

Data Science Resources

1201 Data Science: Tim Raiswell

March 6, 2019

General Learning Resources

Resources From 1201 Datascience

1. Materials from the first five sessions in 2018: <https://github.com/timraiswell/1201DS>
2. Jitendra's machine learning course materials: <https://github.com/jsub10/Machine-Learning-Course-2018>

Online Courses

I'll only comment on the courses I have used:

1. Datacamp is worth every penny. It lets you choose from hundreds of customizable options in multiple programming languages.
2. Coursera is excellent, although gives the impression of letting some of its courses get a little dated.
3. YouTube is full of great, complete courses. Search for 'MIT data science', 'MIT deep learning', etc.
4. For general stats, probability and machine learning mathematics, Penn State's online course materials are *really* high-quality: <https://newonlinecourses.science.psu.edu/statprogram/>
5. Lectures for the book: 'An Introduction to Statistical Learning':
<http://fs2.american.edu/alberto/www/analytics/ISLRLectures.html>
6. Amazon offers access to the training that its own engineers create and use:
<https://aws.amazon.com/training/course-descriptions/machine-learning/>

Language Resources

Recall that Python is a multi-purpose programming language and is not *just* for data science, hence the broader range of support resources.

R

- <https://www.r-bloggers.com/>
- <https://github.com/qinwf/awesome-R>
- <https://r4ds.had.co.nz/index.html>
- <https://www.tidytextmining.com/>
- <https://blog.revolutionanalytics.com>

Python

- <https://awesome-python.com/>
- <https://towardsdatascience.com/> (R articles too, but mostly Python)
- <https://realpython.com/>
- <https://medium.com/@Mybridge/learn-python-from-top-50-articles-for-the-past-year-v-2019-3225f4643d9c>

Data

- <https://dataverse.harvard.edu/>
- <https://archive.ics.uci.edu/ml/index.php>
- <https://gengo.ai/datasets/24-best-ecommerce-retail-datasets-for-machine-learning/>

Books

In no particular order:

1. R for Data Science, Wickham: <https://r4ds.had.co.nz/index.html>
2. Python for Data Analysis, McKinney, 2nd Edition
3. Data Science for Business, Provost and Fawcett
4. An Introduction to Statistical Learning, <https://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf>
5. The Elements of Statistical Learning, <https://web.stanford.edu/~hastie/Papers/ESLII.pdf>
6. Machine Learning Yearning, Ng: <https://www.mlyearning.org>

Miscellanea

1. <https://paperswithcode.com/>
2. <https://git.news/>
3. <https://discourse.datamethods.org/>
4. [Kaggle.com](https://www.kaggle.com/)
5. <https://ai.google/education/>
6. <https://colab.research.google.com>

Gartner Research Data Science Use Cases

This is not a comprehensive list, just some examples I use in my own workflow.

1. Making quick presentations in code.
2. Manipulating large, slow Excel files.
3. Analysis of client problems.
4. Data generation and natural language processing.