Shortlink: https://tinyurl.com/data-science-links
Links curated by Jerry Chi (jerrychi123@gmail.com, jerrychi.com) ; feel free to add!

General data science

- Data science cheatsheet PDF
- http://datasciencemasters.org/ Open Data Science master's curriculum

Data science and organizations / leadership

- **Awesome data leadership** - list of resources
- Scientific Debt
- Data dictionary / glossary best practices
- How to be an awesome data scientist
- How to be an awesome data scientist manager

Statistics / math

- https://seeing-theory.brown.edu/
- Ritvik Math https://youtube.com/c/ritvikmath
- StatQuest https://www.youtube.com/channel/UCtYLUTtgS3k1Fg4y5tAhLbw
- https://brilliant.org/courses/statistics/
- https://alexanderetz.com/2016/02/07/understanding-bayes-how-to-become-a-bayesian-in-eight-easy-steps/
- 3blue1brown YouTube
- Bayes Rules ! book https://www.bayesrulesbook.com/
- https://mathlets.org/mathlets/
- Linear algebra
    - MIT intro to linear algebra
- 2022 ML math summer school
- Independent components analysis (ICA) - a tutorial
- A/B testing
    - Evan Miller's A/B testing tools / articles
        - how not to run an A/B test
    - A/B test significance, power, & effect size Interactive visualization
    - (2-sample normal dist) Sample size / power calculator

Python

- Intro to interactive programming in Python (Coursera)
- https://github.com/r0f1/datascience python tools for data science
- Templated formatting of strings in Python3
- Data Classes in Python3.7

- 2021 python data structures and algos https://t.co/IGP1c6W0du
- Pandas
    - Pandas and pivoting/reshaping
    - Pandas functions w/ examples
    - 101 Pandas exercises
    - Modern Pandas series

Newsletters - data science

- The Batch by deeplearning.ai (founded by Andrew Ng)
- TWIML (This Week in Machine Learning and AI) newsletter
- The Machine Learning Engineer newsletter
- ML News https://mln.dev/ . Like Hacker News except for ML.
- ChinAI . Curated selection/translation of articles etc. about ML/AI in China.
- Data Science Weekly
- Python Weekly includes a short list of cool new Python packages
- Data Science Central
- https://dataengweekly.com/
- https://www.kdnuggets.com/
- O'reilly Data News
- Air bnb data science blog

Book recommendations

- Data science book list
- Data Smart (overview of various data science methods demonstrated in spreadsheets)

Podcasts

- Gradient Dissent (host is Lukas Biewald, founer of Weights & Biases)
- No Priors - Elad Gil and Sarah Guo
- The Data Skeptic
- O'Reilly Data Show
- Partially Derivative
- Linear Digressions
- DataFramed
- Practical AI
- Microsoft Research
- Machine learning street talk
- Data science at home
- Software engineering daily

Youtube

- 2-minute papers . doesn't go much into the algos, mainly summarizes the results.
- Machine Learning Tokyo (all in English)
- Yannic Kilcher (paper explanations etc.)
- AI Epiphany (paper explanations, similar to Yannic Kilcher)

Online classes / online interactive coding courses / MOOCs platforms

- https://www.dataquest.io/ (more Python than R. also, some data engineering)
- https://www.datacamp.com/
- https://www.codecademy.com/
- https://www.brilliant.org (website and mobile app; interactive modules on probability statistics, neural networks, etc.)
- Also see: list of courses that Jerry has taken / will take
- List of recommended courses https://aman.ai/watch

Conferences

- tokyo
    - Tokyo R
    - PyData Tokyo
    - Data Analyst Meetup (Tokyo)
- KDD Conference
- Strata Conference
- PyData
- ICML
- NIPS

General technical tips for productivity

- https://www.alfredapp.com/ more powerful version of Mac spotlight
- Learn Slack's advanced functionality (keyboard shortcuts etc.)

