



Data learning

Курс “Машинное обучение”
Лабораторная работа



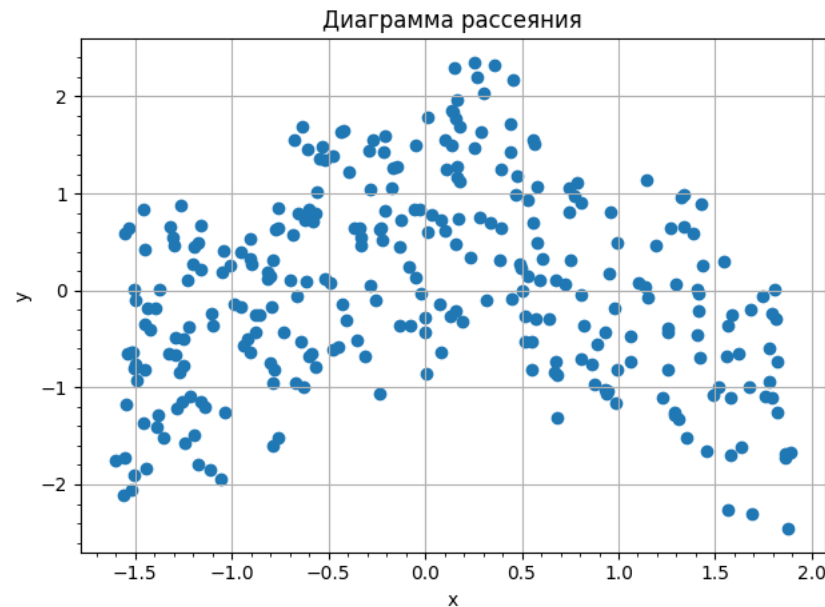
Ridge regression

Шаров К.И., М18-504
Вариант 1-04

Исходные данные

Представляют собой пары
свободной переменной — x и
зависимой переменной — y
Общее количество пар — 300

Кросс-валидация методом Holdout
(70/30).



Используемые методы и формулы

Класс регрессионных моделей: $h(x) = \sum_{i=1}^m \beta_i x^i$

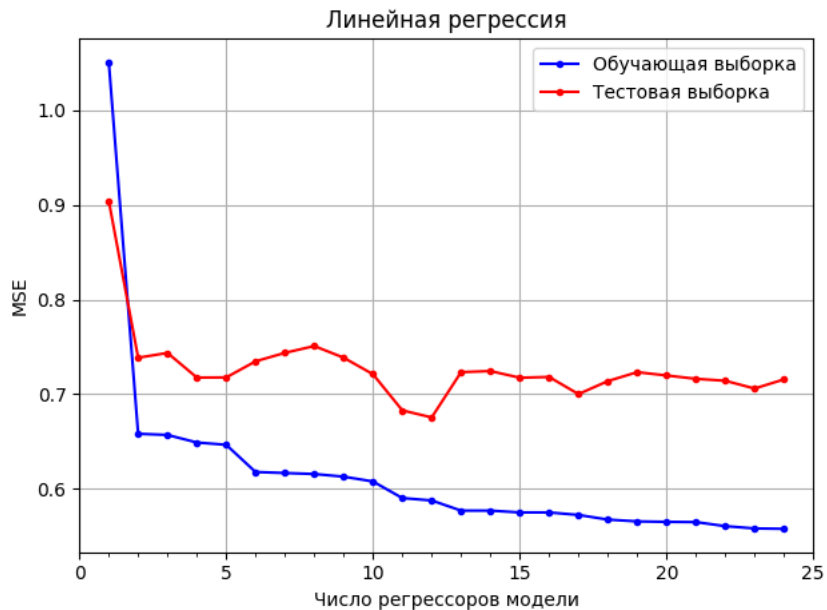
МНК: $Q(\beta) = \|F\beta - y\|^2$
 $\beta^* = \arg \min (Q(\beta))$, где $F = \begin{bmatrix} x_1 & \cdots & x_n \\ \vdots & \ddots & \vdots \\ x_1^m & \cdots & x_n^m \end{bmatrix}$, $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

Решение: $\beta^* = (F^T F)^{-1} F^T y$

Гребневая регрессия: $Q_\lambda(\beta) = \|F\beta - y\|^2 + \lambda \|\beta\|^2$
 $\beta^* = (F^T F + \lambda I_n)^{-1} F^T y.$

