Awesome Resources

Deep Learning Math Requirement	
Awesome Courses	7
Awesome Artificial Intelligence (AI)	53
Awesome Deep Learning	59
Deep Learning Papers Reading Roadmap	72
Machine Learning & Deep Learning Tutorials	85
Data science Python notebooks	107
data-science-ipython-notebooks	107
Awesome Machine Learning	124
Awesome Public Datasets	163
Awesome TensorFlow	178
Awesome Mac	185
Awesome Big Data	215
Top-down learning path: Machine Learning for Software Engineers	237
Motivation	240

Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow, 2nd Edition

ty4z2008/Qix

Machine Learning Deep Learning PostgreSQL Distributed System Node.Js Golang

terryum/awesome-deep-learning-papers

sbrugman/deep-learning-papers

Papers about deep learning ordered by task, date. Current state-of-the-art papers are labelled.

eriklindernoren/ML-From-Scratch

aymericdamien/TopDeepLearning:

Top Deep Learning Projects: A list of popular github projects related to deep learning

kailashahirwar/cheatsheets-ai

Spandan-Madan/DeepLearningProject

An in-depth machine *learning* tutorial introducing readers to a whole machine *learning* pipeline from scratch.

SamDeepLearning/The-Terrible-Deep-Learning-List

15 working examples to get you started with Deep Learning without learning any of the math.

vinta/awesome-python

A curated list of awesome Python frameworks, libraries, software and resources

bulutyazilim/awesome-datascience

An awesome Data Science repository to learn and apply for real world problems.

kjw0612/awesome-deep-vision

A curated list of deep learning resources for computer vision

Contents

Awesome Courses	7
Introduction	7
Table of Contents	7
Legend	8
Courses	8
Systems	8
Programming Languages / Compilers	15
Algorithms	20
CS Theory	25
Introduction to CS	27
Machine Learning	33
Security	39
Artificial Intelligence	42
Computer Graphics	43
Misc	45
Awesome Artificial Intelligence (AI)	53
Contents	53
Courses	53
Books	54
Programming	54
Philosophy	54
Free Content	55
Code	56
Videos	56
Learning	56
Organizations	56
Journals	57
Competitions	57
Movies	58
Misc	58
Awesome Deep Learning	59

Table of Contents	59
Free Online Books	59
Courses	59
Videos and Lectures	60
Papers	61
Tutorials	62
Researchers	62
WebSites	64
Datasets	64
Frameworks	69
Miscellaneous	70
Deep Learning Papers Reading Roadmap	72
1 Deep Learning History and Basics	
1.0 Book	72
1.1 Survey	72
1.2 Deep Belief Network(DBN)(Milestone of Deep Learning Eve)	72
1.3 ImageNet Evolution (Deep Learning broke out from here)	73
1.4 Speech Recognition Evolution	73
2.1 Model	74
2.2 Optimization	75
2.3 Unsupervised Learning / Deep Generative Model	75
2.4 RNN / Sequence-to-Sequence Model	
2.5 Neural Turing Machine	76
2.6 Deep Reinforcement Learning	77
2.7 Deep Transfer Learning / Lifelong Learning / especially for RL	77
2.8 One Shot Deep Learning	78
3 Applications	
3.1 NLP(Natural Language Processing)	
3.2 Object Detection	79
3.3 Visual Tracking	80
3.4 Image Caption	81
3.5 Machine Translation	81
3.6 Robotics	82
3.7 Art	83
3.8 Object Segmentation	83
Machine Learning & Deep Learning Tutorials	85

Contents	85
Miscellaneous	85
Interview Resources	86
Artificial Intelligence	87
Genetic Algorithms	87
Statistics	87
Useful Blogs	88
Resources on Quora	89
Kaggle Competitions WriteUp	89
Cheat Sheets	
Classification	90
Linear Regression	90
Logistic Regression	91
Model Validation using Resampling	91
Deep Learning	92
Natural Language Processing	98
Computer Vision	100
Support Vector Machine	100
Reinforcement Learning	101
Decision Trees	101
Random Forest / Bagging	103
Boosting	103
Ensembles	104
Stacking Models	104
Vapnik–Chervonenkis Dimension	105
Bayesian Machine Learning	105
Semi Supervised Learning	
Optimization	106
Other Tutorials	
Data science Python notebooks	107
data-science-ipython-notebooks	107
Index	107
deep-learning	107
tensor-flow-tutorials	
tensor-flow-exercises	
theano-tutorials	

keras-tutorials	110
deep-learning-misc	111
scikit-learn	111
statistical-inference-scipy	112
pandas	113
matplotlib	
numpy	
python-data	
kaggle-and-business-analyses	119
spark	119
mapreduce-python	120
aws	120
commands	
misc	
notebook-installation	122
anaconda	122
dev-setup	123
running-notebooks	123
credits	123

Deep Learning Math Requirement

- · Linear algebra
- Multi-variable (vector) calculus
- Probability and statistics
- Basic optimization
- Statistics
- differential calculus
- numeral optimization
 - Probability and Statistics (for linear and logistic regressions and the maximum likelihood estimator)
 - Multivariate Calculus (for all the partial derivatives and chain rules in back propagation)
 - Convex Optimization (for stochastic gradient descent)
 - Linear Algebra(this is a must for machine learning in general, to be comfortable with the feature matrix, matrix notation etc)
 - Calculus (I, II and III)
 - Differential Equations
 - Linear Algebra
- Statistics & **probability** (or a good covering of Bayes)

Awesome Courses

https://github.com/prakhar1989/awesome-courses

Introduction

There is a lot of hidden treasure lying within university pages scattered across the internet. This list is an attempt to bring to light those awesome courses which make their high-quality material i.e. assignments, lectures, notes, readings & examinations available online for free.

Table of Contents

- Algorithms
- Artificial Intelligence
- Computer Graphics
- <u>CS Theory</u>
- Introduction to CS
- Machine Learning
- Misc
- Programming Languages / Compilers
- <u>Security</u>

Systems

Legend

- If Lecture Videos
- Lecture Notes
- Assignments / Labs
- 🔹 똘 Readings

Courses

Systems

- CS 61C Great Ideas in Computer Architecture (Machine Structures) UC Berkeley
 - The subjects covered in this course include: C and assembly language programming, translation of high-level programs into machine language, computer organization, caches, performance measurement, parallelism, CPU design, warehouse-scale computing, and related topics.
 - Lecture Videos
 - Lecture Notes
 - Resources
 - Old Exams
- CS 107 Computer Organization & Systems Stanford University 📷 💻
 - CS107 is the third course in Stanford's introductory programming sequence. The course will work from the C programming language down to the microprocessor to de-mystify the machine. With a complete understanding of how computer systems execute programs and manipulate data, you will become a more effective programmer, especially in dealing with issues of debugging, performance, portability, and robustness.
 - Lecture Videos
 - Assignments
- CS 140 Operating Systems Stanford University
 - o This class introduces the basic facilities provided in modern operating systems. The course divides into three major sections. The first part of the course discusses concurrency. The second part of the course addresses the problem of memory management. The third major part of the course concerns file systems.
 - Lecture Notes
 - o **Assignments**
- 6.004 Computation Structures MIT **E** 📝 📷
 - Introduces architecture of digital systems, emphasizing structural principles common to a wide range of technologies. Multilevel implementation strategies;

definition of new primitives (e.g., gates, instructions, procedures, processes) and their mechanization using lower-level elements. Analysis of potential concurrency; precedence constraints and performance measures; pipelined and multidimensional systems. Instruction set design issues; architectural support for contemporary software structures. 4 Engineering Design Points. 6.004 offers an introduction to the engineering of digital systems. Starting with MOS transistors, the course develops of series of building blocks logic gates, combinational and sequential circuits, finite-state machines, computers and finally complete systems. Both hardware and software mechanisms are explored through a series of design examples.

- **Youtube Playlist**
- **Lecture Notes**
- Labs-Assignments

CS 162 Operating Systems and Systems Programming UC Berkeley 🖬 🥟 💻 答

- The purpose of this course is to teach the design of operating systems and operating systems concepts that appear in other advanced systems. Topics we will cover include concepts of operating systems, systems programming, networked and distributed systems, and storage systems, including multiple-program systems (processes, interprocess communication, and synchronization), memory allocation (segmentation, paging), resource allocation and scheduling, file systems, basic networking (sockets, layering, APIs, reliability), transactions, security, and privacy.
 - Operating Systems course by the Chair of EECS, UC Berkeley David Culler
 - Lecture Videos Spring 2015 lectures
 - **Lecture Notes** Spring 2015 lectures

CS 168 Introduction to the Internet: Architecture and Protocols UC Berkeley



- This course is an introduction to the Internet architecture. We will focus on the concepts and fundamental design principles that have contributed to the Internet's scalability and robustness and survey the various protocols and algorithms used within this architecture. Topics include layering, addressing, intradomain routing, interdomain routing, reliable delivery, congestion control, and the core protocols (e.g., TCP, UDP, IP, DNS, and HTTP) and network technologies (e.g., Ethernet, wireless).
- **Lecture Notes & Assignments**
- **Discussion Notes**
- CS 179 GPU Programming Caltech
 - This course will cover programming techniques for the GPU. The course will introduce NVIDIA's parallel computing language, CUDA. Beyond covering the CUDA programming model and syntax, the course will also discuss GPU architecture, high performance computing on GPUs, parallel algorithms, CUDA libraries, and applications of GPU computing.
 - <u>Assignments</u>
 - **Lecture Notes**
- CS 186 Introduction to Database Systems UC Berkeley 🖬 💻 答 🦻

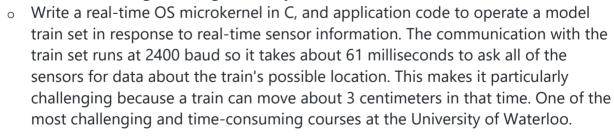
- In the project assignments in CS186, you will write a basic database management system called SimpleDB. For this project, you will focus on implementing the core modules required to access stored data on disk; in future projects, you will add support for various query processing operators, as well as transactions, locking, and concurrent queries.
- **Lecture Videos**
- **Lecture Notes**
- **Projects**
- CS 241 Systems Programming (Spring 2016) Univ of Illinois, Urbana-Champaign



- System programming refers to writing code that tasks advantage of operating system support for programmers. This course is designed to introduce you to system programming. By the end of this course, you should be proficient at writing programs that take full advantage of operating system support. To be concrete, we need to fix an operating system and we need to choose a programming language for writing programs. We chose the C language running on a Linux/UNIX operating system (which implements the POSIX standard interface between the programmer and the OS).
- <u>Assignments</u>
- Labs
- Github Page
- **Crowd Sourced Book**
- CS 425 Distributed Systems Univ of Illinois, Urbana-Champaign



- Brilliant set of lectures and reading material covering fundamental concepts in distributed systems such as Vector clocks, Consensus and Paxos. This is the 2016 version by Prof Indranil Gupta.
- Lectures
- <u>Assignments</u>
- CS 452 Real-Time Programming University of Waterloo



- <u>Assignments</u>
- Lecture notes
- CS 2043 Unix Tools & Scripting Cornell University
 - UNIX-like systems are increasingly being used on personal computers, mobile phones, web servers, and many other systems. They represent a wonderful family of programming environments useful both to computer scientists and to people in many other fields, such as computational biology and computational linguistics, in which data is naturally represented by strings. This course provides an intensive training to develop skills in Unix command line tools and scripting that enable the accomplishment and automation of large and challenging computing tasks. The

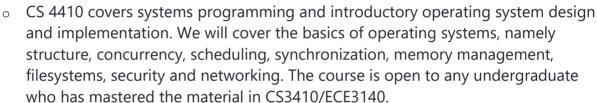
syllabus takes students from shell basics and piping, to regular-expression processing tools, to shell scripting and Python.

