Практика №1. Вариант 8

Андрей Владимиров

Февраль 28 2021

# Задача №1

Дана функция: .

Необходимо построить три графика:

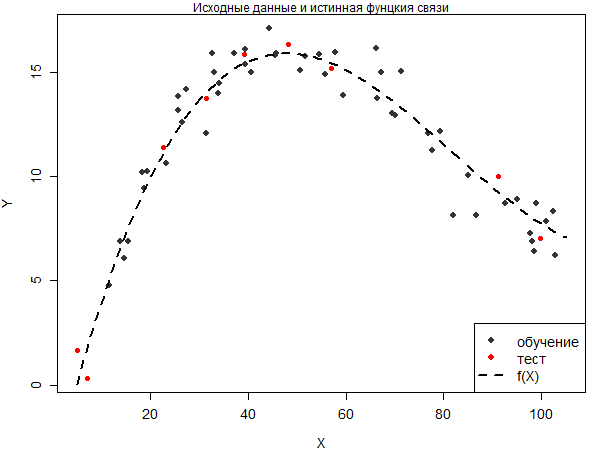
* график истинной взаимосвязи;
* график зависимости MSE от гибкости модели;
* график лучшей модели (компромисс между гибкостью и точностью.

При этом необходимо выбрать число степеней свободы как компромисс между точностью (оценкой ошибки на тестовой выборке) и простотой модели (числом степеней свободы).

Распределение X и Y:

* ;
* , где .

Изобразим исходные данные на графике:

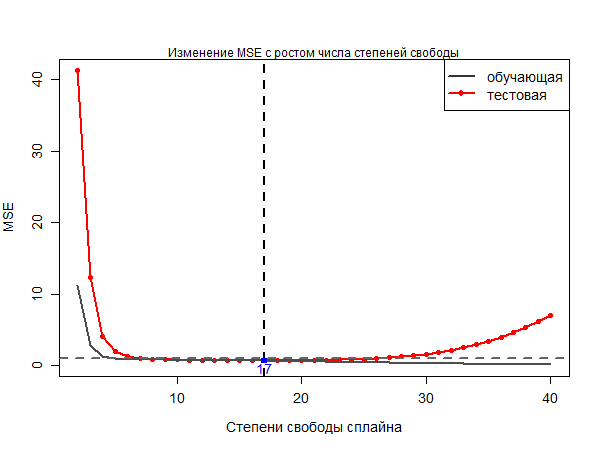


В качестве модели используем сплайны со степенями свободы от 2 (прямая) до 40 (количесество узлов равно 2/3 наблюдений). Строим модели с различным количеством степеней свободы, и в каждом случае считаем среднеквадратическую ошибку модели на обучающей и тестовой выборках.

Выведем первые десять строчек таблицы с моделями.

|  |  |  |
| --- | --- | --- |
| df | MSE.train | MSE.test |
| 2 | 11.2647447 | 41.2593342 |
| 3 | 2.7422900 | 12.2559954 |
| 4 | 1.2861904 | 4.1140168 |
| 5 | 0.9980967 | 1.9525987 |
| 6 | 0.8942784 | 1.2299809 |
| 7 | 0.8406261 | 0.9542322 |
| 8 | 0.8091578 | 0.8448843 |
| 9 | 0.7879560 | 0.8003990 |
| 10 | 0.7716512 | 0.7798648 |
| 11 | 0.7572892 | 0.7648446 |

Изобразим на графике поведение ошибок при различном количестве степеней свободы.

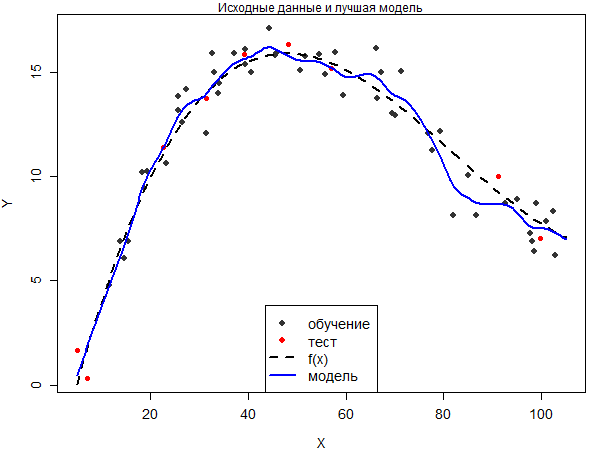


На этом графике:

