Практика №2. Вариант 8

Андрей Владимиров

Февраль 28 2021

# Задача 1

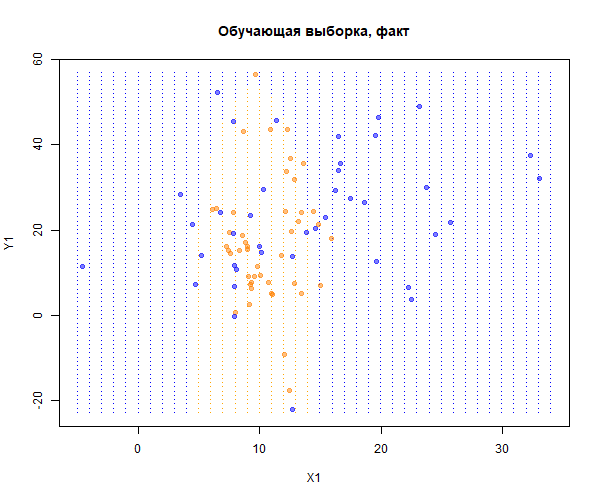
Необходимо построить модели на данных **третьего примера** с параметрами распределения:

* x1.mean.y0 = 12;
* x2.mean.y0 = 23;
* x1.sd.y0 = 7;
* x2.sd.y0 = 15;
* x1.mean.y1 = 10;
* x2.mean.y1 = 19;
* x1.sd.y1 = 3;
* x2.sd.y1 = 14.

Получим:

100, при этом доля обучающей выборки равна 0.85;

Покажем обучающую выборку на графике. Сеткой точек показаны области классов, соответсвующие истинным дискриминирующим правилам. Это правило создаем, зная истинные законы распределения классов, как максиммум из двух плотностей распределения.



Обучим модель **наивного байесовского классификатор** и оценим ее точность (верность) на обучающей выборке. Поскольку объясняющие переменные для классов сгенерированы как двумерные нормальные распределения и сами классы не перекрываются, следует ожидать, что данная окажется точной.

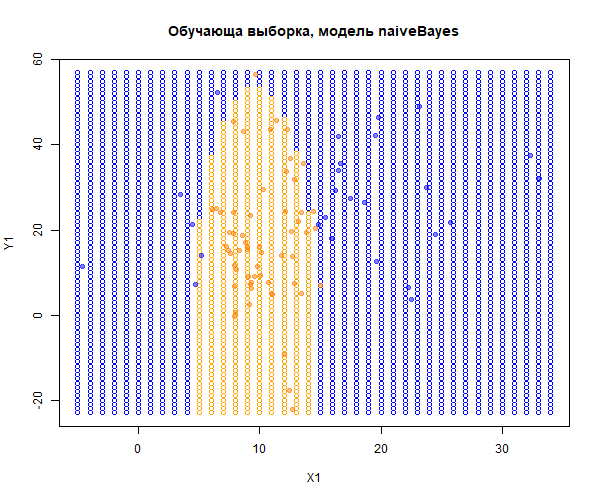
## Warning in plot.window(...): "phc" -- не графический параметр

## Warning in plot.xy(xy, type, ...): "phc" -- не графический параметр

## Warning in axis(side = side, at = at, labels = labels, ...): "phc" -- не  
## графический параметр  
  
## Warning in axis(side = side, at = at, labels = labels, ...): "phc" -- не  
## графический параметр

## Warning in box(...): "phc" -- не графический параметр

## Warning in title(...): "phc" -- не графический параметр



Теперь построим матрицу неточностей на обучающей выборке:

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
| 0 | 24 | 16 |
| 1 | 2 | 43 |

Также выведем точность (верность) на обучающей выборке:

|  |
| --- |
| x |
| 0.7882353 |

Точность на обучающей выборке высокая, поэтому сделаем прогноз классов на тестовую выборку, а также оценим точность модели.

Матрица неточностей на тестовой выборке:

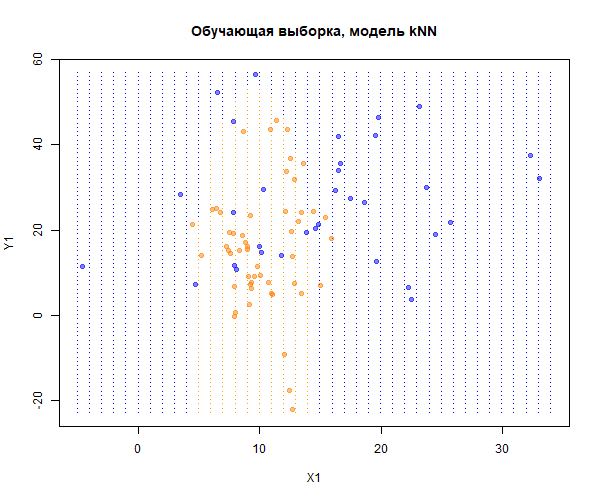
|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
| 0 | 4 | 1 |
| 1 | 4 | 6 |

Точность (верность) на тестовой выборке:

|  |
| --- |
| x |
| 0.6666667 |

Наивный байесовский метод разделяет классы на обучающей выборке, ошибаясь в двух случаях.

Построим модель **kNN**. С этими данными у метода не должно возникнуть проблем, так как он не проводит четкой границы между классами, а в каждом случае ориентируется на соседние наблюдения.



Рассчитаем матрицу неточностей на обучающей выборке для метода **kNN**:

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
| 0 | 29 | 11 |
| 1 | 4 | 41 |

Выведем точность (верность) на обучающей выборке для метода **kNN**:

|  |
| --- |
| x |
| 0.8235294 |

Рассчитаем матрицу неточностей на тестовой выборке для метода **kNN**:

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
| 0 | 4 | 1 |
| 1 | 4 | 6 |

Выведем точность (верность) на тестовой выборке для метода **kNN**:

|  |
| --- |
| x |
| 0.6666667 |

**Вывод:**

Как видно из точности обоих моделей, в обучащей выборке точность выше у kNN метода (0.8235294 (kNN) > 0.7882353 (NB)), однако на тестовых выборках точность у обоих моделей равна (0.6666667 (kNN) = 0.6666667), при этом в обоих случаях точность довольно низкая, близко к 0.5. В качестве модели для последующих вычислений будет использоваться моделей kNN.

# Задача №2

По матрице неточностей модели kNN рассчитать характеристики качеств и ошибки: TPR, SPC, PPV, NPV, FNR, FPR, FDR, MCC. Данные характеристики для обучающей и тестовой выборок приведены в таблице ниже.

|  |  |  |
| --- | --- | --- |
|  | обучающая выборка | тестовая выборка |
| TPR | 0.91 | 0.60 |
| SPC | 0.72 | 0.80 |
| PPV | 0.79 | 0.86 |
| NPV | 0.88 | 0.50 |
| FNR | 0.09 | 0.40 |
| FPR | 0.28 | 0.20 |
| FDR | 0.21 | 0.14 |
| MCC | 0.65 | 0.38 |