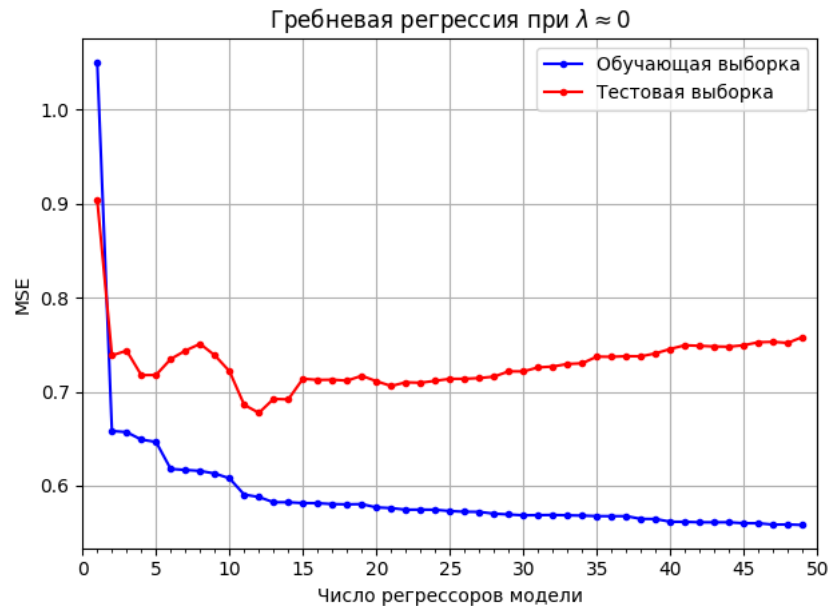
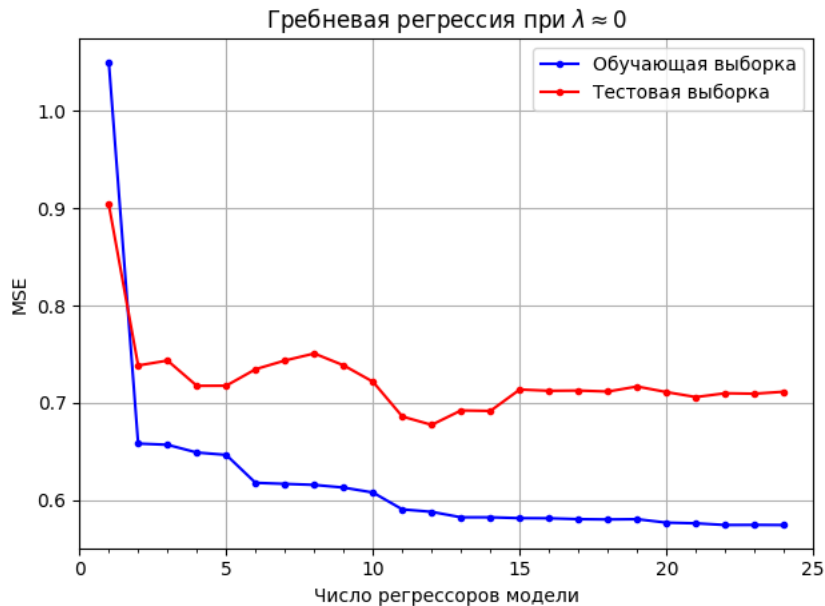
Результаты исследований