* При движении слева направо MSE на обучающей выборке (серая кривая) сокращается, так как с ростом числа степеней свободы растет число узлов, по которым строится сплайн. При этом модельная кривая подгоняется по все возрастающему количеству точек и становится все более гибкой. В результате индивидуальные расстояния от фактических наблюдений за до их модельных оценок сокращаются, что приводит к сокращению MSE.
* При движении слева направо, MSE на тестовой выборке (красная кривая) сначала резко сокращается, затем растет. Нам известна истинная форма связи с , она описывается кубической функцией. Число степеней свободы такой модели равно числу оцениваемых параметров, то есть четырем (коэффициенты перед , , и константа). Поэтому резкое падение ошибки на тестовой выборке при небольшом числе степеней свободы связано с тем, что модель приближается по гибкости к истинной функции связи. Затем MSE на тестовой выборке довольно долго остается стабильной, а после, начинает расти. Этот рост объясняется эффектом переобучения модели: она все лучше описывает обучающую выборку, и при этом постепенно становится неприменимой ни к одному другому набору наблюдений.

Наименьшее значение MSE на тестовой выборке соответствует числу степеней свободы 18, и равно 0.6703458 Визуально по графику мы можем установить, что первое значение , близкое к стабильно низким, соответсвует . Ошибка здесь меньше единицы, что ненамного отличается от минимума. Именно было выбрано в качестве компромисса между точность. (минимальное MSE на тестовой выборке) и простотой модели (чем меньше степеней свободы, тем модель проще).

График с моделью, выбранной в качестве лучшей, показан на рисунке ниже.

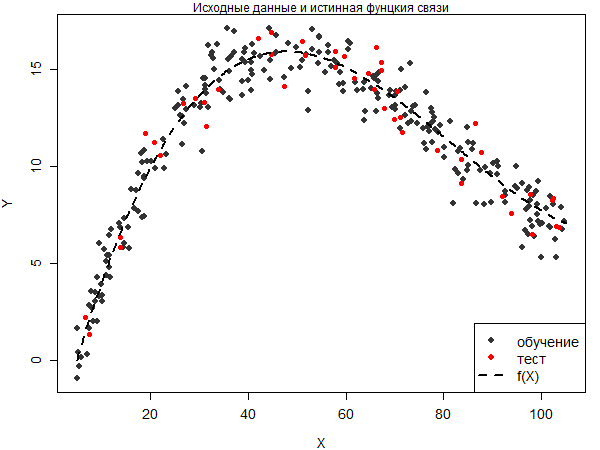


# Задача №2

Необходимо проделать действия из “Задачи №1” с условием, что в трех испытаниях , и соответсвенно.

## 1)

Изобразим исходные данные на графике:



Выведем первые десять строчек таблицы с моделями.

|  |  |  |
| --- | --- | --- |
| df | MSE.train | MSE.test |
| 2 | 15.9973926 | 14.525996 |
| 3 | 3.4553506 | 2.883003 |
| 4 | 1.4049066 | 1.341230 |
| 5 | 1.0888700 | 1.124775 |
| 6 | 1.0153207 | 1.076553 |
| 7 | 0.9905200 | 1.061455 |
| 8 | 0.9795270 | 1.058075 |
| 9 | 0.9728086 | 1.060827 |
| 10 | 0.9674408 | 1.067073 |
| 11 | 0.9625035 | 1.075128 |

Изобразим на графике поведение ошибок при различном количестве степеней свободы.

