Data Science Notebook
PRINCIPLES OF DATA SCIENCE

Yigal Weinstein

$\mathcal{YW}$

2016

Prologue

There is content that I'm trying to learn in the pursuit of data science. This note book serves the following purposes:

1. Continuously analyze and synthesize new material as I'm brought into contact with it. Each chapter is a topic referenced in Zipfian's Data Science Primer [3].

2. Have a catalog of what I've studied up until entering into the boot camp.

3. Be honest with the quantity and quality of work I've produced.

Chapter 1

Machine Learning

1.1 Quotes

"models lend the machine learning field diversity, but tasks and features give it unity." [1, p. 13] Machine learning is all about using the proper algorithms to accomplish the desired tasks.

Chapter 2

Coding - Python

There is a Pythonic way of coding. This is a culturally obtained set of standards. Values are,

- explicit

- choose simple over complex

- maximize readability

LeBlanc's law: Later equals never.
Martin, Robert C. (2008-08-01). Clean Code: A Handbook of Agile Software Craftsmanship (Kindle Location 553). Pearson Education. Kindle Edition.
What is good code?
From the introduction it appears there are some qualities that if enough are satisfied it should be called good code. What is interesting is that good code can be distinguished quite easily. Good code should be,

- readable, and understandable

- elegantly solves the task it was built to complete

- has a full suite of tests which it passes

The stress of the authors of [2] appear to be making the connection between good and clean code. What good code is, is still open for debate what the authors provide are behaviors they've learned through their study. That is read it to learn but don't expect it to contain absolute truth about coding well.

Chapter 3

Algorithms

Chapter 4

SQL

Chapter 5

Probability

Chapter 6

Statistics

Chapter 7

Linear Algebra

Glossary

task An abstract representation of a problem we want to solve regarding the domain objects of import..

Bibliography

[1] Peter Flach. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data.* Cambridge University Press, 2012.

[2] Robert Martin. *Clean code : a handbook of agile software craftsmanship.* Prentice Hall, Upper Saddle River, NJ, 2009.

[3] zipfian. data-science-primer. `https://github.com/zipfian/data-science-primer`, 2015.