Зависимость MSE от числа регрессоров линейной регрессионной модели



Результаты исследований

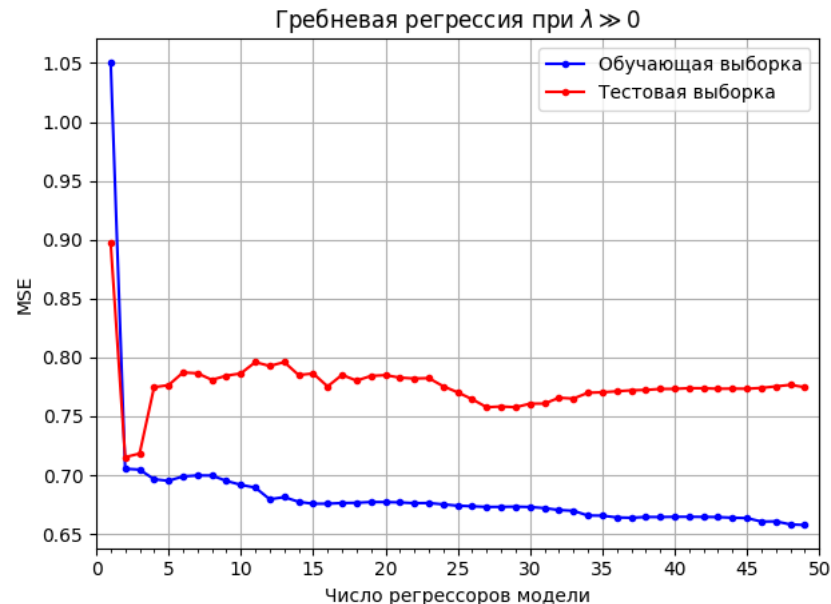
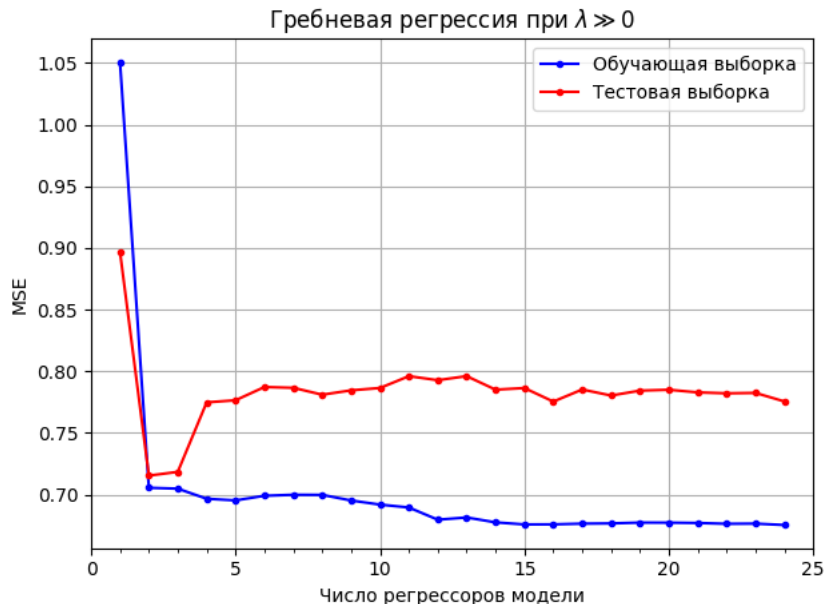
Зависимость MSE от числа регрессоров гребневой регрессионной модели при $\lambda \approx 0$



$\lambda = 0.001$

Результаты исследований

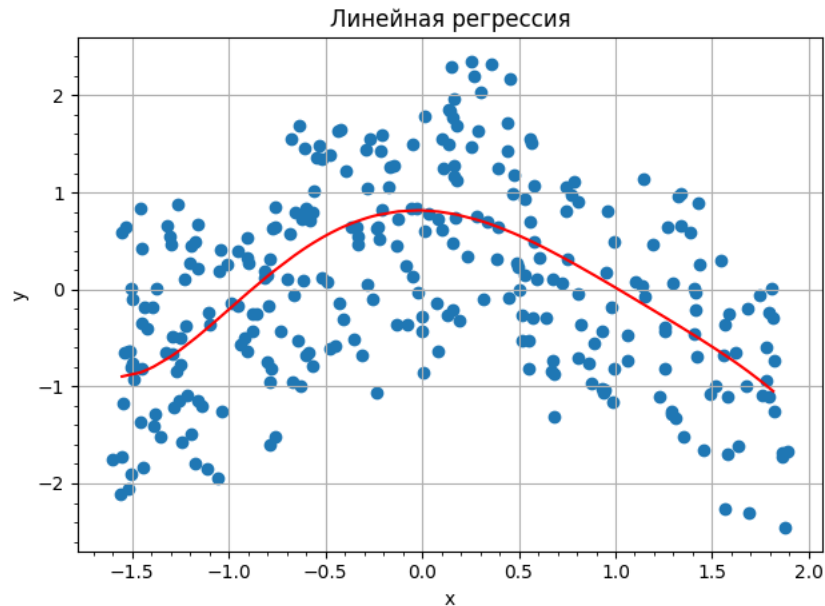
Зависимость MSE от числа регрессоров гребневой регрессионной модели при $\lambda \gg 0$