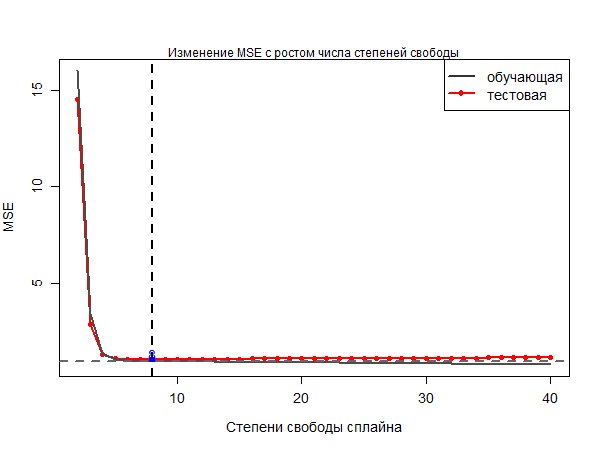
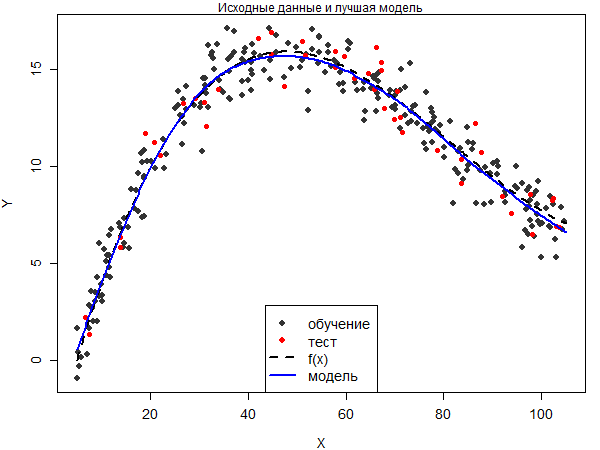
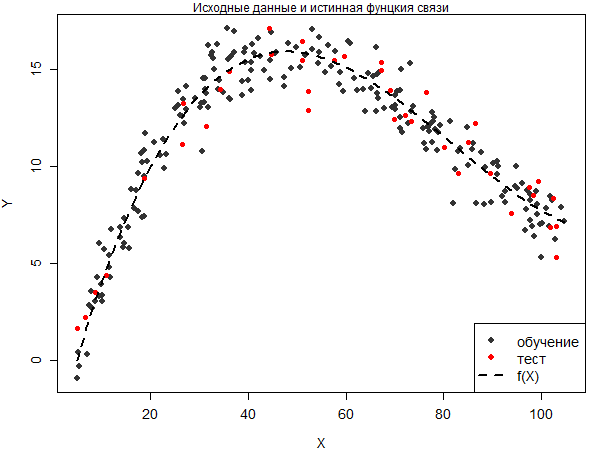


График с моделью, выбранной в качестве лучшей, показан на рисунке ниже.



## 2)

Изобразим исходные данные на графике:



Выведем первые десять строчек таблицы с моделями.

|  |  |  |
| --- | --- | --- |
| df | MSE.train | MSE.test |
| 2 | 14.9531846 | 15.943072 |
| 3 | 3.3963607 | 2.878555 |
| 4 | 1.4320110 | 1.257532 |
| 5 | 1.1131665 | 1.150162 |
| 6 | 1.0351199 | 1.178469 |
| 7 | 1.0075928 | 1.215568 |
| 8 | 0.9954539 | 1.247952 |
| 9 | 0.9884621 | 1.275100 |
| 10 | 0.9829645 | 1.298973 |
| 11 | 0.9777575 | 1.320926 |

Изобразим на графике поведение ошибок при различном количестве степеней свободы.

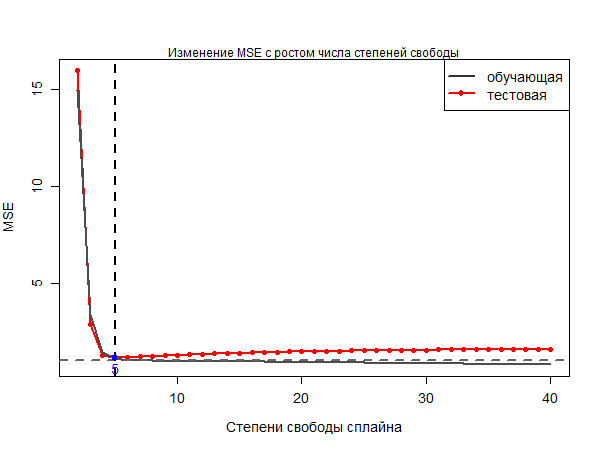
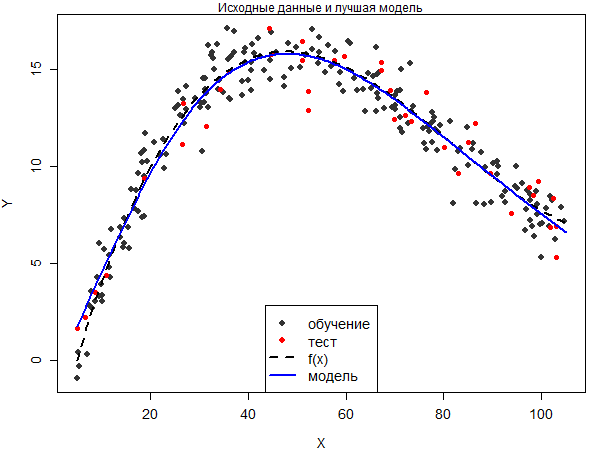
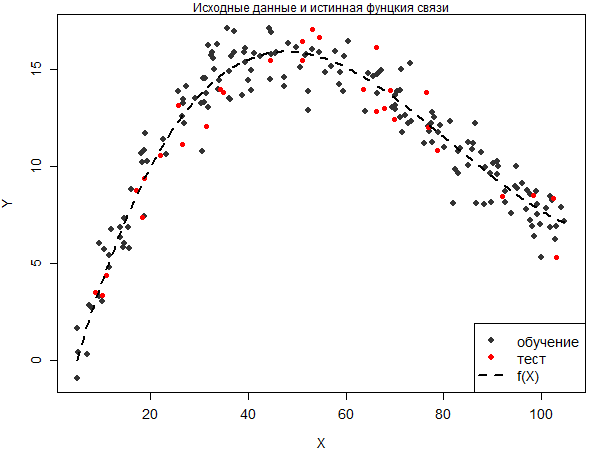


