу $(X^T X + \alpha I)^{-1}$,
- определите значение $X^T y$,

- определите вектор весов w ,
- запишите уравнение полученной регрессии.

Ответы для проверки

Можно заметить, что уравнение регрессии имеет вид:

$$y = 10 + x_1 + 2 * x_2 = 10 + 3 * x_1$$

Найдём значения коэффициентов с помощью метода наименьших квадратов.

Дополним фиктивную переменную x_0 . Это и есть матрица X :

X_0	X_1	X_2
1	1	2
1	3	6
1	5	10
1	7	14
1	9	18

Тогда транспонированная матрица X^T :

1	1	1	1	1
1	3	5	7	9
2	6	10	14	18

Тогда матрица $X^T X$:

5	25	50
25	165	330
50	330	660

Определитель этой матрицы равен 0.

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

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	25	50
25	165,1	330
50	330	660,1

Определитель этой матрицы равен 108,3. Продолжаем решать.

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

0,76185815	-0,0230838	-0,0461676
-0,0230838	8,00094182	-3,9981164
-0,0461676	-3,9981164	2,00376728

Значение $X^T y$:

125
745
1490

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

9,24506699
0,62280127
1,24560253

Таким образом, уравнение регрессии имеет вид:

$$y = 9,24 * 1 + 0,62 * x_1 + 1,24 * x_2 = 9,24 + 0,62 * x_1 + 1,24 * x_2$$

Заметив, что $x_2 = 2 * x_1$, можем сделать следующее преобразование:

$$y = 9,24 + 0,62 * x_1 + 1,24 * x_2 = 9,24 + 0,62 * x_1 + 1,24 * 2 * x_1 = 9,24 + 3,1 * x_1$$

Ранее мы установили, что уравнение выглядит так: $y = 10 + x_1 + 2 * x_2 = 10 + 3 * x_1$

Мы нашли искомый вектор w , но отметим, что оценки получились смещёнными.