Ciencia de datos desde cero

Principios básicos con Python

2.ª Edición

Joel Grus

ANAYA MULTIMEDIA

Índice

bien la forma en la que yo hackeo cosas, que no tiene por qué ser necesariamente la suya. También conocerán bastante bien algunas de las herramientas que utilizo, que no han de ser obligadamente las mejores para ellos. Y entenderán bien el modo en que yo abordo los problemas de datos, que tampoco tiene por qué ser el mejor modo para ellos. La intención (y la esperanza) es que mis ejemplos les inspiren a probar las cosas a su manera. Todo el código y los datos del libro están disponibles en GitHub[3] para que puedan ponerse manos a la obra.

De forma similar, la mejor manera de aprender matemáticas es haciendo matemáticas. Este no es rotundamente un libro de mates, y en su mayor parte no estaremos "haciendo matemáticas". Sin embargo, no se puede hacer ciencia de datos de verdad sin ciertos conocimientos de probabilidad, estadística y álgebra lineal. Esto significa que, donde corresponda, profundizaremos en ecuaciones matemáticas, intuición matemática, axiomas matemáticos y versiones caricaturizadas de grandes ideas matemáticas. Espero que los lectores no teman sumergirse conmigo.

A lo largo de todo el libro también espero dar a entender que jugar con datos es divertido porque, bueno, ¡jugar con datos realmente lo es! (especialmente si lo comparamos con algunas alternativas, como hacer la declaración de la renta o trabajar en una mina).

Partir de cero

Hay muchísimas librerías de ciencia de datos, *frameworks*, módulos y kits de herramientas que implementan de forma eficaz los algoritmos y las técnicas de ciencia de datos más conocidas (así como las menos habituales). Si alguno de mis lectores llega a ser científico de datos, acabará estando íntimamente familiarizado con NumPy, scikit-learn, pandas y todas las demás librerías existentes. Son fabulosas para hacer ciencia de datos, pero también suponen una buena forma de empezar a hacer ciencia de datos sin realmente comprender lo que es.

En este libro nos acercaremos a la ciencia de datos desde el principio de los principios. Esto significa que crearemos herramientas e implementaremos