$\lambda = 100$

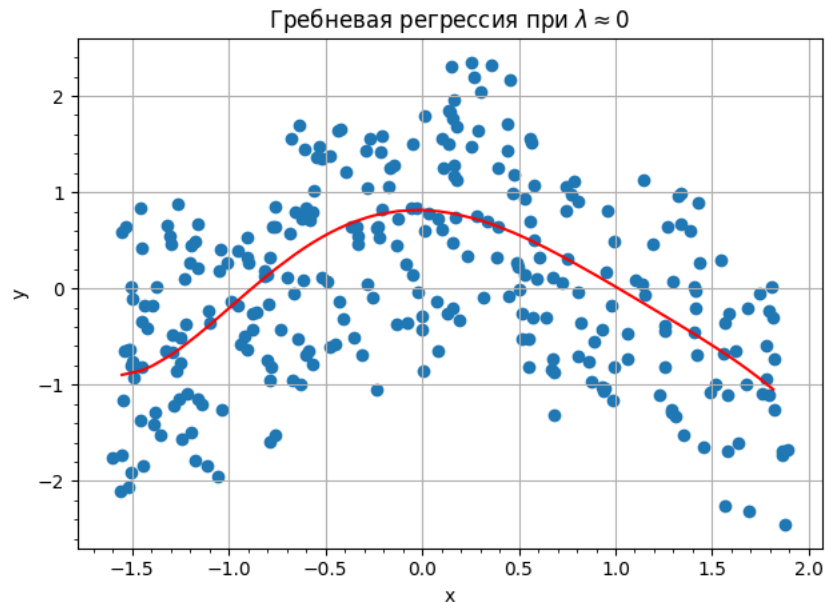
Результаты исследований

График зависимости выхода линейной регрессионной модели от x , наложенный на диаграмму рассеяния ($m = 5$)



Результаты исследований

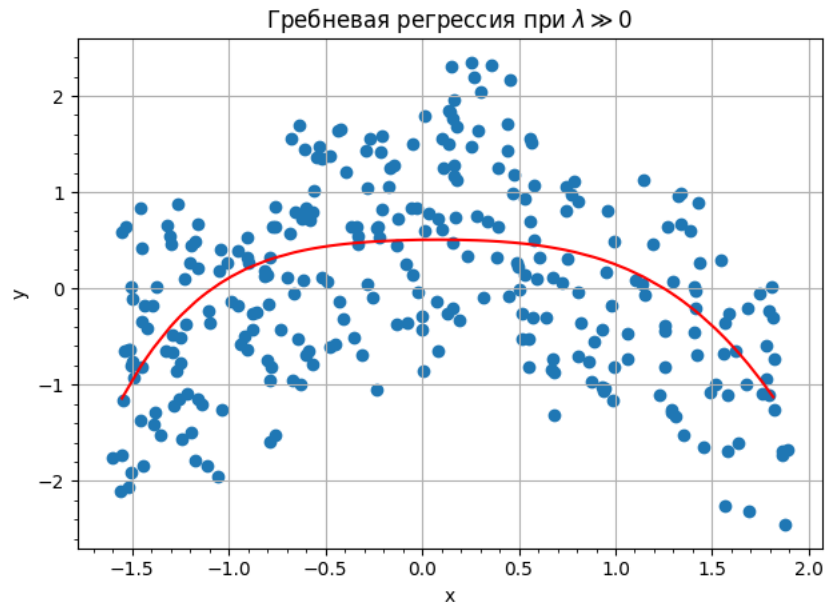
График зависимости выхода гребневой регрессионной модели $\lambda \approx 0$ от x , наложенный на диаграмму рассеяния ($m = 5$)



