



DEEP
LEARNING
INSTITUTE



DLI Accelerated Data Science Teaching Kit

Lecture 6.1 - Break Complex Problems into Simpler Ones: Part 1

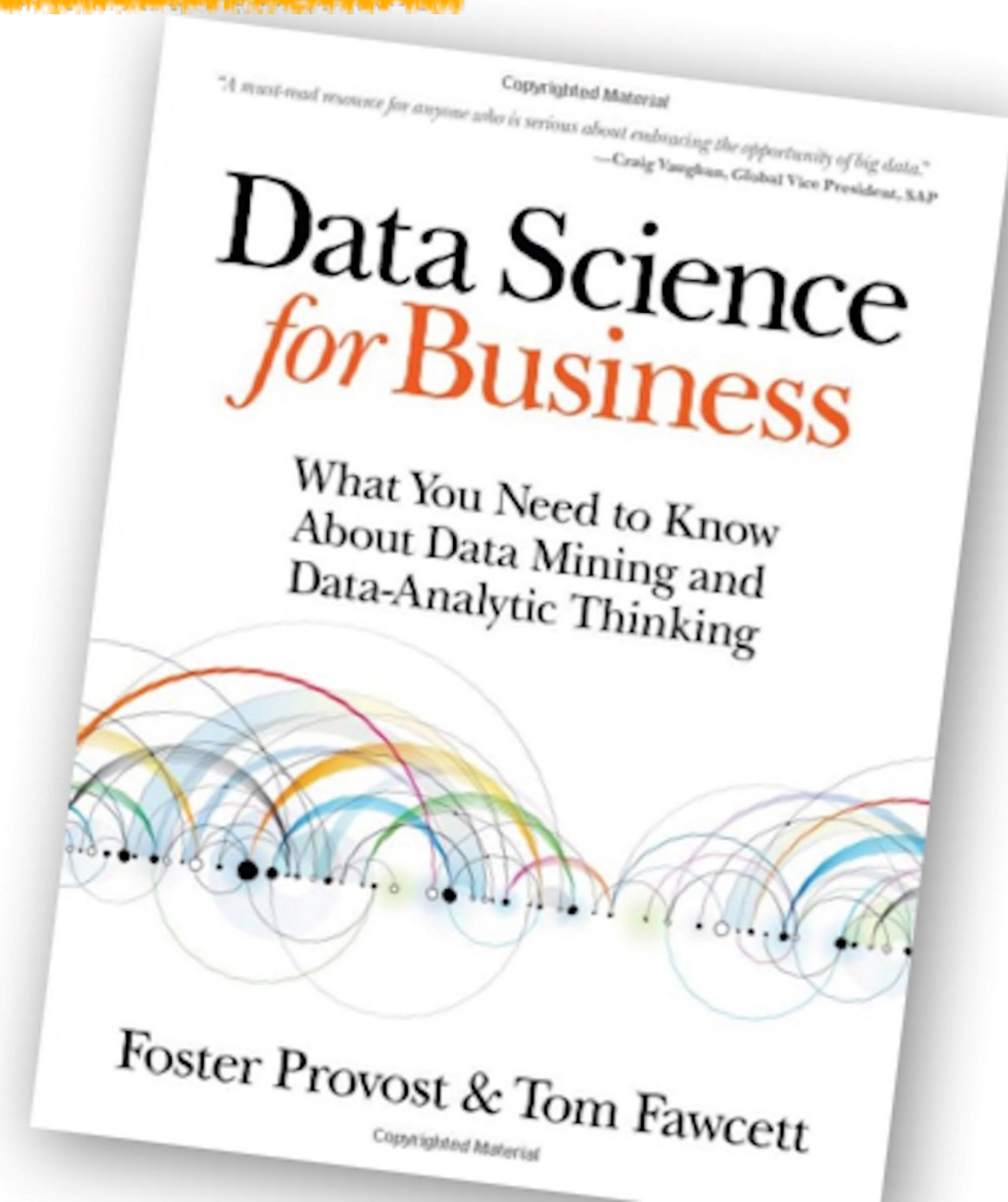


The Accelerated Data Science Teaching Kit is licensed by NVIDIA, Georgia Institute of Technology, and Prairie View A&M University under the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

A critical skill in data science is the ability to decompose a data-analytics problem into pieces such that each piece matches a known task for which tools are available. Recognizing familiar problems and their solutions avoids wasting time and resources reinventing the wheel. It also allows people to focus attention on more interesting parts of the process that require human involvement—parts that have not been automated, so human creativity and intelligence must come in-to play.

8 non-mutually exclusive concept classes

<http://www.amazon.com/Data-Science-Business-data-analytic-thinking/dp/1449361323>



1. Classification

Predict which of a (small) set of classes an entity belongs to:

- email spam (y, n)
- sentiment analysis (+, -, neutral)
- news (politics, sports, ...)
- medical diagnosis (cancer or not)
- shirt size (s, m, l)
- cat detection
- face detection (baby, middle-aged, etc)
- buy/not buy - commerce

2. Regression

(“value estimation”)

Predict the numerical value of some variable for an entity.

- point value of wine (50-100)
- credit scores
- stock prices
- relationship between price and sales
- weather
- sports and game scores

3. Similarity Matching

Find similar entities (from a large dataset) based on what we know about them.

- find similar gene sequences (that may be repeating, or does similar things)
- online dating
- patent search
- carpool matching (find people to carpool)

4. Clustering (For most algorithms, user provides # of clusters)

(unsupervised learning)

Group entities together by their similarity.

- groupings of similar bugs in code
- topical analysis (tweets?)
- land cover: tree/road/...
- for advertising: grouping users for marketing purposes
- cluster people by accents (y'all, you all)



DEEP
LEARNING
INSTITUTE



DLI Accelerated Data Science Teaching Kit

Thank You