

УДК 373.167.1:004.42+004.42(075.3)
ББК 32.973.721
Н87

Нуньес-Иглесиас Х., Уолт ван дер Ш., Дэшноу Х.
Н87 Элегантный SciPy / пер. с англ. А. В. Логунова. – М.: ДМК Пресс, 2018. – 266 с.: ил.

ISBN 978-5-97060-600-1

Книга познакомит вас с основополагающими компонентами библиотеки SciPy языка Python. Вы научитесь писать элегантный, ясный, краткий и эффективный программный код благодаря примерам из обширной научной экосистемы Python. Кроме SciPy, вы узнаете много нового про сопутствующие библиотеки, такие как NumPy, Pandas, scikit-image.

Издание будет полезно всем программистам на Python, желающим использовать научные библиотеки в своей работе.

УДК 373.167.1:004.42+004.42(075.3)
ББК 32.973.721

Authorized Russian translation of the English edition of Elegant SciPy ISBN 9781491922873
© 2017 Juan Nunez-Iglesias, Stéfan van der Walt, and Harriet Dashnow.

This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

ISBN 978-1-491-92287-3 (анг.)

ISBN 978-5-97060-600-1 (рус.)

Copyright © 2017 Juan Nunez-Iglesias, Stéfan van der Walt, and Harriet Dashnow

© Оформление, издание, перевод, ДМК Пресс, 2018

Содержание

| | |
|---|-----------|
| Предисловие | 9 |
| Глава 1. Элегантный NumPy: фундамент научного программирования на Python | 32 |
| Введение в данные: что такое экспрессия гена? | 34 |
| N-мерные массивы NumPy | 38 |
| Зачем использовать массивы ndarray вместо списков Python? | 39 |
| Векторизация | 41 |
| Транслирование | 41 |
| Исследование набора данных экспрессии генов | 43 |
| Чтение данных при помощи библиотеки pandas | 43 |
| Нормализация | 46 |
| Нормализация между образцами | 46 |
| Нормализация между генами | 52 |
| Нормализация по образцам и генам: RPKM | 54 |
| Подведение итогов | 61 |
| Глава 2. Квантильная нормализация с NumPy и SciPy | 62 |
| Получение данных | 64 |
| Разница в распределении экспрессии генов между индивидуумами | 65 |
| Бикластеризация количественных данных | 68 |
| Визуализация кластеров | 70 |
| Предсказание выживаемости | 72 |
| Дальнейшая работа: использование кластеров пациентов TCGA | 77 |
| Дальнейшая работа: воспроизведение кластеров TCGA | 77 |
| Глава 3. Создание сетей из областей изображений при помощи ndimage | 78 |
| Изображения – это просто массивы NumPy | 79 |
| Задача: добавление сеточного наложения | 84 |
| Фильтры в обработке сигналов | 84 |
| Фильтрация изображений (двумерные фильтры) | 90 |
| Универсальные фильтры: произвольные функции от соседних значений | 92 |
| Задача: игра «Жизнь» Конуэя | 93 |
| Задача: магнитуда градиента Собела | 94 |
| Графы и библиотека NetworkX | 94 |
| Задача: подбор кривой при помощи SciPy | 98 |

| | |
|---|-----|
| Графы смежности областей..... | 98 |
| Элегантный пакет ndimage: как строить графы из областей изображений ... | 102 |
| Собираем все вместе: сегментация по среднему цвету | 105 |

Глава 4. Частота и быстрое преобразование Фурье.....107

| | |
|--|-----|
| Введение в частоту | 107 |
| Иллюстрация: спектрограмма пения птиц | 110 |
| История | 115 |
| Реализация | 115 |
| Выбор длины ДПФ | 116 |
| Дополнительные понятия ДПФ | 118 |
| Частоты и их упорядочивание | 118 |
| Оконное преобразование..... | 124 |
| Практическое применение: анализ радарных данных..... | 128 |
| Свойства сигнала в частотной области | 133 |
| Оконное преобразование на практике | 136 |
| Радарные изображения..... | 138 |
| Дополнительные применения БПФ | 142 |
| Дополнительные материалы для чтения..... | 143 |
| Задача: свертывание изображения | 143 |

