

Регрессионный анализ 2.

- 1 Рассмотрим линейную модель $Y = \theta + \varepsilon$, где $E\varepsilon = 0$. Докажите, что значение θ , при котором величина $\|Y - \theta\|^2$ минимизируется при ограничении $A\theta = 0$, где A – известная матрица размера $q \times n$ ранга q , $q < n$, равно

$$\hat{\theta} = (I_n - A^T(AA^T)^{-1}A)Y.$$

- 2 Дана модель линейной регрессии $y_i = a + bx_i + \varepsilon_i$, $i = 1, \dots, n$, где $\theta = (a, b)$ – неизвестный параметр. Пусть

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n |y_i - a - bx_i|.$$

Найдите оценку \hat{a} через \hat{b} . Будет ли такой метод поиска оценки неизвестного параметра давать единственное решение? Можно полагать, что все x_i положительны.

- 3 Выданы данные $\{(y_i, x_{ij}), i = 1, \dots, n+q, j = 1, \dots, k\}$, причем y_{n+1}, \dots, y_{n+q} неизвестны. Используя пройденные методы регрессионного анализа, в рамках линейной регрессионной модели произвести отбор признаков по первым n объектам и предсказать значения откликов объектов с номерами $n+1, \dots, n+q$. Описать и объяснить проделанные процедуры.
- 4 Выданы данные $\{(y_i, x_{ij}), i = 1, \dots, n+q, j = 1, \dots, k\}$, причем y_{n+1}, \dots, y_{n+q} неизвестны. Используя пройденные методы регрессионного анализа, в рамках линейной регрессионной модели произвести отбор признаков по первым n объектам и предсказать значения откликов объектов с номерами $n+1, \dots, n+q$. Описать и объяснить проделанные процедуры.
- 5 Пусть вектор ошибок в линейной регрессионной модели $Y = X\theta + \varepsilon$ состоит из независимых компонент, имеющих распределение Лапласа с параметром σ^2 . Найти фактический уровень значимости критерия Голдфелда-Квандта в данной модели в зависимости от n . Как часто будет отвергаться гипотеза гомоскедастичности, если распределение i -той компоненты вектора ошибок – $Laplace(\sigma^2(1 + i/n))$?
- 6 Выданы данные $\{(y_i, x_{ij}), i = 1, \dots, n+1, j = 1, \dots, k\}$, причем y_{n+1} известен. Используя стандартную и робастные регрессионные модели, произвести отбор признаков по первым n объектам и предсказать значение $n+1$ -го отклика. Какая модель дала наилучший результат и почему? Можно ли пользоваться стандартными процедурами, такими, как критерий Фишера и информационные критерии, для отбора признаков в робастных моделях?