- **Syllabus**
- Lectures
- <u>Assignments</u>
- CS 3410 Computer System Organization and Programming Cornell University

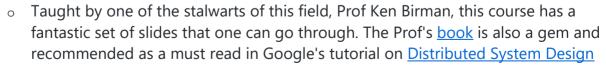


CS3410 provides an introduction to computer organization, systems programming and the hardware/software interface. Topics include instruction sets, computer arithmetic, datapath design, data formats, addressing modes, memory hierarchies including caches and virtual memory, I/O devices, bus-based I/O systems, and multicore architectures. Students learn assembly language programming and design a pipelined RISC processor.

- Lectures
- **Assignments**
- CS 4410 Operating Systems Cornell University



- Syllabus
- Lectures
- CS 4414 Operating Systems University of Virginia
 - o A course (that) covers topics including: Analysis process communication and synchronization; resource management; virtual memory management algorithms; file systems; and networking and distributed systems. The primary goal of this course is to improve your ability to build scalable, robust and secure computing systems. It focuses on doing that by understanding what underlies the core abstractions of modern computer systems.
 - Syllabus
 - Lectures
- CS 5412 Cloud Computing Cornell University



- Slides
- CSCE 3613 Operating Systems University of Arkansas (Fayetteville) 📕 💓 🛎 An introduction to operating systems including topics in system structures, process management, storage management, files, distributed systems, and case studies.
 - **Syllabus**
 - <u>Assignments</u>
 - **Lecture Notes**

- o **Readings**
- CSCI-UA.0202: Operating Systems (Undergrad) Operating Systems NYU



- o NYU's operating system course. It's a fundamental course focusing basic ideas of operating systems, including memory management, process sheeduling, file system, ect. It also includes some recomended reading materials. What's more, there are a series of hands-on lab materials, helping you easily understand OS.
- **Assignments**
- Lectures
- Old Exams
- CSCI 360 Computer Architecture 3 CUNY Hunter College



- A course that covers cache design, buses, memory hierarchies, processorperipheral interfaces, and multiprocessors, including GPUs.
- CSCI 493.66 UNIX System Programming (formerly UNIX Tools) CUNY Hunter College 💻 🦻
 - o A course that is mostly about writing programs against the UNIX API, covering all of the basic parts of the kernel interface and libraries, including files, processes, terminal control, signals, and threading.
- CSCI 493.75 Parallel Computing CUNY Hunter College
 - The course is an introduction to parallel algorithms and parallel programming in C and C++, using the Message Passing Interface (MPI) and the OpenMP application programming interface. It also includes a brief introduction to parallel architectures and interconnection networks. It is both theoretical and practical, including material on design methodology, performance analysis, and mathematical concepts, as well as details on programming using MPI and OpenMP.
- Hack the Kernel Introduction to Operating Systems SUNY University at Buffalo, NY 📷 💻 🦻
 - o This course is an introduction to operating system design and implementation. We study operating systems because they are examples of mature and elegant solutions to a difficult design problem: how to safely and efficiently share system resources and provide abstractions useful to applications.
 - For the processor, memory, and disks, we discuss how the operating system allocates each resource and explore the design and implementation of related abstractions. We also establish techniques for testing and improving system performance and introduce the idea of hardware virtualization. Programming assignments provide hands-on experience with implementing core operating system components in a realistic development environment. Course by Dr.Geoffrey Challen
 - **Syllabus**
 - Slides
 - Video lectures
 - <u>Assignments</u>
 - **Old Exams**
- ECE 459 Programming for Performance University of Waterloo 📝 💻



- Learn techniques for profiling, rearchitecting, and implementing software systems that can handle industrial-sized inputs, and to design and build critical software infrastructure. Learn performance optimization through parallelization, multithreading, async I/O, vectorization and GPU programming, and distributed computing.
- Lecture slides

PODC Principles of Distributed Computing ETH-Zurich =

- Explore essential algorithmic ideas and lower bound techniques, basically the "pearls" of distributed computing in an easy-to-read set of lecture notes, combined with complete exercises and solutions.
- o Book
- o Assignments and Solutions
- SPAC Parallelism and Concurrency Univ of Washington 📚
 - Technically not a course nevertheless an awesome collection of materials used by Prof Dan Grossman to teach parallelism and concurrency concepts to sophomores at UWash
- 6.824 Distributed Systems MIT
 - MIT's graduate-level DS course with a focus on fault tolerance, replication, and consistency, all taught via awesome lab assignments in Golang!
 - o Assignments Just do git clone git://g.csail.mit.edu/6.824-golabs-2014 6.824
 - o Readings
- <u>6.828</u> Operating Systems MIT
 - MIT's operating systems course focusing on the fundamentals of OS design including booting, memory management, environments, file systems, multitasking, and more. In a series of lab assignments, you will build JOS, an OS exokernel written in C.
 - Assignments
 - Lectures
 - <u>Videos</u> Note: These are student recorded cam videos of the 2011 course. The videos explain a lot of concepts required for the labs and assignments.
- - CSEP552 is a graduate course on distributed systems. Distributed systems have become central to many aspects of how computers are used, from web applications to e-commerce to content distribution. This course will cover abstractions and implementation techniques for the construction of distributed systems, including client server computing, the web, cloud computing, peer-to-peer systems, and distributed storage systems. Topics will include remote procedure call, maintaining consistency of distributed state, fault tolerance, high availability, and other topics. As we believe the best way to learn the material is to build it, there will be a series of hands-on programming projects.
 - <u>Lectures</u> of a previous session are available to watch.
- - The ICS course provides a programmer's view of how computer systems execute programs, store information, and communicate. It enables students to become

more effective programmers, especially in dealing with issues of performance, portability and robustness. It also serves as a foundation for courses on compilers, networks, operating systems, and computer architecture, where a deeper understanding of systems-level issues is required. Topics covered include: machine-level code and its generation by optimizing compilers, performance evaluation and optimization, computer arithmetic, memory organization and management, networking technology and protocols, and supporting concurrent computation.

- This is the must-have course for everyone in CMU who wants to learn some computer science no matter what major are you in. Because it's CMU (The course number is as same as the zip code of CMU)!
- Lecture Notes
- Videos
- Assignments
- <u>15-418</u> Parallel Computer Architecture and Programming Carnegie-Mellon University **III III III**
 - The goal of this course is to provide a deep understanding of the fundamental principles and engineering trade-offs involved in designing modern parallel computing systems as well as to teach parallel programming techniques necessary to effectively utilize these machines. Because writing good parallel programs requires an understanding of key machine performance characteristics, this course will cover both parallel hardware and software design.
 - o **Assignments**
 - Lecture Notes
 - Lecture Videos
 - Readings
- <u>15-440</u> **Distributed Systems** Carnegie-Mellon University **E**
 - o Introduction to distributed systems with a focus on teaching concepts via projects implemented in the Go programming language.
 - Assignments
- <u>15-721</u> **Database Systems** Carnegie-Mellon University 📷 💻 📚
 - This course is a comprehensive study of the internals of modern database management systems. It will cover the core concepts and fundamentals of the components that are used in both high-performance transaction processing systems (OLTP) and large-scale analytical systems (OLAP). The class will stress both efficiency and correctness of the implementation of these ideas. All class projects will be in the context of a real in-memory, multi-core database system. The course is appropriate for graduate students in software systems and for advanced undergraduates with strong systems programming skills.
 - Assignments
 - Lecture Videos
 - Readings
- <u>15-445/645</u> **Database Systems** Carnegie-Mellon University **III**
 - This course covers fundamental topics of DBMS, compared to CMU 15-721.

