Data Science in Ten Minutes[*]

Spencer Tipping

May 19, 2018

Contents

1 Introduction

Data science is not machine learning.

---

[*]for extremely large values of ten

2 Linux

Oh yes, we are totally going here. Here's why.

Backend programs and processing pipelines and stuff (basically, "big data" things) operate entirely by talking to the kernel, which, in big-data world, is usually Linux; and this is true regardless of the language, libraries, and framework(s) you're using. You can always throw more hardware at a problem[1], but if you understand system-level programming you'll often have a better/cheaper option.

Some quick background reading:

- How to write a UNIX shell

- How to write a JIT compiler

- ELF executable binary spec

- Intel machine code documentation

2.1 Homework (if that's your thing)

1. Write an ELF Linux executable that prints `hello world` to stdout, then exits successfully.

---

[1]Until you can't