$$\lambda = 0.001$$

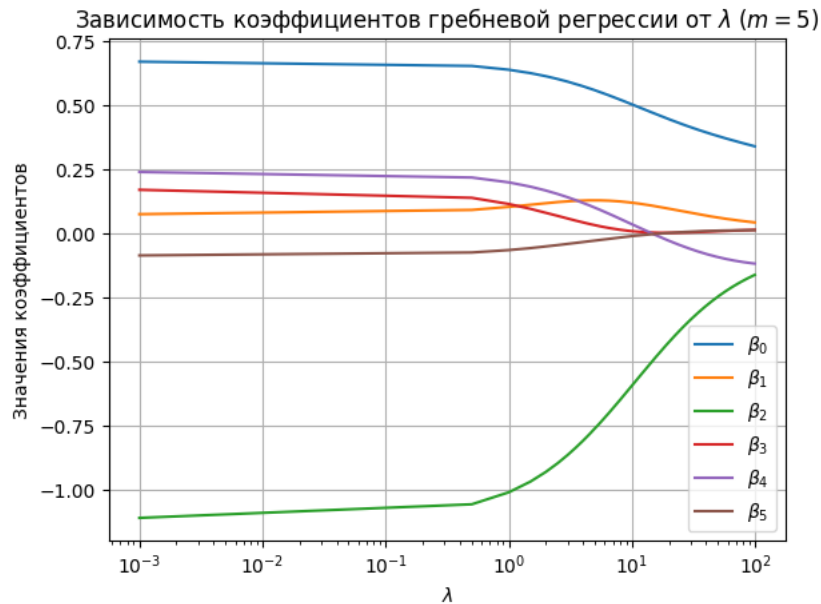
Результаты исследований

График зависимости выхода гребневой регрессионной модели $\lambda \gg 0$ от x , наложенный на диаграмму рассеяния ($m = 5$)