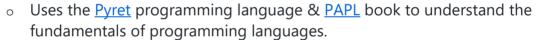
- Assignments
- Lecture Videos
- o Readings
- <u>15-749</u> Engineering Distributed Systems Carnegie-Mellon University 管



- o **Readings**
- 18-447 Introduction to Computer Architecture CMU 📷 💻 📚
 - Very comprehensive material on Computer Architecture definitely more than just "introduction". Online material is very user-friendly, even the recitation videos available online. This is the Spring'15 version by Prof. Onur Mutlu
 - Lectures and Recitation
 - o Homeworks 7 HWs with answer set as well
 - Readings

Programming Languages / Compilers

- <u>CS 75</u> **Principles of Compiler Design** *Swathmore College* **I**
 - Modelled after the influential paper on <u>incremental approach to compiler design</u>, this course teaches how to build a compiler in OCaml
 - Course on Github
 - Notes
- CS 91 Introduction to Programming Languages Swathmore College



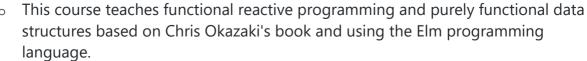
- o <u>Labs</u>
- CIS 194 Introduction to Haskell Penn Engineering
 - Explore the joys of functional programming, using Haskell as a vehicle. The aim of the course will be to allow you to use Haskell to easily and conveniently write practical programs.
 - o Previous semester also available, with more exercises
- CIS 198 Rust Programming UPenn 📝 💻
 - This course covers what makes Rust so unique and applies it to practical systems programming problems. Topics covered include traits and generics; memory safety (move semantics, borrowing, and lifetimes); Rust's rich macro system; closures; and concurrency.
 - Assignments
- Clojure Functional Programming with Clojure University of Helsinki
 - The course is an introduction to functional programming with a dynamically typed language Clojure. We start with an introduction to Clojure; its syntax and development environment. Clojure has a good selection of data structures and we

cover most of them. We also go through the basics of recursion and higher-order functions. The course material is in English.

o Github Page

CMSC 430 Introduction to Compilers Univ of Maryland

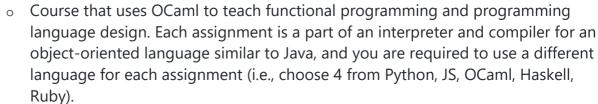
- The goal of CMSC 430 is to arm students with the ability to design, implement, and extend a programming language. Throughout the course, students will design and implement several related languages, and will explore parsing, syntax querying, dataflow analysis, compilation to bytecode, type systems, and language interoperation.
- Lecture Notes
- Assignments
- COS 326 Functional Programming Princeton University
 - Covers functional programming concepts like closures, tail-call recursion & parallelism using the OCaml programming language
 - Lectures
 - Assignments
- CS 143 Compiler construction Stanford University
 - CS143 is a Stanford's course in the practical and theoretical aspects of compiler construction.
 - o Home
 - Syllabus
 - o **Lectures**
 - Assignments
 - o CS143 2011
- 🔹 CS 164 Hack your language! UC Berkeley 🖬 💻 🦻
 - Introduction to programming languages by designing and implementing domainspecific languages.
 - Lecture Videos
 - o Code for Assignments
- CS 173 Programming Languages Brown University 🖬 💻
 - Course by Prof. Krishnamurthi (author of <u>HtDP</u>) and numerous other <u>awesome books</u> on programming languages. Uses a custom designed <u>Pyret</u> programming language to teach the concepts. There was an <u>online</u> <u>class</u> hosted in 2012, which includes all lecture videos for you to enjoy.
 - o Videos
 - o **Assignments**
- CS 223 Purely Functional Data Structures In Elm University of Chicago



- Lectures
- Assignments



- CS 240h Functional Systems in Haskell Stanford University
 - Building software systems in Haskell
 - Lecture Slides
 - o 3 Assignments: Lab1, Lab2, Lab3
- <u>CS 421</u> **Programming Languages and Compilers** *Univ of Illinois, Urbana-Champaign* Course that uses OCaml to teach functional programming and programming language design.
 - o Lectures
 - Videos
 - o **Assignments**
 - o Exams
- CS 3110 Data Structures and Functional Programming Cornell University
 - Another course that uses OCaml to teach alternative programming paradigms, especially functional and concurrent programming.
 - Lecture Slides
 - o **Assignments**
- CS 4120 Introduction to Compilers Cornell University
 - An introduction to the specification and implementation of modern compilers. Topics covered include lexical scanning, parsing, type checking, code generation and translation, an introduction to optimization, and compile-time and run-time support for modern programming languages. As part of the course, students build a working compiler for an object-oriented language.
 - o Syllabus
 - o Lectures
 - Assignments
- CS 4400 Programming Languages Northeastern University 💻 🥟 管
 - This is a course on the study, design, and implementation of programming languages.
 - The course works at two simultaneous levels: first, we will use a programming language that can demonstrate a wide variety of programming paradigms. Second, using this language, we will learn about the mechanics behind programming languages by implementing our own language(s). The two level approach usually means that we will often see how to use a certain feature, and continue by implementing it.
 - Syllabus
 - Lecture Notes/Resources
- CS 4610 Programming Languages and Compilers University of Virginia = 2



Lecture Notes

- Assignments
- CS 5114 Network Programming Languages Cornell University



- This course provides an introduction to the languages used to program computer networks. It will examine recent proposals based on logic, functional, and distributed languages, as well as tools for establishing correctness using automatic solvers, model checkers, and proof assistants.
- **Syllabus**
- Lectures
- CS 5142 Scripting Languages Cornell University



- Perl, PHP, JavaScript, VisualBasic -- they are often-requested skills for employment, but most of us do not have the time to find out what they are all about. In this course, you learn how to use scripting languages for rapid prototyping, web programming, data processing, and application extension. Besides covering traditional programming languages concepts as they apply to scripting (e.g., dynamic typing and scoping), this course looks at new concepts rarely found in traditional languages (e.g., string interpolation, hashes, and polylingual code). Through a series of small projects, you use different languages to achieve programming tasks that highlight the strengths and weaknesses of scripting. As a side effect, you practice teaching yourself new languages.
- **Syllabus**
- Lectures
- **Assignments**
- CS 5470 Compilers University of Utah = 📝 管



- If you're a fan of Prof Matt's writing on his fantastic blog you ought to give this a shot. The course covers the design and implementation of compilers, and it explores related topics such as interpreters, virtual machines and runtime systems. Aside from the Prof's witty take on cheating the page has tons of interesting links on programming languages, parsing and compilers.
- Lecture Notes
- **Projects**
- CS 6118 Types and Semantics Cornell University
 - Types and Semantics is about designing and understand programming languages, whether they be domain specific or general purpose. The goal of this class is to provide a variety of tools for designing custom (programming) languages for whatever task is at hand. Part of that will be a variety of insights on how languages work along with experiences from working with academics and industry on creating new languages such as Ceylon and Kotlin. The class focuses on types and semantics and the interplay between them. This means category theory and constructive type theory (e.g. Coq and richer variations) are ancillary topics of the class. The class also covers unconventional semantic domains such as classical linear type theory in order to both break students from convential thinking and to provide powerful targets capable of formalizing thinks like networking protocols, resource-sensitive computation, and concurrency constructs. The class project is to design and formalize a (programming) language for a purpose of the student's choosing, and assignments are designed to ensure students have had a chance to

practice applying the techniques learned in class before culminating these skills in the class project.

- <u>Syllabus</u>
- Lectures
- CSC 253 CPython internals: A ten-hour codewalk through the Python interpreter source code University of Rochester 📷 📚
 - o Nine lectures walking through the internals of CPython, the canonical Python interpreter implemented in C. They were from the Dynamic Languages and Software Development course taught in Fall 2014 at the University of Rochester.
- CSE 341 Programming Languages University of Washington 📰 💻 🥟
 - o Covers non-imperative paradigms and languages such as Ruby, Racket, and ML and the fundamentals of programming languages.
 - Lectures and Videos
 - **Assignments and Tests**
- CSE P 501 Compiler Construction University of Washington 🖬 💻
 - o Teaches understanding of how a modern compiler is structured and the major algorithms that are used to translate code from high-level to machine language. The best way to do this is to actually build a working compiler, so there will be a significant project to implement one that translates programs written in a core subset of Java into executable x86 assembly language. The compilers themselves will use scanner and parser generator tools and the default implementation language is Java.
 - Lectures
 - Assignments, Tests, and Solutions
- **DMFP** Discrete Mathematics and Functional Programming Wheaton College



- A course that teaches discrete maths concepts with functional programming
- Lecture Videos
- <u>Assignments</u>
- PCPP Practical Concurrent and Parallel Programming IT University of Copenhagen



- In this MSc course you learn how to write correct and efficient concurrent and parallel software, primarily using Java, on standard shared-memory multicore hardware.
- The course covers basic mechanisms such as threads, locks and shared memory as well as more advanced mechanisms such as parallel streams for bulk data, transactional memory, message passing, and lock-free data structures with compare-and-swap.
- It covers concepts such as atomicity, safety, liveness and deadlock.
- It covers how to measure and understand performance and scalability of parallel programs.
- It covers tools and methods to find bugs in concurrent programs.
- 6.945 Adventures in Advanced Symbolic Programming MIT

- Taught by Gerald Sussman of SICP fame, this class deals with concepts and techniques for the design an implementation of large software systems that can be adapted to uses not anticipated by the designer. Applications include compilers, computer-algebra systems, deductive systems, and some artificial intelligence applications.
- Assignments: Extensive programming assignments, using MIT/GNU Scheme. Students should have significant programming experience in Scheme, Common Lisp, Haskell, CAML or other "functional" language.
- Readings
- CS 696 Functional Design and Programming San Diego State University



- Covers functional programming basis using Clojure.
- Topics include testing, functional programming, immutable collections and concurrency.
- Also includes assignments covering Clojurescript, [Reagent](Reagent Github) etc.
- **L28** Advanced Functional Programming University of Cambridge
 - This module aims to teach students how to use the features of modern typed functional programming languages (e.g. OCaml, Haskell) to design and implement libraries and DSLs. It aims to demonstrate how such techniques can improve both correctness and efficiency. Students wishing to take the module should have some experience of a typed functional programming language and an understanding of type inference.
 - This particular session was taught by a prominent OCaml programmer, open Source contributor & author of real world OCaml - Dr Anil Madhavapeddy.

