

Лабораторная работа №1

Использование библиотеки `pymru`

При решении задач данной лабораторной работы необходимо использовать функции библиотеки `pymru` (обязательно). Также можно (и рекомендуется) использовать библиотеки `scipy` и `matplotlib` (либо любую другую для построения графиков).

По данной работе необходимо подготовить отчет в формате блокнота `Jupyter Notebook (.ipynb)` либо в текстовом виде (`.pdf`).

Задача 1 (5 баллов). В файле «`global-electricity-generation.csv`» представлена информация о производстве электроэнергии странами с 1992 по 2021 год. В файле «`global-electricity-consumption.csv`» – информация о потреблении электроэнергии. Все данные – в млрд. кВт*ч.

1. Загрузите информацию из этих файлов.

Указание: рекомендуется воспользоваться функцией `genfromtxt`. Обратите внимание, что данные в файлах разделены запятой, в обоих файлах есть заголовок, который при загрузке нужно пропустить. Для корректной загрузки названий стран необходимо также указать тип данных (по умолчанию загружаемые данные преобразуются в вещественный тип). Вы можете либо собрать все данные в один массив записей, либо работать с тремя разными массивами: массив названий стран (одномерный, строки), массивы производства и потребления энергии (двумерные, вещественные)).

Порядок стран в двух исходных файлах одинаковый.

2. Постройте одномерные массивы ежегодного производства и потребления электроэнергии в среднем за последние 5 лет (один массив – производство по всем странам, второй – потребление по всем странам).

Указание: не забывайте, что в агрегирующих функциях для построчной обработки нужно указывать параметр `axis=1`.

3. Напишите выражения, позволяющие получить ответ на следующие вопросы:

3.1. Суммарное (по всем странам) потребление электроэнергии за каждый год.

3.2. Максимальное количество электроэнергии, которое произвела одна страна за один год (указание: чтобы не учитывать отсутствующие и некорректные данные (`nan`) воспользуйтесь `NaN`-безопасной версией функции `max`, то есть `nanmax`).

3.3. Список стран, которые производят более 500 млрд. кВт*ч электроэнергии ежегодно в среднем за последние 5 лет (воспользуйтесь массивом, полученным на шаге 2).

3.4. 10% стран, которые потребляют больше всего электроэнергии ежегодно в среднем за последние 5 лет (указание: вначале определите соответствующую квантиль в массиве, построенном на шаге 2).

3.5. Список стран, которые увеличили производство электроэнергии в 2021 году по сравнению с 1992 годом более, чем в 10 раз.

3.6. Список стран, которые в сумме за все годы потратили больше 100 млрд. кВт*ч электроэнергии и при этом произвели меньше, чем потратили.

3.7. Какая страна потратила наибольшее количество электроэнергии в 2020 году?

Задача 2 (5 баллов). В файле «`data2.csv`» представлены данные наблюдений о прибыли (второй столбец) в зависимости от установленной скидки (первый столбец).

Комментарий. На самом деле эти данные сгенерированы синтетически, в учебных целях, но смысл значений в этой задаче не важен, и эти данные можно рассматривать как реальный датасет.

Бизнес-консультанты считают, что реальная зависимость прибыли от установленной скидки может быть описана квадратичной или кубической функцией (то есть полиномом

второй или третьей степени, $f(x) = a_2x^2 + a_1x + a_0$ или $f(x) = a_3x^3 + a_2x^2 + a_1x + a_0$.

1. Сформируйте систему линейных уравнений (СЛУ) для полинома 2й степени (для этого нужно выбрать 3 точки, в которых значение полинома должно совпадать с исходными данными; точки лучше выбирать равномерно «разбросанными» по исходным данным, то есть одна в начале имеющегося диапазона данных, одна в конце и одна в середине).

2. Решите СЛУ (с помощью `scipy.linalg.solve`), тем самым найдя коэффициенты полинома.

3. Получите вектор значений построенного полинома для заданных точек.

4. Постройте в одной области два графика: один по заданным в файле точкам, другой – по полученному вектору.

5. Посчитайте значение квадратичного отклонения RSS (оно вычисляется по формуле $RSS = \sum_{i=1}^n (y_i - f(x_i))^2$, где y_i – ожидаемые значения (из исходного файла), $f(x_i)$ – рассчитанные значения полинома.

6. Повторите шаги 1-5 для полинома 3й степени (для этого нужно будет выбрать 4 точки).

7 (по желанию). Можно поэкспериментировать с выбором точек для построения полиномов.

8. Выберите тот вариант, где значение отклонения (RSS) получается наименьшим. Для этого варианта посчитайте ожидаемое значение прибыли при значениях скидки в 6 и 8 процентов.