Command Line

- oh-my-zsh: easily set up your command line to be awesome
- Command line tips for manipulating text / CSV / TSV files
- Search Command line history
- **fselect** tool: use SQL to find/sort files on your laptop/server
- Use **screen** command on often-used servers
- Good terminal setup (e.g. http://ohmyz.sh/ and iterm2)
- https://github.com/sharkdp/bat much better version of Mac/Linux "cat" command.
- Setup SSH aliases and command line aliases for common tasks

Other

- General thinking tools / approaches
- Data Science interviews

Machine Learning

General ML

- (2024) resources to break in AI (from Thomas Wolfe co-founder of Hugging Face) https://thomwolf.io/data/Thom_wolf_reading_list.txt
- https://beta.mlsearch.ai/ …created a platform that serves a central place for searching for machine learning resources (code, papers, videos, blog posts, tools, etc.) and bookmarking them in your personal profile. So instead of having to go to many different websites or browse Google endlessly you can look for specific keywords and we fetch relevant information for you from different places in one sweep
- ML papers explained https://github.com/dair-ai/ML-Papers-Explained
- (2022) pen and paper exercises in ML (mostly math)
- https://github.com/louisfb01/best_AI_papers_2021
- https://papers.labml.ai/papers/weekly
- Best practices for ML by Google
- https://explained.ai/decision-tree-viz/index.html decision trees explained visually
- Recommend studying at least a little bit of basic machine learning before studying deep learning. recommend taking at least the first 2 courses in this free online Machine Learning Specialization on Coursera by the University of Washington. The rating is very high, the explanations are good, and the homework is in Python in Jupyter notebooks, which is very convenient.
- Awesome interpretable ML (curated list of interpretable ML resources)
- list of links / resources
- another list of links/resources
- Yet another list of resources
- ML Interactive experiments
- Made with ML - community / feed of ML projects.. Submit your own :)
- good explanation of autoencoders and variational autoencoders (VAE)

ML DevOps, doing ML in production, etc.

- Great overview by Google on MLOps
- https://applyingml.com/
- The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction (2017 paper from Google)
- https://rise.cs.berkeley.edu/blog/a-short-history-of-prediction-serving-systems/
- Applied ML in production curated papers etc
- Docker for data science

Feature selection / feature engineering

- http://efavdb.com/unsupervised-feature-selection-in-python-with-linselect/

Recommender systems (recsys)

- 10 recsys papers everyone should read

Natural language processing (NLP)

- NLP specialization on Coursera / deeplearning.ai

- NLP papers curated: https://github.com/NiuTrans/ABigSurvey

Deep Learning-specific

- Annotated Research Paper Implementations: Transformers, StyleGAN, Stable Diffusion, DDPM/DDIM, LayerNorm, Nucleus Sampling and more
- https://animatedai.github.io/ Animations of neural network architectures (convolution etc.)
- Ilya Sutskevers' AI reading list https://tensorlabbet.com/2024/09/24/ai-reading-list/
- Fast.ai - high-level, easy-to-use library based on PyTorch with state-of-the-art results. Has a free online course.
- https://deep-learning-drizzle.github.io/ list of deep learning courses including Youtube links
- https://github.com/ahkarami/Deep-Learning-in-Production
- General deep learning
    - study list on github
    - Practical advice for building deep neural networks
    - https://github.com/floodsung/Deep-Learning-Papers-Reading-Roadmap
    - Annotated/explained PyTorch implementations of popular papers/models (labml.ai)
- Primer on transformers etc.
- Online book on deep learning for Search / neural retrieval
- Neural Net Loss Landscapes - how can NNs generalize?
- Tensorflow
    - https://github.com/vahidk/EffectiveTensorflow/blob/master/README.md
    - Getting started with Tensorflow 2
    - Intro to Tensorflow for Deep Learning
- https://e2eml.school/transformers.html transformers from scratch
- Gentle intro to graph neural networks https://distill.pub/2021/gnn-intro/
- Transformers / attention
    - 'The Illustrated Transformer' and the follow-on 'The Illustrated GPT-2' for getting a better mental picture of things like self-attention. 'The Annotated GPT-2' is similar with a more code-focused approach.
    - The Neuromatch content for transformers
-

Research / papers

- https://artigopapers.io/home/trending
- Semantic Scholar | AI-Powered Research Tool
- https://paperswithcode.com/
- Annotated pytorch paper / Algo implementations https://nn.labml.ai/

Creative AI / Generative AI

- See list of resources by yours truly