Вариант 24

Упражнение 1 2021

Спиридонова Алина

18 02 2021

# Данные

Функция для задачи 1:

Характеристики для задачи 2:

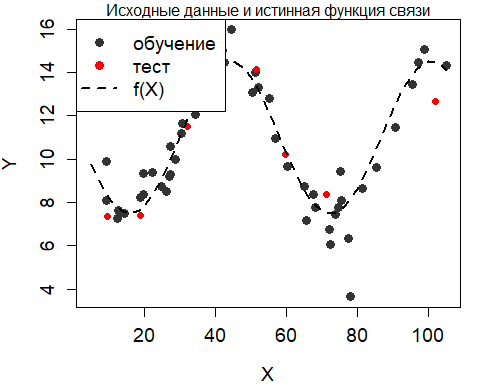
# Задача 1.

На данных своего варианта повторить три графика из первой практики, выбрав число степеней свободы как компромисс между точностью (оценкой ошибки на тестовой выборке) и простотой модели (числом степеней свободы). Все рисунки сохранить в графические файлы в формате png.

Сгенерируем X и Y линейной модели и изобразим исходные данные на графике.

#### Генерация данных

###### График 1: Исходные данные на график



Для начала строим модель №2 из лекции

Далее в качестве модели используем сплайны со степенями свободы от 2 до 40.

строим модели с различным количеством степеней свободы на обучающей и тестовой выборках.

Получаем таблицу вида:

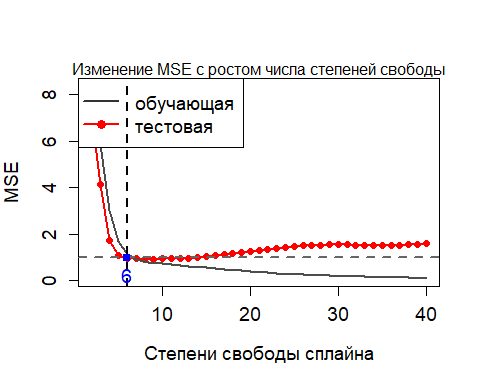
## train test   
## 1.15 0.97

|  |  |  |
| --- | --- | --- |
| df | MSE.train | MSE.test |
| 2 | 8.3671508 | 6.8731839 |
| 3 | 5.8760457 | 4.1319492 |
| 4 | 3.0286712 | 1.7364090 |
| 5 | 1.6844408 | 1.0873055 |
| 6 | 1.1519790 | 0.9711759 |
| 7 | 0.9302515 | 0.9473308 |
| 8 | 0.8283730 | 0.9424579 |
| 9 | 0.7713435 | 0.9420935 |
| 10 | 0.7309576 | 0.9447203 |
| 11 | 0.6967680 | 0.9521146 |

Расчёт ошибки на обучающей выборке:

Расчёт ошибки на тестовой выборке:

###### График 2: Зависимость MSE от гибкости модели

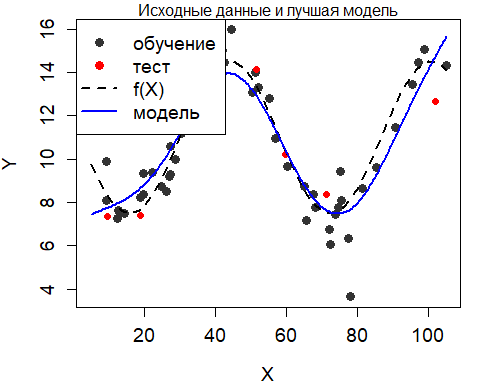


На графике видно что MSE на обучающей выборке (серая) уменьшается сначала резко, затем плавнее

На графике видно что MSEТЕСТ (красная) сначала резко уменьшается, затем плавно растёт.

Наименьшее значение MSEТЕСТ соответствует числу степеней свободы df = 9 равно 0.9420935.По графику можно примерно сказать, что первое значение MSEТЕСТ df = 6. Ошибка ненамного отличается от минимума.  
df = 6 было выбрано в качестве компромисса между точностью (минимальной MSEТЕСТ) и простотой модели (чем меньше степеней свободы, тем модель проще).

###### График 3: Лучшая модель (компромисс между гибкостью и точностью)

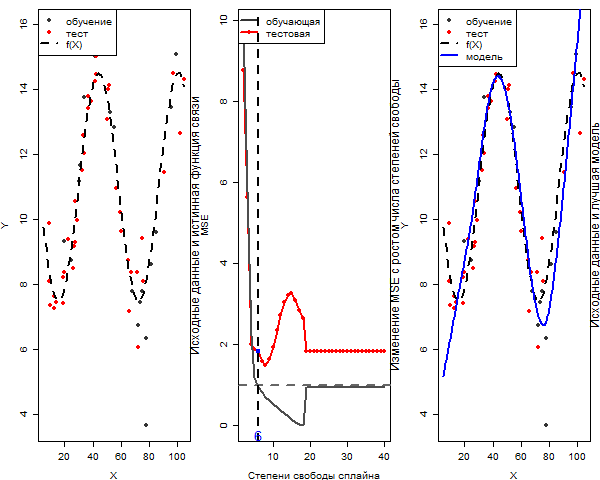


# Задача 2.

*Задание: Решить задачу 1, изменив характеристики данных. Почему при таком изменении данных MSE меняется именно так? Все рисунки сохранить в графические файлы в формате png.*