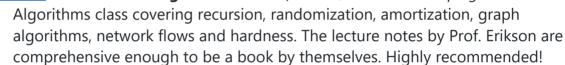
Algorithms

- CS 61B Data Structures UC Berkeley 🖬 💻 🦻 📚
 - o In this course, you will study advanced programming techniques including data structures, encapsulation, abstract data types, interfaces, and algorithms for sorting and searching, and you will get a taste of "software engineering"—the design and implementation of large programs.
 - Full Lecture Materials Lecture of Spring 2016. This website contains full matrials including video links, labs, homeworks, projects. Very good for self-learner. Also a good start for Java. And it includes some other usefull resources for Java Documentation, Data Structure Resources, Git/GitHub and Java Development Resources. Resources
 - <u>Labs</u> The link to labs and projects is included in the website.
 - Lecture Videos
- CS 97SI Introduction to Competitive Programming Stanford University



- Fantastic repository of theory and practice problems across various topics for students who are interested to participate in ACM-ICPC.
- **Lectures and Assignments**

- - CS 224 is an advanced course in algorithm design, and topics we will cover include the word RAM model, data structures, amortization, online algorithms, linear programming, semidefinite programming, approximation algorithms, hashing, randomized algorithms, fast exponential time algorithms, graph algorithms, and computational geometry.
 - <u>Lecture Videos (Youtube)</u>
 - Assignments
- CS 261 A Second Course in Algorithms Stanford University **I**
 - Algorithms for network optimization: max-flow, min-cost flow, matching, assignment, and min-cut problems. Introduction to linear programming. Use of LP duality for design and analysis of algorithms. Approximation algorithms for NPcomplete problems such as Steiner Trees, Traveling Salesman, and scheduling problems. Randomized algorithms. Introduction to online algorithms.
 - Lecture Notes, Videos & Assignments (Youtube)
- CS 473/573 Fundamental Algorithms Univ of Illinois, Urbana-Champaign 📕 🦻



- Lecture Notes
- Labs and Exams
- CS 2150 Program & Data Representation University of Virginia
 - This data structures course introduces C++, linked-lists, stacks, queues, trees, numerical representation, hash tables, priority queues, heaps, huffman coding, graphs, and x86 assembly.
 - Lectures
 - Assignments
- CS 4820 Introduction to Analysis of Algorithms Cornell University 💻 🥟 管
 - o This course develops techniques used in the design and analysis of algorithms, with an emphasis on problems arising in computing applications. Example applications are drawn from systems and networks, artificial intelligence, computer vision, data mining, and computational biology. This course covers four major algorithm design techniques (greedy algorithms, divide and conquer, dynamic programming, and network flow), computability theory focusing on undecidability, computational complexity focusing on NP-completeness, and algorithmic techniques for intractable problems, including identification of structured special cases, approximation algorithms, and local search heuristics.
 - <u>Lectures</u>
 - Assignments
 - Syllabus
- <u>CSCI 104</u> Data Structures and Object Oriented Design University of Southern California (USC)
 - Lectures
 - o Labs

- Assignments
- Additional Resources
- CSCI 135 Software Design and Analysis I P CUNY Hunter College
 - o It is currently an intensive introduction to program development and problem solving. Its emphasis is on the process of designing, implementing, and evaluating small-scale programs. It is not supposed to be a C++ programming course, although much of the course is spent on the details of C++. C++ is an extremely large and complex programming language with many features that interact in unexpected ways. One does not need to know even half of the language to use it well.
 - Lectures and Assignments
- CSCI 235 Software Design and Analysis II CUNY Hunter College
 - o Introduces algorithms for a few common problems such as sorting. Practically speaking, it furthers the students' programming skills with topics such as recursion, pointers, and exception handling, and provides a chance to improve software engineering skills and to give the students practical experience for more productive programming.
 - Lectures and Assignments
- CSCI 335 Software Design and Analysis III Programme College
 - This includes the introduction of hashes, heaps, various forms of trees, and graphs.
 It also revisits recursion and the sorting problem from a higher perspective than was presented in the prequels. On top of this, it is intended to introduce methods of algorithmic analysis.
 - <u>Lectures and Assignments</u>
- CSE 331 Software Design and Implementation University of Washington
 - Explores concepts and techniques for design and construction of reliable and maintainable software systems in modern high-level languages; program structure and design; program-correctness approaches, including testing.
 - Lectures, Assignments, and Exams
- CSE 373 Analysis of Algorithms Stony Brook University 🖬 💻 🥟 😉
 - Prof Steven Skiena's no stranger to any student when it comes to algorithms. His seminal book has been touted by many to be best for getting that job in Google. In addition, he's also well-known for tutoring students in competitive programming competitions. If you're looking to brush up your knowledge on Algorithms, you can't go wrong with this course.
 - Lecture Videos
- ECS 122A Algorithm Design and Analysis UC Davis 🖬 💻 🦻
 - Taught by <u>Dan Gusfield</u> in 2010, this course is an undergraduate introduction to algorithm design and analysis. It features traditional topics, such as Big Oh notation, as well as an importance on implementing specific algorithms. Also featured are sorting (in linear time), graph algorithms, depth-first search, string matching, dynamic programming, NP-completeness, approximation, and randomization.
 - Syllabus

- **Lecture Videos**
- **Assignments**

ECS 222A Graduate Level Algorithm Design and Analysis UC Davis 📰 💻 🥟



This is the graduate level complement to the ECS 122A undergraduate algorithms course by **Dan Gusfield** in 2011. It assumes an undergrad course has already been taken in algorithms, and, while going over some undergraduate algorithms topics, focuses more on increasingly complex and advanced algorithms.

- Lecture Videos
- **Syllabus**
- **Assignments**

6.INT Hacking a Google Interview MIT

- This course taught in the MIT Independent Activities Period in 2009 goes over common solution to common interview questions for software engineer interviews at highly selective companies like Apple, Google, and Facebook. They cover time complexity, hash tables, binary search trees, and other common algorithm topics you should have already covered in a different course, but goes more in depth on things you wouldn't otherwise learn in class-like bitwise logic and problem solving tricks.
- Handouts
- **Topics Covered**

6.006 Introduction to Algorithms MIT

- This course provides an introduction to mathematical modeling of computational problems. It covers the common algorithms, algorithmic paradigms, and data structures used to solve these problems. The course emphasizes the relationship between algorithms and programming, and introduces basic performance measures and analysis techniques for these problems. This course provides an introduction to mathematical modeling of computational problems. It covers the common algorithms, algorithmic paradigms, and data structures used to solve these problems. The course emphasizes the relationship between algorithms and programming, and introduces basic performance measures and analysis techniques for these problems.
- **Lecture Videos**
- **Assignments**
- **Readings**
- Resources
- **Old Exams**

6.046J/18.410J Design and Analysis of Algorithms MIT 📷 🥟 💻 📚

This is an intermediate algorithms course with an emphasis on teaching techniques for the design and analysis of efficient algorithms, emphasizing methods of application. Topics include divide-and-conquer, randomization, dynamic programming, greedy algorithms, incremental improvement, complexity, and cryptography. This course assumes that students know how to analyze simple algorithms and data structures from having taken 6.006. It introduces students to the design of computer algorithms, as well as analysis of sophisticated algorithms.

- Lecture Videos
- Lecture Notes
- o **Assignments**
- Resources
- o Old Exams

• <u>6.851</u> Advanced Data Structures MIT **III III**

- This is an advanced DS course, you must be done with the <u>Advanced Algorithms</u> course before attempting this one.
- Lectures Contains videos from sp2012 version, but there isn't much difference.
- Assignments contains the calendar as well.

• <u>6.854/18.415J</u> Advanced Algorithms *MIT* **III III**

- Advanced course in algorithms by Dr. David Karger covering topics such as amortization, randomization, fingerprinting, word-level parallelism, bit scaling, dynamic programming, network flow, linear programming, fixed-parameter algorithms, and approximation algorithms.
- Register on <u>NB</u> to access the <u>problem set and lectures</u>.

• 6.854J/18.415J Advanced Algorithms MIT 🔊 💻 😉

- This course is a first-year graduate course in algorithms. Emphasis is placed on fundamental algorithms and advanced methods of algorithmic design, analysis, and implementation. Techniques to be covered include amortization, randomization, fingerprinting, word-level parallelism, bit scaling, dynamic programming, network flow, linear programming, fixed-parameter algorithms, and approximation algorithms. Domains include string algorithms, network optimization, parallel algorithms, computational geometry, online algorithms, external memory, cache, and streaming algorithms, and data structures. The need for efficient algorithms arises in nearly every area of computer science. But the type of problem to be solved, the notion of what algorithms are "efficient," and even the model of computation can vary widely from area to area. In this second class in algorithms, we will survey many of the techniques that apply broadly in the design of efficient algorithms, and study their application in a wide range of application domains and computational models. The goal is for the class to be broad rather than deep. Our plan is to touch upon the following areas. This is a tentative list of topics that might be covered in the class; we will select material adaptively based on the background, interests, and rate of progress of the students.
- <u>Lecture Videos Spring 2016</u>
- Lecture Notes
- Assignments
- Readings
- Resources

15-451/651 Algorithms Carnegie Mellon University

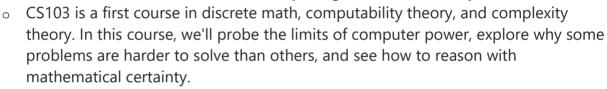
o The required algorithms class that go in depth into all basic algorithms and the proofs behind them. This is one of the heavier algorithms curriculums on this page. Taught by Avrim Blum and Manuel Blum who has a Turing Award due to his

contributions to algorithms. Course link includes a very comprehensive set of reference notes by Avrim Blum.

- 16s-4102 Algorithms University of Virginia
 - Lecture Videos & Homeworks (Youtube)

CS Theory

- CIS 500 Software Foundations University of Pennsylvania
 - An introduction to formal verification of software using the Coq proof assistant.
 Topics include basic concepts of logic, computer-assisted theorem proving, functional programming, operational semantics, Hoare logic, and static type systems.
 - <u>Lectures and Assignments</u>
 - Textbook
- CS 103 Mathematical Foundations of Computing Stanford University



- o Links to all lectures notes and assignments are directly on the course page
- CS 173 Discrete Structures Univ of Illinois Urbana-Champaign 📕 🥡 😉
 - This course is an introduction to the theoretical side of computer science. In it, you
 will learn how to construct proofs, read and write literate formal mathematics, get
 a quick introduction to key theory topics and become familiar with a range of
 standard mathematics concepts commonly used in computer science.
 - o Textbook Written by the professor. Includes Instructor's Guide.
 - o <u>Assignments</u>
 - o Exams
- CS 276 Foundations of Cryptography UC Berkeley
 - This course discusses the complexity-theory foundations of modern cryptography, and looks at recent results in the field such as Fully Homomorphic Encryption, Indistinguishability Obfuscation, MPC and so on.
- CS 278 Complexity Theory UC Berkeley
 - o A graduate level course on complexity theory that introduces P vs NP, the power of randomness, average-case complexity, hardness of approximation, and so on.
- - CS 498 section 374 (unofficially "CS 374") covers fundamental tools and techniques from theoretical computer science, including design and analysis of algorithms, formal languages and automata, computability, and complexity.
 Specific topics include regular and context-free languages, finite-state automata, recursive algorithms (including divide and conquer, backtracking, dynamic

programming, and greedy algorithms), fundamental graph algorithms (including depth- and breadth-first search, topological sorting, minimum spanning trees, and shortest paths), undecidability, and NP-completeness. The course also has a strong focus on clear technical communication.

- Assignments/Exams
- Lecture Notes/Labs
- Lecture videos

CS 3110 Data Structures and Functional Programming Cornell University



CS 3110 (formerly CS 312) is the third programming course in the Computer Science curriculum, following CS 1110/1112 and CS 2110. The goal of the course is to help students become excellent programmers and software designers who can design and implement software that is elegant, efficient, and correct, and whose code can be maintained and reused.

