

## **8.1. Визуализация многомерных данных для классификации. Кластерный анализ для заданной выборки**

### **Задание**

***Воспользуйтесь набором данных USArrests. Выполните иерархическую кластеризацию штатов.***

- 1) Выполните иерархическую кластеризацию штатов, используя полное присоединение и евклидово расстояние.
- 2) Выполните расщепление дендрограммы на высоте, которая приводит к выделению трех кластеров. Какие штаты входят в каждый из этих кластеров?
- 3) Выполните иерархическую кластеризацию штатов на основе полного присоединения и евклидова расстояния после масштабирования переменных, в результате которого их стандартные отклонения становятся равными 1.
- 4) Какой эффект на полученную иерархическую кластеризацию оказывает масштабирование переменных? На ваш взгляд, следует ли выполнять масштабирование переменных перед вычислением различий между наблюдениями? Обоснуйте свой ответ.

## **8.2. Применение метода опорных векторов и деревьев решений**

### **Задание**

***Примените методы, основанные на опорных векторах, для предсказания уровня расхода топлива автомобилем (высокий или низкий) по данным из таблицы Auto.***

- 1) Создайте бинарную переменную, принимающую значение 1 для автомобилей, у которых расход топлива превышает соответствующее медианное значение, и 0 для автомобилей, у которых расход топлива меньше этого медианного значения.
- 2) Постройте по этим данным классификатор на опорных векторах с разными значениями `cost` для предсказания уровня расхода топлива (низкий или высокий). Приведите частоты ошибок, полученные в ходе перекрестной проверки при разных значениях этого параметра. Прокомментируйте свои результаты.

3) Теперь повторите (2), используя SVM с радиальным и полиномиальным ядрами для нескольких значений параметров `gamma`, `degree` и `cost`. Прокомментируйте свои результаты.

4) Постройте графики, подтверждающие Ваши ответы для пунктов (2) и (3).

Подсказка: можно применить `plot()` со следующими параметрами:

```
> plot(svmfit, dat, x1 ~ x4)
```

в которой `svmfit` содержит построенную вами модель, а `dat` — это таблица с вашими данными, Вы можете выполнить для изображения только первой и четвертой переменных. Однако Вы должны заменить `x1` и `x4` на правильные имена переменных.