|  |  |  |
| --- | --- | --- |
| df | MSE.train | MSE.test |
| 2 | 9.8642201 | 8.752888 |
| 3 | 5.7870010 | 5.627139 |
| 4 | 2.2386751 | 2.005324 |
| 5 | 1.1916367 | 1.871586 |
| 6 | 0.9429532 | 1.828994 |
| 7 | 0.8166672 | 1.598766 |
| 8 | 0.7131838 | 1.498982 |
| 9 | 0.6207368 | 1.629121 |
| 10 | 0.5367514 | 1.943016 |
| 11 | 0.4587840 | 2.340599 |

## png   
## 2

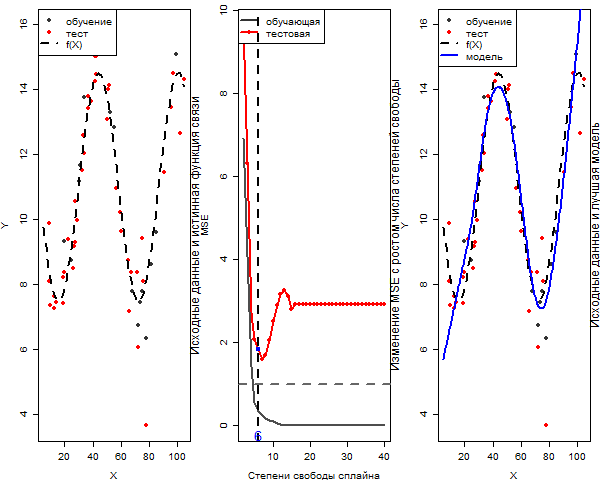


Наименьшее значение MSE на тестовой выборке соответствует числу степеней свободы 9 и равно 1.570066. По графику мы можем установить, что первое значение MSEТЕСТ df = 6.

df = 6 было выбрано в качестве компромисса между точностью (минимальной MSE на тестовой выборке) и простотой модели (чем меньше степеней свободы, тем модель проще).

|  |  |  |
| --- | --- | --- |
| df | MSE.train | MSE.test |
| 2 | 6.9085177 | 9.639307 |
| 3 | 4.0805883 | 6.316725 |
| 4 | 1.4268115 | 2.691400 |
| 5 | 0.5621387 | 2.079397 |
| 6 | 0.3480331 | 1.838944 |
| 7 | 0.2466432 | 1.603365 |
| 8 | 0.1776104 | 1.697176 |
| 9 | 0.1313957 | 2.067741 |
| 10 | 0.0971674 | 2.509176 |
| 11 | 0.0649128 | 2.888099 |

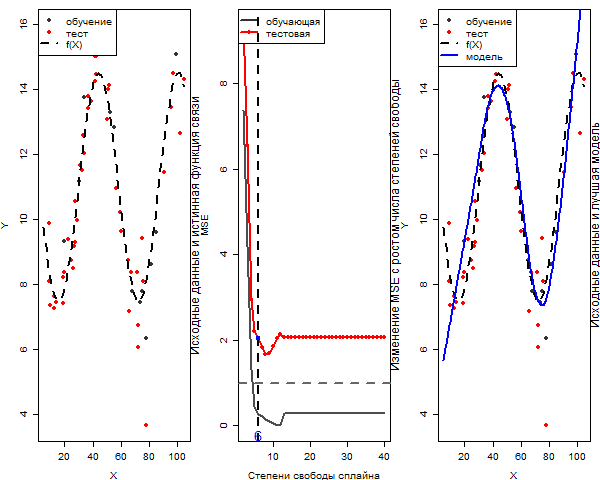
## png   
## 2



Наименьшее значение MSE на тестовой выборке df = 8 и равно 1.845563. По графику можно заметить, что первое значение MSEТЕСТ df = 7.

|  |  |  |
| --- | --- | --- |
| df | MSE.train | MSE.test |
| 2 | 7.3655707 | 9.361782 |
| 3 | 4.1859381 | 6.555726 |
| 4 | 1.3161194 | 2.961928 |
| 5 | 0.4492288 | 2.214202 |
| 6 | 0.2773747 | 2.050421 |
| 7 | 0.2098183 | 1.825458 |
| 8 | 0.1511535 | 1.677540 |
| 9 | 0.0965581 | 1.699875 |
| 10 | 0.0485240 | 1.856400 |
| 11 | 0.0141422 | 2.038058 |

## png   
## 2



Наименьшее значение MSE на тестовой выборке df = 6 и равно 2.246055. По графику можно предположить, что значение MSEТЕСТ, соответствует df = 6.

* Почему при таком изменении данных MSE меняется именно так?

Потому что разброс выборок уменьшается доля обучающей выборки меняется от 0.3 до 0.2 и модель пытается принять во внимание все точки