График с моделью, выбранной в качестве лучшей, показан на рисунке ниже.



## 3)

Изобразим исходные данные на графике:



Выведем первые десять строчек таблицы с моделями.

|  |  |  |
| --- | --- | --- |
| df | MSE.train | MSE.test |
| 2 | 12.366281 | 12.274372 |
| 3 | 3.536769 | 3.033219 |
| 4 | 1.582853 | 1.264379 |
| 5 | 1.245968 | 1.111070 |
| 6 | 1.157404 | 1.152109 |
| 7 | 1.125583 | 1.206044 |
| 8 | 1.111796 | 1.248777 |
| 9 | 1.104096 | 1.280820 |
| 10 | 1.098203 | 1.305955 |
| 11 | 1.092659 | 1.326629 |

Изобразим на графике поведение ошибок при различном количестве степеней свободы.

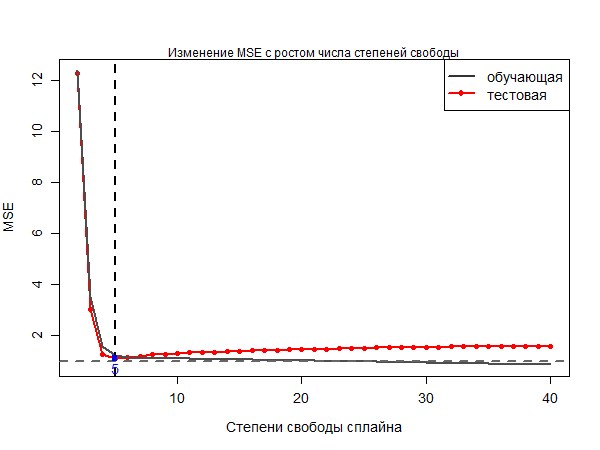
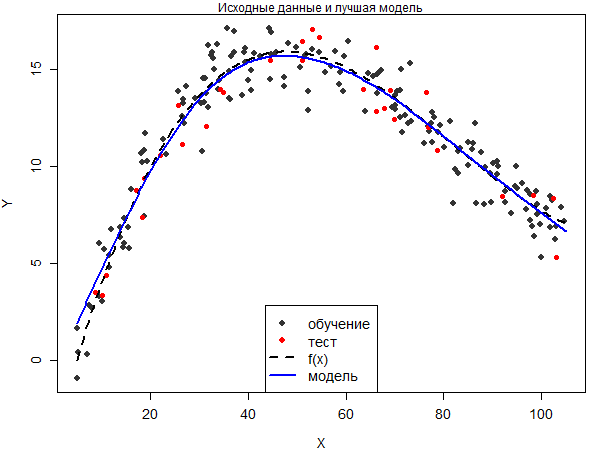


График с моделью, выбранной в качестве лучшей, показан на рисунке ниже.



# Вывод:

Выше описаны аналогичные действия и графики как в “Задачи №1”. Как видно из графиков, при увеличении n (количества наблюдений) в общем плане уменьшаеться при увеличении степеней свободы. В “Задаче №1” стремительно начинает расти при ().Также при увеличении количества наблюдений растет и среднее значение при обучении. Так минимальное значение достигается при меньших степенях свободы, но само значение выше: при , а ; при , ; при , , - в “Задачи №1” минимальное значение равно 0.6703458 при .