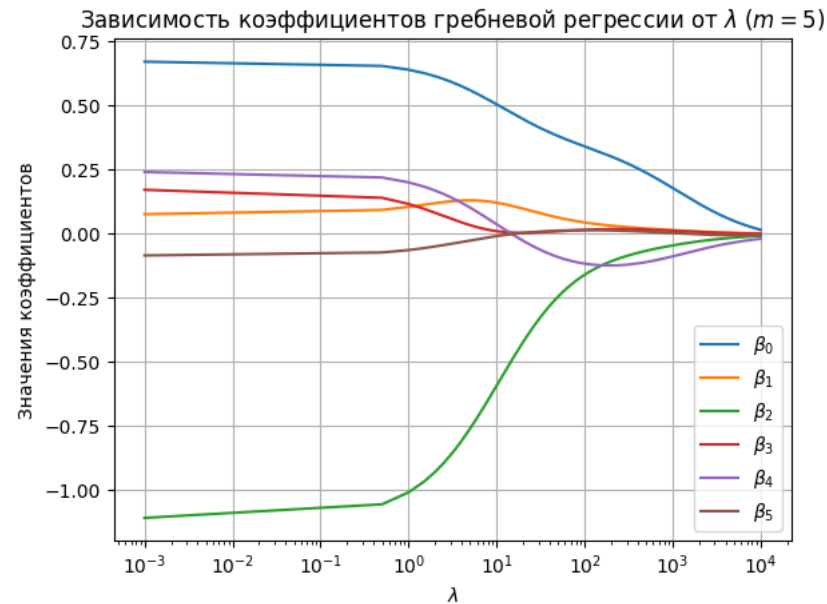


$$\lambda = 100$$

Результаты исследований



$$\lambda = \overline{0.001, 100}$$



$$\lambda = \overline{0.001, 10000}$$

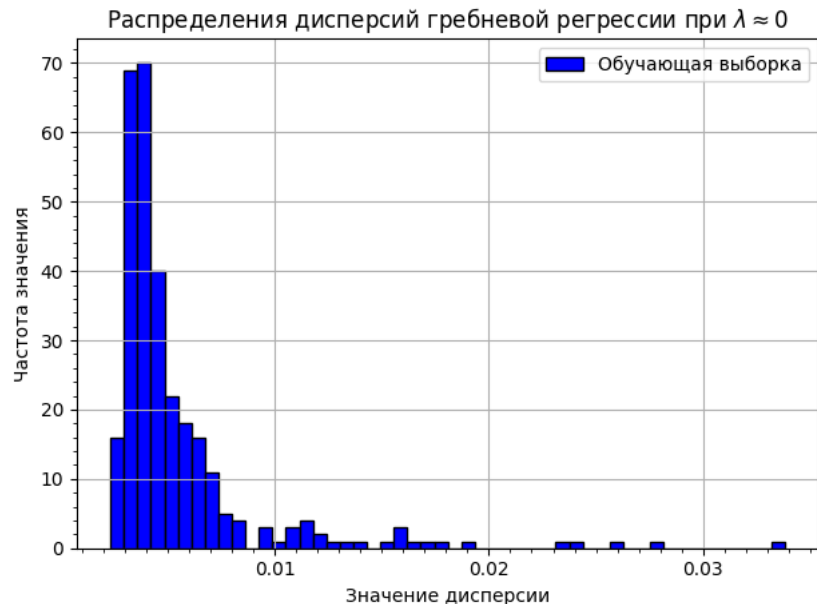
Результаты исследований

Гистограммы распределения дисперсий отклика модели для обучающей и тестовой выборок при линейной регрессии ($m = 5$)



Результаты исследований

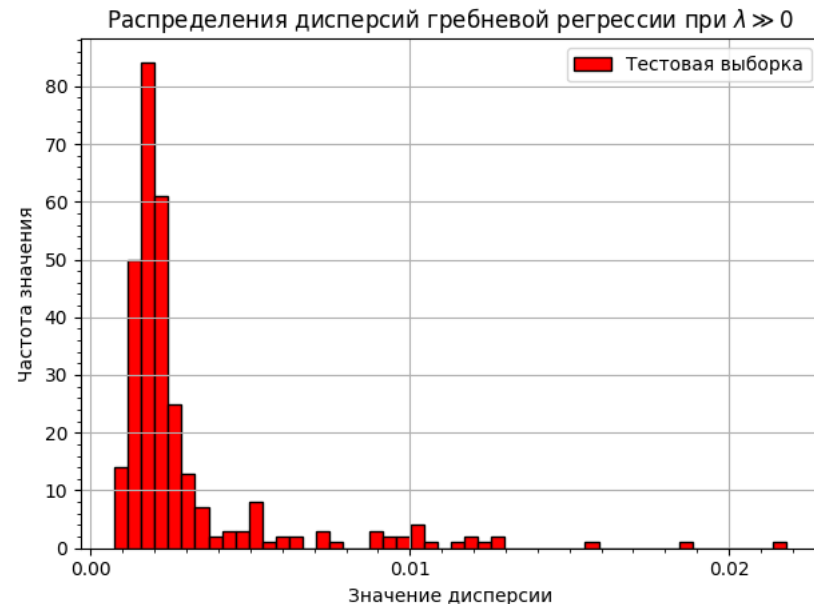
Гистограммы распределения дисперсий отклика модели для обучающей и тестовой выборок при гребневой регрессии $\lambda \approx 0$ ($m = 5$)