- **Syllabus**
- Lectures
- <u>Assignments</u>
- CS 3220 Introduction to Scientific Computing Cornell University



- In this one-semester survey course, we introduce numerical methods for solving linear and nonlinear equations, interpolating data, computing integrals, and solving differential equations, and we describe how to use these tools wisely (we hope!) when solving scientific problems.
- **Syllabus**
- Lectures
- **Assignments**
- CS 4300 Information Retrieval Cornell University



Studies the methods used to search for and discover information in large-scale systems. The emphasis is on information retrieval applied to textual materials, but there is some discussion of other formats. The course includes techniques for searching, browsing, and filtering information and the use of classification systems and thesauruses. The techniques are illustrated with examples from web searching and digital libraries.

- <u>Syllabus</u>
- **Lectures**
- **Assignments**
- CS 4810 Introduction to Theory of Computing Cornell University



- This undergraduate course provides a broad introduction to the mathematical foundations of computer science. We will examine basic computational models, especially Turing machines. The goal is to understand what problems can or cannot be solved in these models.
- Syllabus
- Lectures
- **Assignments**
- CS 6810 Theory of Computing Cornell University

- This graduate course gives a broad introduction to complexity theory, including classical results and recent developments. Complexity theory aims to understand the power of efficient computation (when computational resources like time and space are limited). Many compelling conceptual questions arise in this context. Most of these questions are (surprisingly?) difficult and far from being resolved. Nevertheless, a lot of progress has been made toward understanding them (and also why they are difficult). We will learn about these advances in this course. A theme will be combinatorial constructions with random-like properties, e.g., expander graphs and error-correcting codes. Some examples:
 - Is finding a solution inherently more difficult than verifying it?
 - Do more computational resources mean more computing power?
 - Is it easier to find approximate solutions than exact ones?
 - Are randomized algorithms more powerful than deterministic ones?
 - Is it easier to solve problems in the average case than in the worst case?
 - Are quantum computers more powerful than classical ones?
- **Syllabus**
- Lectures
- **Assignments**
- CSCE 3193 Programming Paradigms University of Arkansas (Fayetteville)



- Programming in different paradigms with emphasis on object oriented programming, network programming and functional programming. Survey of programming languages, event driven programming, concurrency, software validation.
- **Syllabus**
- **Notes**
- **Assignments**
- **Practice Exams**
- 6.045 Great Ideas in Theoretical Computer Science MIT 📝 🖬 隓



- This course provides a challenging introduction to some of the central ideas of theoretical computer science. Beginning in antiquity, the course will progress through finite automata, circuits and decision trees, Turing machines and computability, efficient algorithms and reducibility, the P versus NP problem, NPcompleteness, the power of randomness, cryptography and one-way functions, computational learning theory, and quantum computing. It examines the classes of problems that can and cannot be solved by various kinds of machines. It tries to explain the key differences between computational models that affect their power.
- **Syllabus**
- **Lecture Notes**
- **Lecture Videos**

Introduction to CS

- CS 10 The Beauty and Joy of Computing UC Berkeley 📷 💻 🦻
 - CS10 is UCB's introductory computer science class, taught using the beginners' drag-and-drop language. Students learn about history, social implications, great principles, and future of computing. They also learn the joy of programming a computer using a friendly, graphical language, and will complete a substantial team programming project related to their interests.
 - o Snap*!* (based on Scratch by MIT).
 - o Curriculum
- - CS50x is Harvard College's introduction to the intellectual enterprises of computer science and the art of programming for majors and non-majors alike, with or without prior programming experience. An entry-level course taught by David J. Malan.
 - Lectures
 - Problem Sets
 - The course can also be taken from edX.
- - o In CS 61A, we are interested in teaching you about programming, not about how to use one particular programming language. We consider a series of techniques for controlling program complexity, such as functional programming, data abstraction, and object-oriented programming. Mastery of a particular programming language is a very useful side effect of studying these general techniques. However, our hope is that once you have learned the essence of programming, you will find that picking up a new programming language is but a few days' work.
 - Lecture Resources by Type
 - <u>Lecture Resources by Topic</u>
 - Additional Resources
 - Practice Problems
 - Extra Lectures
- <u>CS 61AS</u> Structure & Interpretation of Computer Programs [Racket] *UC Berkeley*
 - A self-paced version of the CS61 Course but in Racket / Scheme. 61AS is a great introductory course that will ease you into all the amazing concepts that future CS courses will cover, so remember to keep an open mind, have fun, and always respect the data abstraction
 - o Lecture Videos
 - o Assignments and Notes
- - CS101 teaches the essential ideas of Computer Science for a zero-prior-experience audience. Participants play and experiment with short bits of "computer code" to bring to life to the power and limitations of computers.

- Lectures videos will available for free after registration.
- - This course is the largest of the introductory programming courses and is one of the largest courses at Stanford. Topics focus on the introduction to the engineering of computer applications emphasizing modern software engineering principles: object-oriented design, decomposition, encapsulation, abstraction, and testing. Programming Methodology teaches the widely-used Java programming language along with good software engineering principles.
 - Lecture Videos
 - Assignments
 - o All materials in a zip file
- CS 106B Programming Abstractions Stanford University **T**
 - This course is the natural successor to Programming Methodology and covers such advanced programming topics as recursion, algorithmic analysis, and data abstraction using the C++ programming language, which is similar to both C and Java.
 - Lectures
 - Assignments
 - o All materials in a zip file
- CS 107 Programming Paradigms Stanford University 🖬 💻 🦻
 - Topics: Advanced memory management features of C and C++; the differences between imperative and object-oriented paradigms. The functional paradigm (using LISP) and concurrent programming (using C and C++)
 - Lectures
 - Assignments
- CS 109 Programming Practice Using Scala KAIST \(\bar{\textit{M}} \)
 - This course introduces basic concepts of programming and computer science, such as dynamic and static typing, dynamic memory allocation, objects and methods, binary representation of numbers, using an editor and compiler from the command line, running programs with arguments from the command line, using libraries, and the use of basic data structures such as arrays, lists, sets, and maps. We will use Scala for this course.
 - o [Lectures] (http://otfried.org/courses/cs109/index.html)
 - [Assignments] (http://otfried.org/courses/cs109/index.html)
- CS 1109 Fundamental Programming Concepts Cornell University
 - This course provides an introduction to programming and problem solving using a high-level programming language. It is designed to increase your knowledge level to comfortably continue to courses CS111x. Our focus will be on generic programming concepts: variables, expressions, control structures, loops, arrays, functions, pseudocode and algorithms. You will learn how to analyze problems and convert your ideas into solutions interpretable by computers. We will use MATLAB; because it provides a productive environment, and it is widely used by all engineering communities.
 - o <u>Syllabus</u>

- Lectures
- **Assignments**
- CS 1110 Introduction to Computing Using Python Cornell University



Programming and problem solving using Python. Emphasizes principles of software development, style, and testing. Topics include procedures and functions, iteration, recursion, arrays and vectors, strings, an operational model of procedure and function calls, algorithms, exceptions, object-oriented programming, and GUIs (graphical user interfaces). Weekly labs provide guided practice on the computer, with staff present to help. Assignments use graphics and GUIs to help develop fluency and understanding.

- **Syllabus**
- Lectures
- **Assignments**
- CS 1112 Introduction to Computing Using Matlab Cornell University



- Programming and problem solving using MATLAB. Emphasizes the systematic development of algorithms and programs. Topics include iteration, functions, arrays and vectors, strings, recursion, algorithms, object-oriented programming, and MATLAB graphics. Assignments are designed to build an appreciation for complexity, dimension, fuzzy data, inexact arithmetic, randomness, simulation, and the role of approximation. NO programming experience is necessary; some knowledge of Calculus is required.
- **Syllabus**
- Lectures
- **Assignments**
- **Projects**
- **CS 1115** Introduction to Computational Science and Engineering Using Matlab Graphical User Interfaces Cornell University
 - Programming and problem solving using MATLAB. Emphasizes the systematic development of algorithms and programs. Topics include iteration, functions, arrays and vectors, strings, recursion, algorithms, object-oriented programming, and MATLAB graphics. Assignments are designed to build an appreciation for complexity, dimension, fuzzy data, inexact arithmetic, randomness, simulation, and the role of approximation. NO programming experience is necessary; some knowledge of Calculus is required.
 - **Syllabus**
 - Lectures
 - **Projects**
- CS 1130 Transition to OO Programming Cornell University
 - Introduction to object-oriented concepts using Java. Assumes programming knowledge in a language like MATLAB, C, C++, or Fortran. Students who have learned Java but were not exposed heavily to OO programming are welcome.
 - <u>Syllabus</u>
 - Lectures

- <u>Assignments</u>
- CS 1133 Transition to Python Cornell University
 - Introduction to the Python programming language. Covers the basic programming constructs of Python, including assignment, conditionals, iteration, functions, object-oriented design, arrays, and vectorized computation. Assumes programming knowledge in a language like Java, Matlab, C, C++, or Fortran.
 - **Syllabus**
 - Lectures
 - <u>Assignments</u>
- CS 1410-2 and CS2420-20 Computer Science I and II for Hackers University of Utah 📕 🦻 🝃
 - o An intro course in the spirit of SICP designed by Professor Matthew Flatt (one of the lead designers of Racket and author of HtDP). Mostly Racket and C, and a bit of Java, with explanations on how high level functional programming concepts relate to the design of OOP programs. Do this one before SICP if SICP is a bit too much...
 - <u>Lectures and Assignments 1</u>
 - <u>Lectures and Assignments 2</u>
 - Textbook
 - Racket Language
- CS 2110 Object-Oriented Programming and Data Structures Cornell University



- CS 2110 is an intermediate-level programming course and an introduction to computer science. Topics include program design and development, debugging and testing, object-oriented programming, proofs of correctness, complexity analysis, recursion, commonly used data structures, graph algorithms, and abstract data types. Java is the principal programming language. The course syllabus can easily be extracted by looking at the link to lectures.
- <u>Syllabus</u>
- Lectures
- **Assignments**
- CS 4302 Web Information Systems Cornell University **E**
 - This course will introduce you to technologies for building data-centric information systems on the World Wide Web, show the practical applications of such systems, and discuss their design and their social and policy context by examining cross-cutting issues such as citizen science, data journalism and open government. Course work involves lectures and readings as well as weekly homework assignments, and a semester-long project in which the students demonstrate their expertise in building data-centric Web information systems.
 - <u>Syllabus</u>
 - <u>Lectures</u>
 - **Assignments**
- CSCE 2004 Programming Foundations I University of Arkansas (Fayetteville) 💻 📝 🛎



