

Конспекты лекций для курса «Введение в машинное обучение и анализ данных на Python»

Омелюсик Владимир

28 июля 2020 г.

Содержание

1. Установка программного обеспечения	1
1.1. Anaconda и Jupyter Notebook	1
1.2. Инструкция по установке новых библиотек	2
2. Тема 1: Введение в анализ данных и статистику	3
3. Тема 2: Элементы линейной алгебры	3
4. Тема 3: Введение в машинное обучение	3
5. Тема 4: Линейные методы регрессии	3
6. Тема 5: kNN	3
7. Тема 6: Решающие деревья. Случайный лес	3
8. Тема 7: Нейронные сети	3
9. Тема 8: Кластеризация	3

1 Установка программного обеспечения

1.1 Anaconda и Jupyter Notebook

На нашем курсе мы будем использовать `Python` – современный и достаточно продвинутый язык программирования, который предоставляет много удобных инструментов для анализа данных и машинного обучения. Как и многие другие языки, `Python` поставляется бесплатно, и установить его на свой компьютер может каждый желающий. Тем не менее, «сырой» `Python` представляет из себя набор методов для работы с простыми объектами: числами, строками и так далее: мы можем написать программу в любом текстовом редакторе и попросить `Python` выполнить её. Понятно, что для высокоуровневых задач, таких как машинное обучение и анализ данных, такой подход не очень удобен.

Поэтому нам также потребуются дополнительные инструменты: во-первых, удобная *среда*, в которой мы будем программировать. Во-вторых, уже написанные методы для работы с анализом

данных – без них нам придётся программировать всё самостоятельно с нуля, а это займёт слишком много времени и усилий. Такие уже написанные кем-то методы обычно распространяются в виде *библиотек*, которые можно установить, подключить и использовать.

В этом курсе мы будем использовать **Anaconda** – потрясающий дистрибутив типа «всё в одном». При установке **Anaconda** на компьютер устанавливается **Python**, среда программирования **Spyder**, удобный инструмент для анализа данных **Jupyter Notebook**, значительное число полезных библиотек и некоторые другие программы.

Инструкция по установке Anaconda.

1. Установщик **Anaconda** можно загрузить с [официального сайта](#). Почти в самом низу страницы выберите вашу операционную систему и скачайте Graphical Installer для последней доступной версии **Python** (на момент написания это версия **3.8**).
2. Запустите скачанный файл и установите **Anaconda** в соответствии с инструкциями.
3. После установки на Вашем компьютере появится приложение **Anaconda Navigator**. Запустите его.

Anaconda Navigator показывает входящие в **Anaconda** программы в удобном виде.

Непосредственно работать с анализом данных мы будем в программе **Jupyter Notebook**. Эта программа проста в освоении и позволяет выполнять код кусочками, что оказывается очень удобным во многих задачах. Также она предоставляет простые средства для красивого оформления результатов исследований. Программа поставляется вместе с **Anaconda**.

Инструкция по установке Jupyter Notebook.

1. В **Anaconda Navigator** найдите **Jupyter Notebook**. Если он не установлен, нажмите кнопку **Install**.
2. Запустите **Jupyter Notebook**.

После запуска **Jupyter Notebook** вы, скорее всего, увидите длинный список папок. Это вид *корневой папки* вашего компьютера. Её можно найти и обычным образом при использовании файловой системы. Например, на **Windows** корневая папка находится по адресу **C:\Users\imya_polzovatelya**.

Рекомендуется создать в корневой папке подпапку и назвать её так, как вам нравится. Это будет наша рабочая папка, в которой мы будем хранить все материалы по ходу курса. Сделать это можно как в файловой системе (просто создать папку в корневой папке), так и в **Jupyter Notebook** (кнопка **New** → **Folder**).

Файлы, в которых мы будем работать, называются «тетрадками» (**notebooks**). Чтобы создать новую тетрадку в той папке, которая отображается в интерфейсе **Jupyter Notebook**, нажмите **New** → **Python 3**. Тетрадка автоматически откроется в новом окне.

1.2 Установка новых библиотек

Если вы обнаружили, что требуемая библиотека не установлена на вашем компьютере, то его можно установить следующим образом:

Инструкция по установке новых библиотек.

1. Найдите на вашем компьютере программу `Anaconda Prompt` (быстрее всего это можно сделать через поиск). Запустите её. Откроется интерфейс командной строки.

2. Введите

```
1 conda install package_name  
2
```

где `package_name` – название пакета. Нажмите `Enter`.

2 Тема 1: Введение в анализ данных и статистику

3 Тема 2: Элементы линейной алгебры

4 Тема 3: Введение в машинное обучение

5 Тема 4: Линейные методы регрессии

6 Тема 5: kNN

7 Тема 6: Решающие деревья. Случайный лес

8 Тема 7: Нейронные сети

9 Тема 8: Кластеризация