Лабораторная работа 3
Подготовка исходных данных

1. Сгенерировать вектор (массив, таблица данных) и добавить в него элементы NA. Очистить данные с использованием функцииis.na() [1].
2. Сгенерировать таблицу данных с числовыми и текстовые столбцами. Очистить данные с функции complete.cases() [1].
2. Сгенерировать числовую таблицу данных с пропусками. С использованием функции preProcess из пакета caret заполнить пропуски предсказанными значениями (среднее, медиана) [2].
3. Сгенерировать два числовых набора данных, добавить в них выбросы. С использованием функции boxplot обнаружить выбросы и удалить их [3, 4].
4. Сгенерируйте таблицу данных, в которой дублируются строки. Удалите строки с использованием функций unique(), duplicated(). Сравните результаты [5].

5. Обработать пропуски в данных с использованием пакета mice [6].
6. Разобрать пример с мультиколлинеарностью [7].

Литература

1. http://datascientist.one/removing-na-values-r/
2. https://r-analytics.blogspot.com/2017/01/blog-post.html
3. http://datascientist.one/delete-outliers-with-boxplot-r/
4. https://www.r-bloggers.com/outlier-detection-and-treatment-with-r/
5. https://stackoverflow.com/questions/13967063/remove-duplicated-rows
6. https://habr.com/ru/company/infopulse/blog/305692/
7. https://datascienceplus.com/multicollinearity-in-r/

Лабораторная работа 4

Обработка данных. Выбор признаков (Feature Selection)

1. Установить пакет CARET, выполнить команду names(getModelInfo()), ознакомиться со списком доступных методов выбора признаков. Выполните графический разведочный анализ данных с использование функции featurePlot() для набора данных из справочного файла пакета CARET:

```
x <- matrix(rnorm(50*5),ncol=5)
y <- factor(rep(c("A", "B"), 25))
```

Сохранить полученные графики в *.jpg файлы. Сделать выводы.

2. С использование функций из пакета Fselector [2] определить важность признаков для решения задачи классификации. Использовать набор data(iris). Сделать выводы.
3. Установите пакет Boruta и проведите выбор признаков для набора данных data("Ozone") [3, 4]. Построить график boxplot, сделать выводы.

Литература

1. https://topepo.github.io/caret/train-models-by-tag.html#implicit-feature-selection
2. https://miningthedetails.com/blog/r/fselector/
3. https://www.jstatsoft.org/article/view/v036i11/v36i11.pdf
4. https://www.datacamp.com/community/tutorials/feature-selection-R-boruta
5. https://habr.com/ru/post/264915/
6. http://ai.stanford.edu/~ronnyk/wrappersPrint.pdf