Глава 5. Таблицы сопряженности на основе разреженных координатных матриц144

| | |
|---|-----|
| Таблицы сопряженности..... | 146 |
| Задача: вычислительная сложность матриц ошибок..... | 147 |
| Задача: альтернативный алгоритм вычисления матрицы ошибок..... | 147 |
| Задача: мультиклассовая матрица ошибок | 148 |
| Форматы данных модуля scipy.sparse | 148 |
| Формат COO | 148 |
| Задача: представление в формате COO | 149 |
| Формат сжатой разреженной строки | 150 |
| Применения разреженных матриц: преобразования изображений | 152 |
| Задача: поворот изображения | 156 |
| Назад к таблицам сопряженности..... | 157 |
| Задача: сокращение объема потребляемой оперативной памяти | 158 |
| Таблицы сопряженности в сегментации изображений | 159 |
| Теория информации вкратце | 160 |
| Задача: вычисление условной энтропии | 163 |
| Теория информации применительно к сегментации: изменчивость информации..... | 163 |
| Конвертирование программного кода массивов NumPy под использование разреженных матриц | 166 |

| | |
|---|------------|
| Применение изменчивости информации | 167 |
| Дальнейшая работа: сегментация на практике..... | 173 |
| Глава 6. Линейная алгебра в SciPy | 174 |
| Основы линейной алгебры | 174 |
| Лапласова матрица графа | 175 |
| Задача: матрица поворота | 176 |
| Лапласовы матрицы с данными о мозге..... | 181 |
| Задача: изображение аффинного подобия..... | 186 |
| Задача: линейная алгебра с разреженными матрицами..... | 186 |
| PageRank: линейная алгебра для репутации и важности | 187 |
| Задача: обработка висячих узлов | 192 |
| Задача: эквивалентность разных методов получения собственного вектора | 192 |
| Заключительные замечания | 192 |
| Глава 7. Оптимизация функций в SciPy | 193 |
| Оптимизация в SciPy: <code>scipy.optimize</code> | 195 |
| Пример: вычисление оптимального сдвига изображения..... | 195 |
| Регистрация изображения при помощи <code>optimize</code> | 201 |
| Предотвращение локальных минимумов на основе алгоритма <code>basin hopping</code> | 204 |
| Задача: модификация функции <code>align</code> | 205 |
| «Что лучше?»: выбор правильной целевой функции..... | 205 |
| Глава 8. Большие данные с Toolz в маленьком ноутбуке | 212 |
| Потоковая передача при помощи <code>yield</code> | 214 |
| Введение в потоковую библиотеку Toolz | 217 |
| Подсчет k-мер и исправление ошибок..... | 219 |
| Каррирование: изюминка потоковой обработки..... | 223 |
| Возвращаясь к подсчету k-мер | 226 |
| Задача: анализ главных компонент потоковых данных..... | 227 |
| Марковская модель на основе полного генома | 228 |
| Задача: онлайн-овая распаковка архива..... | 231 |
| Эпилог | 233 |
| Что дальше?..... | 233 |
| Списки рассылок | 233 |
| GitHub | 234 |
| Конференции | 235 |
| За пределами SciPy | 235 |
| Содействие этой книге | 236 |

| | |
|---|------------|
| До следующей встречи..... | 237 |
| Приложение. Решения задач..... | 238 |
| Решение: добавление сеточного наложения | 238 |
| Решение: игра «“Жизнь” Конуэя» | 239 |
| Решение: магнитуа градиента Собела | 240 |
| Решение: подбор кривой при помощи SciPy..... | 241 |
| Решение: свертывание изображения..... | 243 |
| Решение: вычислительная сложность матриц ошибок | 243 |
| Решение: альтернативный алгоритм вычисления матрицы ошибок..... | 243 |
| Решение: вычисление матрицы ошибок | 244 |
| Решение: представление в формате COO | 244 |
| Решение: поворот изображения..... | 245 |
| Решение: сокращение объема потребляемой оперативной памяти | 246 |
| Решение: вычисление условной энтропии..... | 247 |
| Решение: матрица поворота..... | 247 |
| Решение: изображение аффинного подобия..... | 248 |
| Решение: линейная алгебра с разреженными матрицами..... | 249 |
| Решение: обработка висячих узлов..... | 252 |
| Решение: методы проверки | 253 |
| Решение: модификация функции align | 253 |
| Решение: анализ главных компонент потоковых данных при помощи библиотеки scikit-learn | 255 |
| Решение: добавление шага в начало конвейера | 257 |
| Предметный указатель | 259 |