$$\lambda = 0.001$$

Результаты исследований

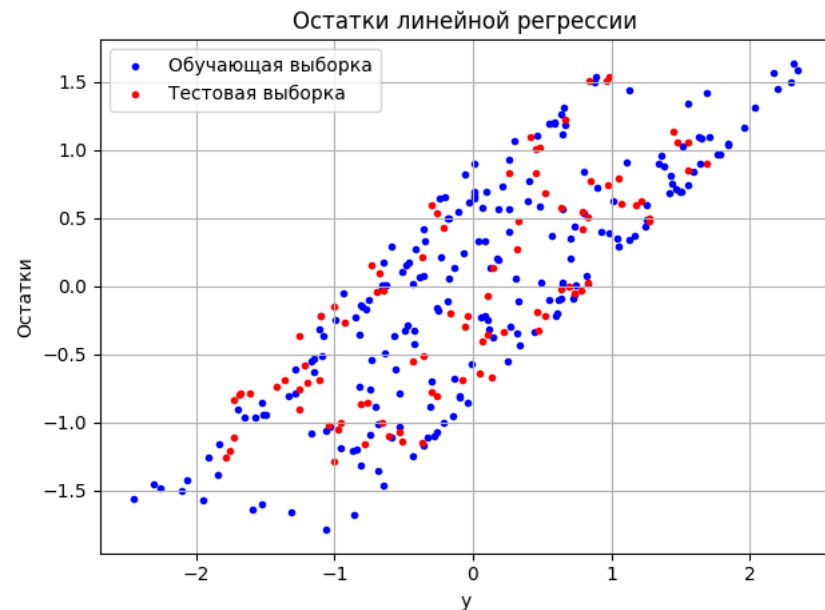
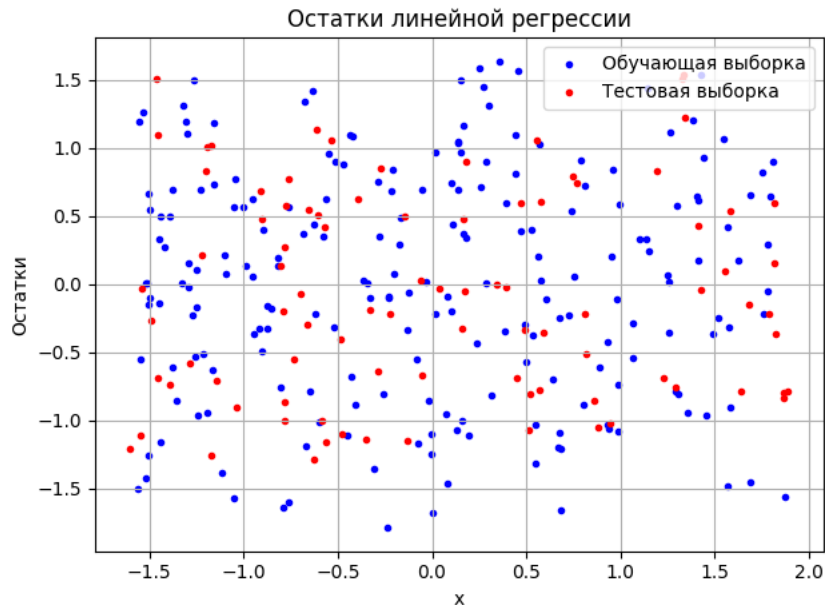
Гистограммы распределения дисперсий отклика модели для обучающей и тестовой выборок при гребневой регрессии $\lambda \gg 0$ ($m = 5$)



$$\lambda = 100$$

Результаты исследований

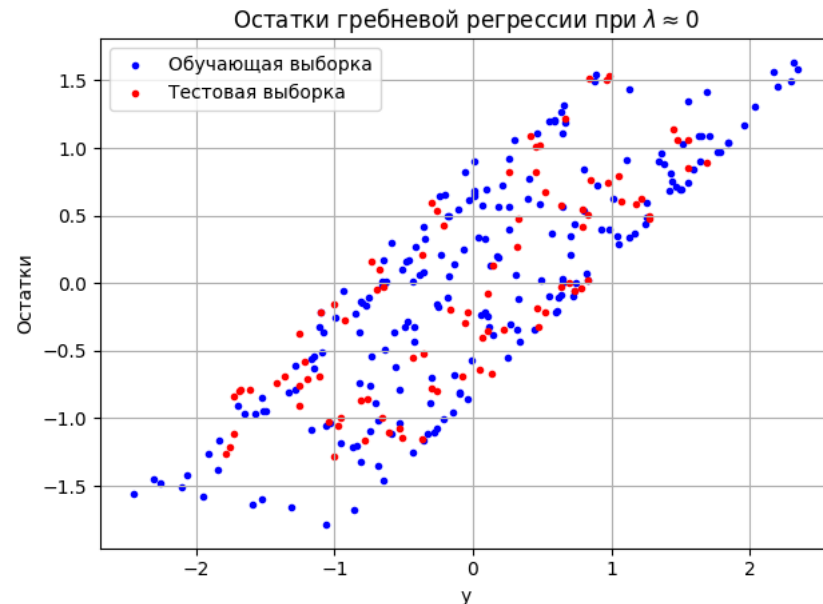
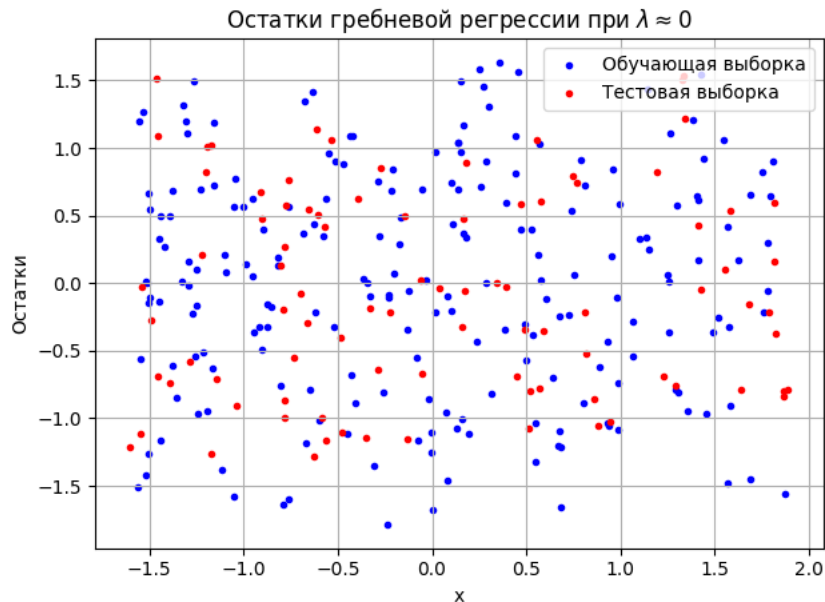
Графики зависимости остатков модели для линейной регрессии ($m = 5$)



