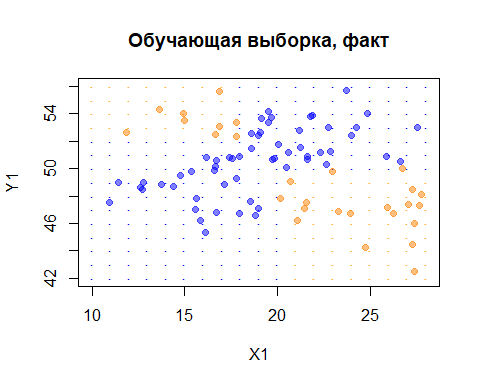
Упражнение 2

Александрова Алексадра

23 02 2021

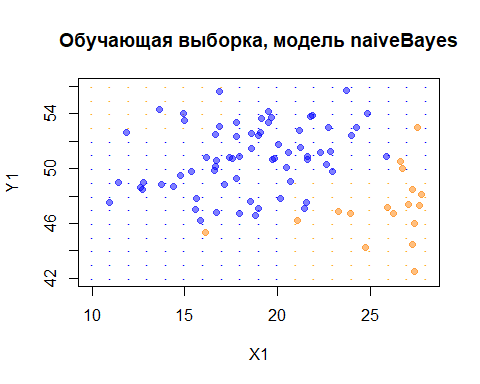
Нарисуем обучающую выборку на графике. Сеткой точек показаны области классов, соответствующие истинным дискриминирующим правилам.

#Рисуем обучающую выборку графике



Обучим модель наивного байесовского классификатора и оценим её точность (верность) на обучающей выборке.

#Байесовский классификатор



## y.nb.train  
## y.train 0 1  
## 0 54 3  
## 1 14 14

## [1] 0.8

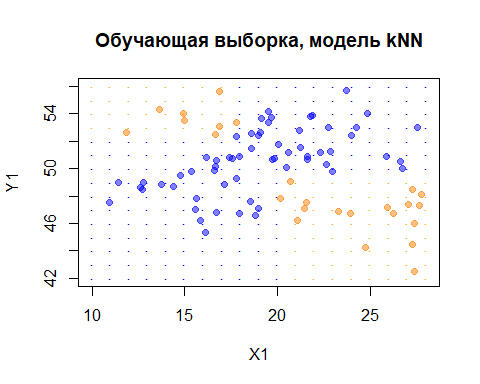
Как можно видеть на графике, ту часть жёлтого класса, которая расположена в левой верхней области пространства координат, модель классифицирует неверно. Таким образом, байесовская решающая граница не моделирует разрыв жёлтого класса синим. Это происходит потому, что в непрерывном случае наивный байесовский метод исходит из допущения о линейной разделимости двух классов и нормальности распределения объясняющих переменных в них. Однако в этом примере это допущение не выполняется.

Сделаем прогноз классов Y на тестовую выборку и оценим точность модели. Как можно убедиться, точность на тестовой оказывается ниже, чем на обучающей выборке. Учитывая, как ведёт себя классификатор на обучающей выборке, такой модели доверять не стоит.

## y.nb.test  
## y.test 0 1  
## 0 9 0  
## 1 2 4

## [1] 0.8666667

Построим модель kNN. Это “ленивый” классификатор, ему не требуется предварительное обучение. А ещё это непараметрический метод, и чем меньше количество ближайших соседей , тем гибче ведёт себя разделяющая граница. Метод хорошо работает с линейно неразделимыми классами.



## y.knn.train  
## y.train 0 1  
## 0 57 0  
## 1 3 25

## [1] 0.9647059

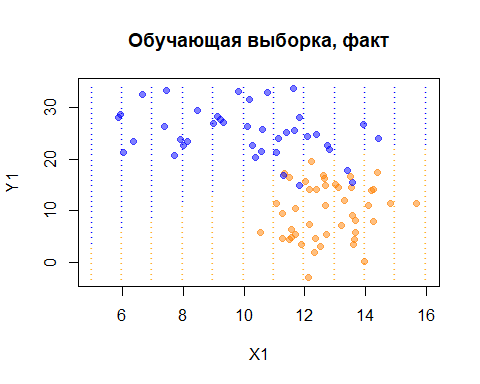
Можно видеть, что классификация обучающей выборки методом kNN не на много, но все же отличается от фактических классов наблюдений. Оценим также точность модели на тестовой выборке.

## y.knn.test  
## y.test 0 1  
## 0 6 3  
## 1 1 5

## [1] 0.7333333

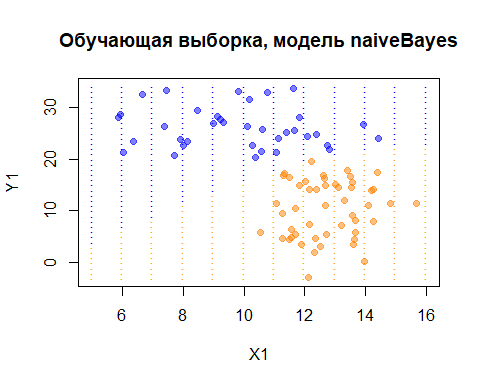
Модель kNN оказалась не достаточно точной на этих данных

Нарисуем обучающую выборку на графике. Сеткой точек показаны области классов, соответствующие истинным дискриминирующим правилам. Это правило создаём, зная истинные законы распределения классов, как максимум из двух плотностей распределения (плотность многомерного закона считаем функцией dmvnorm(), классы точкам сетки присваиваем пользовательской функцией rules.mv()).



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

### Байесовский классификатор



## y.nb.train  
## y.train 0 1  
## 0 36 4  
## 1 0 45

## [1] 0.9529412

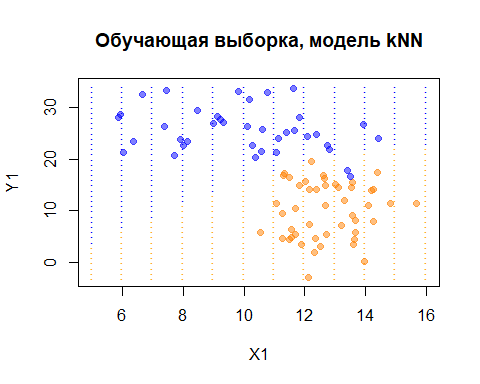
Точность на обучающей выборке высокая и близка к идеалу. Сделаем прогноз классов Y на тестовую выборку и оценим точность модели.

## y.nb.test  
## y.test 0 1  
## 0 4 1  
## 1 0 10

## [1] 0.9333333

Байесовский метод разделяет классы на обучающей выборке, не ошибаясь. Точность практически идеальна.

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



## y.knn.train  
## y.train 0 1  
## 0 37 3  
## 1 1 44

## [1] 0.9529412

Так и есть. Точность на обучающей выборке близка к идеальной. Оценка точности на тестовой выборке также показывает, что модель классифицирует верно практически все наблюдения, кроме четырех, близких к границе разделения классов.

## y.nb.train  
## y.train 0 1  
## 0 36 4  
## 1 0 45

## 0 1   
## 0 45

## [1] 1

## [1] 0.9

## [1] 0.9183673

## [1] 1

## [1] 0

## [1] 0.1

## [1] 0.08163265

## [1] 0.9091373