

Линейные методы регрессии II

Факультатив «Введение в анализ данных и машинное обучение на Python»

14 декабря 2019 г.

Задание

Пусть представлены следующие данные:

$$Y = \begin{pmatrix} 2 \\ 4 \\ 8 \end{pmatrix}, X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{pmatrix}$$

- а) Оценивается следующая модель:

$$Y = X\beta + u,$$

где u – случайная ошибка.

Найдите оценки коэффициентов модели. Найдите прогноз модели.

- б) Нанесите выборку на диаграмму рассеяния. Постройте оценённую линию регрессии.
в) Определите направление корреляции между X и Y .
г) Найдите выборочное среднее и выборочную дисперсию неконстантного признака.
д) Найдите коэффициент детерминации данной регрессии.

Решение

- а)

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T Y \\ X^T X &= \begin{pmatrix} 3 & 5 \\ 5 & 9 \end{pmatrix} \\ (X^T X)^{-1} &= \frac{1}{2} \begin{pmatrix} 9 & -5 \\ -5 & 3 \end{pmatrix} \\ \hat{\beta} &= \frac{1}{2} \begin{pmatrix} 9 & -5 \\ -5 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \\ 8 \end{pmatrix} = \begin{pmatrix} -2 \\ 4 \end{pmatrix} \\ \hat{Y} &= X\hat{\beta} = \begin{pmatrix} 2 \\ 6 \\ 6 \end{pmatrix}\end{aligned}$$

- б) Строим в координатах (X_1, Y) , где X_1 – второй столбец матрицы X . Уравнение оценённой линии:
 $\hat{Y} = -2 + 4X_1$. Можно построить и без уравнения, последовательно нанося точки (X_1, \hat{Y}) .

с) Положительное.

d)

$$\bar{X} = \frac{1 + 2 + 2}{3} = \frac{5}{3}$$

$$s \text{Var}(X) = (1 - \frac{5}{3})^2 + (2 - \frac{5}{3})^2 + (2 - \frac{5}{3})^2.$$

$$\text{e) } R^2 = 1 - \frac{\|Y - \hat{Y}\|^2}{\|Y - \bar{Y}\|^2} = 1 - \frac{0^2 + (-2)^2 + 2^2}{(2 - 14/3)^2 + (4 - 14/3)^2 + (8 - 14/3)^2} = 0.57.$$