Математическое ожидание остатков регрессионной модели: -0.0746

Результаты исследований

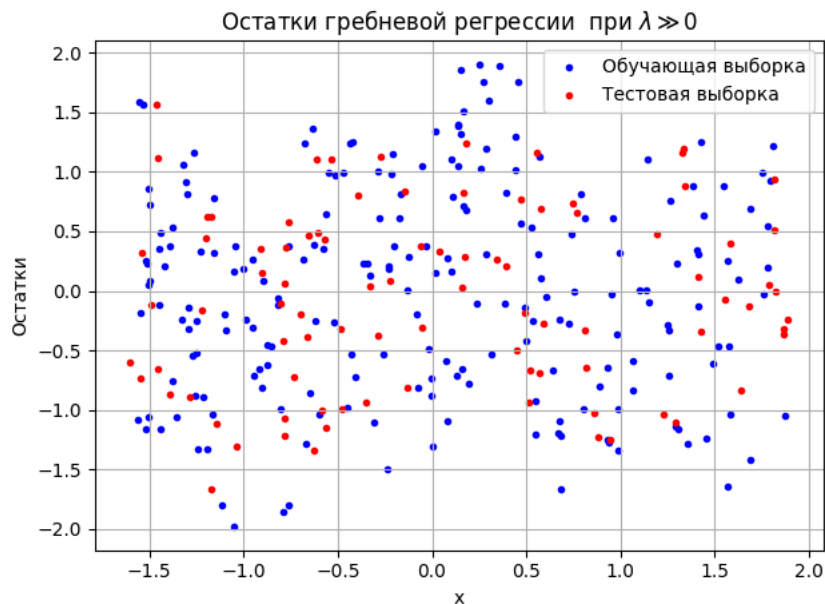
Графики зависимости остатков модели для гребневой регрессии $\lambda \approx 0$ ($m = 5$; $\lambda = 0.001$)



Математическое ожидание остатков регрессионной модели: -0.0746

Результаты исследований

Графики зависимости остатков модели для гребневой регрессии $\lambda \gg 0$ ($m = 5$; $\lambda = 100$)



Математическое ожидание остатков регрессионной модели: -0.063

Выводы

- Использование параметра λ позволяет сгладить выбросы MSE, проявляющиеся при использовании моделей с большим числом регрессоров ($m \geq 30$).
- При малых значениях λ ($\lambda \approx 0$) графики выходов линейной и гребневой регрессионных моделей не отличаются существенно; при больших значениях λ ($\lambda \gg 0$) график выходов гребневой регрессии приближается к прямой.
- С ростом λ коэффициенты моделей гребневой регрессии стремятся к нулю.

Выводы

- Исходя из анализа гистограмм распределения дисперсий отклика моделей можно сделать вывод, что дисперсия откликов не зависит от входа модели.
- Исходя из визуального анализа графиков остатков регрессионных моделей можно сделать следующие выводы:
 - Математическое ожидание остатков всех моделей приблизительно равно нулю;
 - Дисперсия остатков всех моделей постоянна;
 - На графиках всех моделей отсутствуют неслучайные паттерны.