- Introductory course for students majoring in computer science or computer engineering. Software development process: problem specification, program design, implementation, testing and documentation. Programming topics: data representation, conditional and iterative statements, functions, arrays, strings, file I/O, and classes. Using C++ in a UNIX environment.
- o <u>Syllabus</u>
- Notes
- o **Assignments**
- Practice Exams
- - This course is all about understanding: understanding what's going on inside your computer when you flip on the switch, why tech support has you constantly rebooting your computer, how everything you do on the Internet can be watched by others, and how your computer can become infected with a worm just by being turned on. Designed for students who use computers and the Internet every day but don't fully understand how it all works, this course fills in the gaps. Through lectures on hardware, software, the Internet, multimedia, security, privacy, website development, programming, and more, this course "takes the hood off" of computers and the Internet so that students understand how it all works and why. Through discussions of current events, students are exposed also to the latest technologies.
 - Lecture Videos
 - o <u>Syllabus</u>
 - Notes / Recaps
 - Assignments
- <u>CS-for-all</u> **CS for All** Harvey Mudd College **=** 🥒 📚
 - This book (and course) takes a unique approach to "Intro CS." In a nutshell, our objective is to provide an introduction to computer science as an intellectually rich and vibrant field rather than focusing exclusively on computer programming. While programming is certainly an important and pervasive element of our approach, we emphasize concepts and problem-solving over syntax and programming language features.
 - Lectures and Other resources
- - Teaches big-picture computing concepts using the Scheme programming language. Students will implement programs in a variety of different programming paradigms (functional, object-oriented, logical). Heavy emphasis on function composition, code-as-data, control abstraction with continuations, and syntactic abstraction through macros. An excellent course if you are looking to build a mental framework on which to hang your programming knowledge.
 - o **Lectures**
 - o Textbook (epub, pdf)
 - o IDE

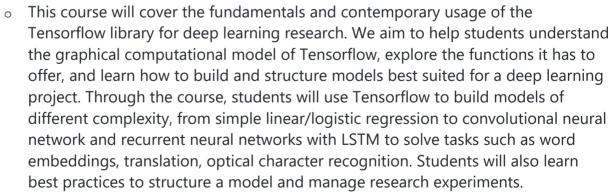
- 6.005 Software Construction, Fall 2016 MIT
 - This course introduces fundamental principles and techniques of software development. Students learn how to write software that is safe from bugs, easy to understand, and ready for change. Topics include specifications and invariants; testing, test-case generation, and coverage; state machines; abstract data types and representation independence; design patterns for object-oriented programming; concurrent programming, including message passing and shared concurrency, and defending against races and deadlock; and functional programming with immutable data and higher-order functions.
 - **Lectures Notes/Assignments**

Machine Learning

<u>DEEPNLP</u> Deep Learning for Natural Language Processing University of Oxford



- o This is an applied course focussing on recent advances in analysing and generating speech and text using recurrent neural networks. We introduce the mathematical definitions of the relevant machine learning models and derive their associated optimisation algorithms. The course covers a range of applications of neural networks in NLP including analysing latent dimensions in text, transcribing speech to text, translating between languages, and answering questions. This course is organised by Phil Blunsom and delivered in partnership with the **DeepMind Natural Language Research Group**.
- Lectures
- Assignments are available on the organisation page titled as "practicals"
- CS20si Tensorflow for Deep Learning Research Stanford University



- Assignments available on Github.
- COMS 4771 Machine Learning Columbia University
 - o Course taught by Tony Jebara introduces topics in Machine Learning for both generative and discriminative estimation. Material will include least squares methods, Gaussian distributions, linear classification, linear regression, maximum likelihood, exponential family distributions, Bayesian networks, Bayesian inference, mixture models, the EM algorithm, graphical models, hidden Markov models, support vector machines, and kernel methods.
 - **Lectures and Assignments**

- CS 109 Data Science Harvard University **P** 🥟 管
 - Learning from data in order to gain useful predictions and insights. This course introduces methods for five key facets of an investigation: data wrangling, cleaning, and sampling to get a suitable data set; data management to be able to access big data quickly and reliably; exploratory data analysis to generate hypotheses and intuition; prediction based on statistical methods such as regression and classification; and communication of results through visualization, stories, and interpretable summaries.
 - Lectures
 - Slides
 - Labs and Assignments
 - o 2014 Lectures
 - o 2013 Lectures (slightly better)
- CS 156 Learning from Data Caltech 🖬 💻 🍞 管
 - This is an introductory course in machine learning (ML) that covers the basic theory, algorithms, and applications. ML is a key technology in Big Data, and in many financial, medical, commercial, and scientific applications. It enables computational systems to adaptively improve their performance with experience accumulated from the observed data. ML has become one of the hottest fields of study today, taken up by undergraduate and graduate students from 15 different majors at Caltech. This course balances theory and practice, and covers the mathematical as well as the heuristic aspects.
 - Lectures
 - Homework
 - Textbook
- <u>CS 224d</u> **Deep Learning for Natural Language Processing** *Stanford University*
 - Natural language processing (NLP) is one of the most important technologies of the information age. Understanding complex language utterances is also a crucial part of artificial intelligence. Applications of NLP are everywhere because people communicate most everything in language: web search, advertisement, emails, customer service, language translation, radiology reports, etc. There are a large variety of underlying tasks and machine learning models powering NLP applications. Recently, deep learning approaches have obtained very high performance across many different NLP tasks. These models can often be trained with a single end-to-end model and do not require traditional, task-specific feature engineering. In this spring quarter course students will learn to implement, train, debug, visualize and invent their own neural network models. The course provides a deep excursion into cutting-edge research in deep learning applied to NLP.
 - o Syllabus
 - Lectures and Assignments

- Big data is data so large that it does not fit in the main memory of a single machine, and the need to process big data by efficient algorithms arises in Internet search, network traffic monitoring, machine learning, scientific computing, signal processing, and several other areas. This course will cover mathematically rigorous models for developing such algorithms, as well as some provable limitations of algorithms operating in those models.
- <u>Lectures</u> (Youtube)
- o **Assignments**
- <u>CS 231n</u> Convolutional Neural Networks for Visual Recognition Stanford University
 - Computer Vision has become ubiquitous in our society, with applications in search, image understanding, apps, mapping, medicine, drones, and self-driving cars. This course is a deep dive into details of the deep learning architectures with a focus on learning end-to-end models for these tasks, particularly image classification. During the 10-week course, students will learn to implement, train and debug their own neural networks and gain a detailed understanding of cutting-edge research in computer vision.
 - Lecture Notes
 - Lecture Videos
 - Github Page
- CS 287 Advanced Robotics UC Berkeley
 - The course introduces the math and algorithms underneath state-of-the-art robotic systems. The majority of these techniques are heavily based on probabilistic reasoning and optimization---two areas with wide applicability in modern Artificial Intelligence. An intended side-effect of the course is to generally strengthen your expertise in these two areas.
 - Lectures Notes
 - o **Assignments**
- CS 395T Statistical and Discrete Methods for Scientific Computing University of Texas
 - o Practical course in applying modern statistical techniques to real data, particularly bioinformatic data and large data sets. The emphasis is on efficient computation and concise coding, mostly in MATLAB and C++. Topics covered include probability theory and Bayesian inference; univariate distributions; Central Limit Theorem; generation of random deviates; tail (p-value) tests; multiple hypothesis correction; empirical distributions; model fitting; error estimation; contingency tables; multivariate normal distributions; phylogenetic clustering; Gaussian mixture models; EM methods; maximum likelihood estimation; Markov Chain Monte Carlo; principal component analysis; dynamic programming; hidden Markov models; performance measures for classifiers; support vector machines; Wiener filtering; wavelets; multidimensional interpolation; information theory.
 - Lectures and Assignments
- CS 4780 Machine Learning Cornell University
 - This course will introduce you to technologies for building data-centric information systems on the World Wide Web, show the practical applications of

such systems, and discuss their design and their social and policy context by examining cross-cutting issues such as citizen science, data journalism and open government. Course work involves lectures and readings as well as weekly homework assignments, and a semester-long project in which the students demonstrate their expertise in building data-centric Web information systems.

- Syllabus
- Lectures
- CS 4786 Machine Learning for Data Science Cornell University
 - An introductory course in machine learning, with a focus on data modeling and related methods and learning algorithms for data sciences. Tentative topic list:
 - Dimensionality reduction, such as principal component analysis (PCA) and the singular value decomposition (SVD), canonical correlation analysis (CCA), independent component analysis (ICA), compressed sensing, random projection, the information bottleneck. (We expect to cover some, but probably not all, of these topics).
 - Clustering, such as k-means, Gaussian mixture models, the expectation-maximization (EM) algorithm, link-based clustering. (We do not expect to cover hierarchical or spectral clustering.).
 - Probabilistic-modeling topics such as graphical models, latent-variable models, inference (e.g., belief propagation), parameter learning.
 - Regression will be covered if time permits.
 - Assignments
 - Lectures
- CVX 101 Convex Optimization Stanford University **E**
 - The course concentrates on recognizing and solving convex optimization problems that arise in applications. Topics addressed include the following. Convex sets, functions, and optimization problems. Basics of convex analysis. Least-squares, linear and quadratic programs, semidefinite programming, minimax, extremal volume, and other problems. Optimality conditions, duality theory, theorems of alternative, and applications. Interior-point methods. Applications to signal processing, statistics and machine learning, control and mechanical engineering, digital and analog circuit design, and finance.
 - o <u>Textbook</u>
 - <u>Lectures and Assignments</u>
- - This increasingly popular course is taught through the Data Science Center at NYU. Originally introduced by <u>Yann Lecun</u>, it is now led by <u>Zaid Harchaoui</u>, although Prof. Lecun is rumored to still stop by from time to time. It covers the theory, technique, and tricks that are used to achieve very high accuracy for machine learning tasks in computer vision and natural language processing. The assignments are in Lua and hosted on Kaggle.
 - Course Page
 - Recorded Lectures

