

Лабораторная работа 1. Корреляция и линейная регрессия (часть 2)

Цель: исследовать зависимость количества скачиваний программного продукта от количества посещений сайта средствами *Scikit-Learn*.

Исходные данные и описание проблемы см. по ссылке (часть 1):
<https://habr.com/ru/company/nerepetitor/blog/250633/>

Задания

1. Создать CSV-файл с данными.
2. Напечатать таблицу с данными (библиотека *pandas*) по посещаемости и регистрации на сайте, а также их сводные статистические показатели (среднее значение, стандартное отклонение, дисперсия, минимальное и максимальное значение), используя функции библиотеки *NumPy*. Сравнить полученные значения показателей со значениями, полученными при использовании функций, написанных самостоятельно (см. Лабораторная работа (часть 1)).
3. Построить зависимость зарегистрировавшихся на сайте от количества посещений (попробовать использовать библиотеку *Seaborn*).
4. Построить модель линейной регрессии средствами *Scikit-Learn*. Вывести значения коэффициентов линии (библиотека *Scikit-Learn*) и сравнить их со значениями, полученными при использовании функций, написанных самостоятельно (см. Лабораторная работа (часть 1), метод наименьших квадратов).
5. Ответить на следующие вопросы средствами *Scikit-Learn*.
 - При количестве посещений сайта в 8000 человек, сколько планируется получить зарегистрировавшихся (или скачавших) программный продукт?
 - Для обеспечения 500 скачиваний продукта, сколько человек должны зайти на сайт?

6. Написать общий отчет о проделанной лабораторной работе 1 (часть 1 + часть 2).

Отчет должен включать:

- 1) Постановку задачи. Представление исходных данных в виде таблицы (библиотека *pandas*) и графиков (с названиями, легендой, подписанными осями).
- 2) Описание алгоритма с основными формулами (в том числе вывод формул для метода наименьших квадратов).
- 3) Программную реализацию (исходный код программы с описанием основных «самописных» функций, а также код, заменяющий «самописные» функции встроенными функциями библиотеки *NumPy*).

- 4) Построение модели линейной регрессии с использованием собственных функций, а также средствами *Scikit-Learn*. Сравнить значения коэффициентов линии (библиотека *Scikit-Learn*) со значениями, полученными при использовании функций, написанных самостоятельно.
- 5) Основные выводы и ответы на вопросы.
- При количестве посещений сайта в 8000 человек, сколько планируется получить зарегистрировавшихся (или скачавших) программный продукт?
 - Для обеспечения 500 скачиваний продукта, сколько человек должны зайти на сайт?
- 6) Список литературы (в тексте отчета ОБЯЗАТЕЛЬНО должны быть ссылки на используемую литературу)