• <u>EECS E6893 & EECS E6895</u> Big <u>Data Analytics & Advanced Big Data</u>

Analytics Columbia University = 💆 💆

- Students will gain knowledge on analyzing Big Data. It serves as an introductory course for graduate students who are expecting to face Big Data storage, processing, analysis, visualization, and application issues on both workplaces and research environments.
- Taught by <u>Dr. Ching-Yung Lin</u>
- o Course Site
- Assignments Assignments are present in the Course Slides
- <u>EECS E6894</u> **Deep Learning for Computer Vision and Natural Language Processing** *Columbia University*
 - This graduate level research class focuses on deep learning techniques for vision and natural language processing problems. It gives an overview of the various deep learning models and techniques, and surveys recent advances in the related fields. This course uses Theano as the main programming tool. GPU programming experiences are preferred although not required. Frequent paper presentations and a heavy programming workload are expected.
 - Readings
 - o **Assignments**
 - Lecture Notes
- <u>EE103</u> Introduction to Matrix Methods Stanford University **= ****
 - The course covers the basics of matrices and vectors, solving linear equations, least-squares methods, and many applications. It'll cover the mathematics, but the focus will be on using matrix methods in applications such as tomography, image processing, data fitting, time series prediction, finance, and many others. EE103 is based on a book that Stephen Boyd and Lieven Vandenberghe are currently writing. Students will use a new language called Julia to do computations with matrices and vectors.
 - Lectures
 - o Book
 - o **Assignments**
 - o Code
- Info 290 Analyzing Big Data with Twitter UC Berkeley school of information 📷
 - In this course, UC Berkeley professors and Twitter engineers provide lectures on the most cutting-edge algorithms and software tools for data analytics as applied to Twitter's data. Topics include applied natural language processing algorithms such as sentiment analysis, large scale anomaly detection, real-time search, information diffusion and outbreak detection, trend detection in social streams, recommendation algorithms, and advanced frameworks for distributed computing.
 - Lecture Videos
 - o Previous Years coursepage

- The course focusses on neural networks and uses the <u>Torch</u> deep learning library (implemented in Lua) for exercises and assignments. Topics include: logistic regression, back-propagation, convolutional neural networks, max-margin learning, siamese networks, recurrent neural networks, LSTMs, hand-writing with recurrent neural networks, variational autoencoders and image generation and reinforcement learning
- <u>Lectures and Assignments</u>
- Source code
- StatLearning Intro to Statistical Learning Stanford University 📕 🥟 管 📹
 - This is an introductory-level course in supervised learning, with a focus on regression and classification methods. The syllabus includes: linear and polynomial regression, logistic regression and linear discriminant analysis; cross-validation and the bootstrap, model selection and regularization methods (ridge and lasso); nonlinear models, splines and generalized additive models; tree-based methods, random forests and boosting; support-vector machines.
 - The lectures cover all the material in <u>An Introduction to Statistical Learning</u>, with <u>Applications in R</u> which is a more approachable version of the <u>Elements of Statistical Learning</u> (or ESL) book.
- <u>10-601</u> Machine Learning Carnegie Mellon University **E** 📝 管 📹
 - o This course covers the theory and practical algorithms for machine learning from a variety of perspectives. It covers topics such as Bayesian networks, decision tree learning, Support Vector Machines, statistical learning methods, unsupervised learning and reinforcement learning. The course covers theoretical concepts such as inductive bias, the PAC learning framework, Bayesian learning methods, margin-based learning, and Occam's Razor. Short programming assignments include hands-on experiments with various learning algorithms. This course is designed to give a graduate-level student a thorough grounding in the methodologies, technologies, mathematics and algorithms currently needed by people who do research in machine learning.
 - o Taught by one of the leading experts on Machine Learning Tom Mitchell
 - Lectures
 - o Project Ideas and Datasets
- <u>10-708</u> Probabilistic Graphical Models Carnegie Mellon University 🖬 🥟 💻 答
 - Many of the problems in artificial intelligence, statistics, computer systems, computer vision, natural language processing, and computational biology, among many other fields, can be viewed as the search for a coherent global conclusion from local information. The probabilistic graphical models framework provides a unified view for this wide range of problems, enabling efficient inference, decision-making and learning in problems with a very large number of attributes and huge datasets. This graduate-level course will provide you with a strong foundation for both applying graphical models to complex problems and for addressing core research topics in graphical models.
 - Lecture Videos
 - Assignments
 - Lecture notes

- Readings
- <u>11-785</u> **Deep Learning** Carnegie Mellon University **=** 隓
 - o The course presents the subject through a series of seminars and labs, which will explore it from its early beginnings, and work themselves to some of the state of the art. The seminars will cover the basics of deep learning and the underlying theory, as well as the breadth of application areas to which it has been applied, as well as the latest issues on learning from very large amounts of data. We will concentrate largely, although not entirely, on the connectionist architectures that are most commonly associated with it. *Lectures* and *Reading Notes* are available on the page.
- CS246 Mining Massive Data Sets Stanford University 🖬 🥟 💻 📚
 - o The course will discuss data mining and machine learning algorithms for analyzing very large amounts of data. The emphasis will be on Map Reduce as a tool for creating parallel algorithms that can process very large amounts of data.
 - Lecture Videos
 - Assignments
 - <u>Lecture notes</u>
 - Readings
- CS276 Information Retrieval and Web Search Stanford University 📝 📽
 - Basic and advanced techniques for text-based information systems: efficient text indexing; Boolean and vector space retrieval models; evaluation and interface issues; Web search including crawling, link-based algorithms, and Web metadata; text/Web clustering, classification; text mining.
 - Lecture notes
 - Readings
- Practical RL Reinforcement Learning in the Wild Yandex SDA 🖬 🥟 💻 📚
 - A course on reinforcement learning in the wild. Taught on-campus in HSE and Yandex SDA (russian) and maintained to be friendly to online students (both english and russian).
 - o Syllabus

Security

- CIS 4930 / CIS 5930 Offensive Computer Security Florida State University 🖬 💻 🦻
 - Course taught by <u>W. Owen Redwood</u> and <u>Xiuwen Liu</u>. It covers a wide range of computer security topics, starting from Secure C Coding and Reverse Engineering to Penetration Testing, Exploitation and Web Application Hacking, both from the defensive and the offensive point of view.
 - <u>Lectures and Videos</u>
 - Assignments
- CS 155 Computer and Network Security Stanford > Security Stanford



Principles of computer systems security. Attack techniques and how to defend against them. Topics include: network attacks and defenses, operating system holes, application security (web, email, databases), viruses, social engineering attacks, privacy, and digital rights management. Course projects focus on building reliable code. Recommended: Basic Unix. Primarily intended for seniors and firstyear graduate students.

• CS 161 Computer Security UC Berkeley

 Introduction to computer security. Cryptography, including encryption, authentication, hash functions, cryptographic protocols, and applications.
 Operating system security, access control. Network security, firewalls, viruses, and worms. Software security, defensive programming, and language-based security. Case studies from real-world systems.

• CS 259 Security Modeling and Analysis Stanford 🥟 💻 管

The course will cover a variety of contemporary network protocols and other systems with security properties. The course goal is to give students hands-on experience in using automated tools and related techniques to analyze and evaluate security mechanisms. To understand security properties and requirements, we will look at several network protocols and their properties, including secrecy, authentication, key establishment, and fairness. In parallel, the course will look at several models and tools used in security analysis and examine their advantages and limitations. In addition to fully automated finite-state model checking techniques, we will also study other approaches, such as constraint solving, process algebras, protocol logics, probabilistic model checking, game theory, and executable models based on logic programming.

• CS 261 Internet/Network Security UC Berkeley 🥟 🛎

This class aims to provide a thorough grounding in network security suitable for those interested in conducting research in the area, as well as students more generally interested in either security or networking. We will also look at broader issues relating to Internet security for which networking plays a role. Topics include: denial-of-service; capabilities; network intrusion detection; worms; forensics; scanning; traffic analysis / inferring activity; architecture; protocol issues; legality and ethics; web attacks; anonymity; honeypots; botnets; spam; the underground economy; research pitfalls. The course is taught with an emphasis on seminal papers rather than bleeding-edge for a given topic.

• CS 5430 System Security Cornell University

- This course discusses security for computers and networked information systems.
 We focus on abstractions, principles, and defenses for implementing military as well as commercial-grade secure systems.
- Syllabus
- o **Lectures**
- Assignments
- CSCI 4968 Modern Binary Exploitation Rensselaer Polytechnic Institute
 - This repository contains the materials as developed and used by <u>RPISEC</u> to teach Modern Binary Exploitation at <u>Rensselaer Polytechnic Institute</u> in Spring 2015. This

was a university course developed and run solely by students to teach skills in vulnerability research, reverse engineering, and binary exploitation.

- **Lectures Notes**
- Labs
- **Projects**
- CSCI 4976 Malware Analysis Rensselaer Polytechnic Institute



o This repository contains the materials as developed and used by RPISEC to teach Malware Analysis at Rensselaer Polytechnic Institute in Fall 2015. This was a university course developed and run soley by students, primarily using the

EECS 588 Computer & Network Security University of Michigan 🥒 管



- Taught by J. Alex Halderman who has analyzed the security of Electronic Voting Machines in the US and over seas.
- This intensive research seminar covers foundational work and current topics in computer systems security.
- Readings Practical Malware Analysis book by Michael Sikorski and Andrew Honig, to teach skills in reverse engineering, malicious behaviour, malware, and antianalysis techniques.
- **Lectures Notes**
- Labs
- **Projects**
- 6.857 Computer and Network Security MIT
 - o Emphasis on applied cryptography and may include: basic notion of systems security, cryptographic hash functions, symmetric cryptography (one-time pad, stream ciphers, block ciphers), cryptanalysis, secret-sharing, authentication codes, public-key cryptography (encryption, digital signatures), public-key attacks, web browser security, biometrics, electronic cash, viruses, electronic voting, Assignments include a group final project. Topics may vary year to year. Lecture **Notes References**
- 6.858 Computer Systems Security MIT **1**
 - Design and implementation of secure computer systems. Lectures cover threat models, attacks that compromise security, and techniques for achieving security, based on recent research papers. Topics include operating system (OS) security, capabilities, information flow control, language security, network protocols, hardware security, and security in web applications.
 - Taught by <u>James Mickens</u> and <u>Nickolai Zeldovich</u>
 - Video Lectures and Labs
 - Quizzes
 - Readings
 - **Final Projects**
- 18-636 Browser Security Stanford 🐨 🦻
 - The Web continues to grow in popularity as platform for retail transactions, financial services, and rapidly evolving forms of communication. It is becoming an increasingly attractive target for attackers who wish to compromise users' systems

or steal data from other sites. Browser vendors must stay ahead of these attacks by providing features that support secure web applications. This course will study vulnerabilities in existing web browsers and the applications they render, as well as new technologies that enable web applications that were never before possible. The material will be largely based on current research problems, and students will be expected to criticize and improve existing defenses. Topics of study include (but are not limited to) browser encryption, JavaScript security, plug-in security, sandboxing, web mashups, and authentication.

Artificial Intelligence

- CS 188 Introduction to Artificial Intelligence UC Berkeley 🖬 📕 🦻
 - This course will introduce the basic ideas and techniques underlying the design of intelligent computer systems. A specific emphasis will be on the statistical and decision-theoretic modeling paradigm. By the end of this course, you will have built autonomous agents that efficiently make decisions in fully informed, partially observable and adversarial settings. Your agents will draw inferences in uncertain environments and optimize actions for arbitrary reward structures. Your machine learning algorithms will classify handwritten digits and photographs. The techniques you learn in this course apply to a wide variety of artificial intelligence problems and will serve as the foundation for further study in any application area you choose to pursue.
 - Lectures
 - **Projects**
 - **Exams**
- CS 4700 Foundations of Artificial Intelligence Cornell University
 - This course will provide an introduction to computer vision, with topics including image formation, feature detection, motion estimation, image mosaics, 3D shape reconstruction, and object and face detection and recognition. Applications of these techniques include building 3D maps, creating virtual characters, organizing photo and video databases, human computer interaction, video surveillance, automatic vehicle navigation, and mobile computer vision. This is a project-based course, in which you will implement several computer vision algorithms throughout the semester.
 - **Assignments**
 - Lectures
- CS 6700 Advanced Artificial Intelligence Cornell University 🥒 둘
 - The design of systems that are among top 10 performers in the world (human, computer, or hybrid human-computer).
 - <u>Syllabus</u>
 - Lectures
 - Readings
- 6.868J The Society of Mind MIT 📝 💻



- This course is an introduction, by Prof. Marvin Minsky, to the theory that tries to explain how minds are made from collections of simpler processes. It treats such aspects of thinking as vision, language, learning, reasoning, memory, consciousness, ideals, emotions, and personality. It incorporates ideas from psychology, artificial intelligence, and computer science to resolve theoretical issues such as wholes vs. parts, structural vs. functional descriptions, declarative vs. procedural representations, symbolic vs. connectionist models, and logical vs. common-sense theories of learning.
- Lectures
- o **Assignments**
- Readings

Computer Graphics

- CAP 5415 Computer Vision University of Central Florida 🐨 🥟 💻
 - o An introductory level course covering the basic topics of computer vision, and introducing some fundamental approaches for computer vision research.
 - Lectures and Videos
 - o <u>Assignments</u>
- CIS 581 Computer Vision and Computational Photography University of Pennsylvania
 - An introductory course in computer vision and computational photography focusing on four topics: image features, image morphing, shape matching, and image search.
 - o **Lectures**
 - Assignments
- CMU 462 Computer Graphics Carnegie Mellon University 📝 📕 管
 - This course provides a comprehensive introduction to computer graphics. Focuses on fundamental concepts and techniques, and their cross-cutting relationship to multiple problem domains in graphics (rendering, animation, geometry, imaging). Topics include: sampling, aliasing, interpolation, rasterization, geometric transformations, parameterization, visibility, compositing, filtering, convolution, curves & surfaces, geometric data structures, subdivision, meshing, spatial hierarchies, ray tracing, radiometry, reflectance, light fields, geometric optics, Monte Carlo rendering, importance sampling, camera models, high-performance ray tracing, differential equations, time integration, numerical differentiation, physically-based animation, optimization, numerical linear algebra, inverse kinematics, Fourier methods, data fitting, example-based synthesis.
 - o Lectures and Readings
 - o Assignments and Quizes
- CS 123 Introduction to Computer Graphics Brown University

- This course offers an in-depth exploration of fundamental concepts in 2D and 3D computer graphics. It introduces 2D raster graphics techniques, including scan conversion, simple image processing, interaction techniques and user interface design. The bulk of the course is devoted to 3D modeling, geometric transformations, and 3D viewing and rendering.
- Lectures
- Labs
- **Demos**
- CS 378 3D Reconstruction with Computer Vision UTexas



- In this lab-based class, we'll dive into practical applications of 3D reconstruction, combining hardware and software to build our own 3D environments from scratch. We'll use open-source frameworks like OpenCV to do the heavy lifting, with the focus on understanding and applying state-of-the art approaches to geometric computer vision
- Lectures
- CS 4620 Introduction to Computer Graphics Cornell University
 - The study of creating, manipulating, and using visual images in the computer.
 - **Assignments**
 - Exams
- CS 4670 Introduction to Computer Vision Cornell University
 - This course will provide an introduction to computer vision, with topics including image formation, feature detection, motion estimation, image mosaics, 3D shape reconstruction, and object and face detection and recognition. Applications of these techniques include building 3D maps, creating virtual characters, organizing photo and video databases, human computer interaction, video surveillance, automatic vehicle navigation, and mobile computer vision. This is a project-based course, in which you will implement several computer vision algorithms throughout the semester.
 - **Assignments**
 - Lectures
- CS 6670 Computer Vision Cornell University
 - Introduction to computer vision. Topics include edge detection, image segmentation, stereopsis, motion and optical flow, image mosaics, 3D shape reconstruction, and object recognition. Students are required to implement several of the algorithms covered in the course and complete a final project.
 - <u>Syllabus</u>
 - Lectures
 - **Assignments**
- CSCI-GA.2270-001 Graduate Computer Graphics New York University



- Step-by-step study computer graphics, with reading and homework at each lecture (Fall2015)
- Lectures

Misc

- AM 207 Monte Carlo Methods and Stochastic Optimization Harvard University
 - This course introduces important principles of Monte Carlo techniques and demonstrates the power of these techniques with simple (but very useful) applications. All of this in Python!
 - Lecture Videos
 - o **Assignments**
 - Lecture Notes
- CS 75 Introduction to Game Development Tufts University
 - o The course taught by Ming Y. Chow teaches game development initially in PyGame through Python, before moving on to addressing all facets of game development. Topics addressed include game physics, sprites, animation, game development methodology, sound, testing, MMORPGs and online games, and addressing mobile development in Android, HTML5, and iOS. Most to all of the development is focused on PyGame for learning principles
 - Text Lectures
 - Assignments
 - o Labs
- CS 100 Open Source Software Construction UC Riverside
 - o This is a course on how to be a hacker. Your first four homework assignments walk you through the process of building your own unix shell. You'll be developing it as an open source project, and you will collaborate with each other at various points.
 - o Github Page
 - Assignments
- CS 108 Object Oriented System Design Stanford
 - Software design and construction in the context of large OOP libraries. Taught in Java. Topics: OOP design, design patterns, testing, graphical user interface (GUI) OOP libraries, software engineering strategies, approaches to programming in teams.
- CS 168 Computer Networks UC Berkeley 🚆 篖 🦻
 - This is an undergraduate level course covering the fundamental concepts of networking as embodied in the Internet. The course will cover a wide range of topics; see the lecture schedule for more details. While the class has a textbook, we will not follow its order of presentation but will instead use the text as a reference when covering each individual topic. The course will also have several projects that involve programming (in Python).
 - You should know programming, data structures, and software engineering. In terms of mathematics, your algebra should be very solid, you need to know basic probability, and you should be comfortable with thinking abstractly. The TAs will spend very little time reviewing material that is not specific to networking. We assume that you either know the material covered in those courses, or are willing to learn the material as necessary. We won't cover any of this material in lecture.





- Course Description: This course provides an introduction to developing applications for the Android mobile platform.
- Prerequisite: CS 106B or equivalent. Java experience highly recommended. OOP highly recommended.
- Devices: Access to an Android phone and/or tablet recommended but not required.
- Videos: Videos list can be found here
- Other materials: Some codes, handsout, homework and lecture notes are not downloadable on the site due to login requirement. Please head to my Github repo here to download them.

CS 193p Developing Applications for iOS Stanford University

- Updated for iOS 7. Tools and APIs required to build applications for the iPhone and iPad platform using the iOS SDK. User interface designs for mobile devices and unique user interactions using multi-touch technologies. Object-oriented design using model-view-controller paradigm, memory management, Objective-C programming language. Other topics include: object-oriented database API, animation, multi-threading and performance considerations.
- Prerequisites: C language and object-oriented programming experience
- Recommended: Programming Abstractions
- Updated courses for iOS8 Swift
- Updated courses for iOS9 Swift

CS 223A Introduction to Robotics Stanford University **1**

- o The purpose of this course is to introduce you to basics of modeling, design, planning, and control of robot systems. In essence, the material treated in this course is a brief survey of relevant results from geometry, kinematics, statics, dynamics, and control.
- Lectures
- <u>Assignments</u>

CS 262a Advanced Topics in Computer Systems UC Berkeley 👺 📝

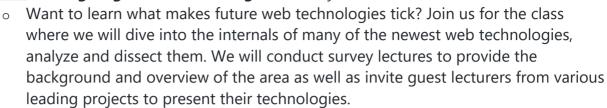


- CS262a is the first semester of a year-long sequence on computer systems research, including operating systems, database systems, and Internet infrastructure systems. The goal of the course is to cover a broad array of research topics in computer systems, and to engage you in top-flight systems research. The first semester is devoted to basic thematic issues and underlying techniques in computer systems, while the second semester goes deeper into topics related to scalable, parallel and distributed systems. The class is based on a discussion of important research papers and a research project.
- Parts: Some Classics, Persistent Storage, Concurrency, Higher-Level Models, Virtual Machines, Cloud Computing, Parallel and Distributed Computing, Potpourri.
- Prerequisites: The historical prerequisite was to pass an entrance exam in class, which covered undergraduate operating systems material (similar to UCB's CS162). There is no longer an exam. However, if you have not already taken a decent

undergrad OS class, you should talk with me before taking this class. The exam had the benefit of "paging in" the undergrad material, which may have been its primary value (since the pass rate was high).

Readings & Lectures

CS 294 Cutting-edge Web Technologies Berkeley 💻 📽 🦻



CS 411 Software Architecture Design Bilkent University

This course teaches the basic concepts, methods and techniques for designing software architectures. The topics include: rationale for software architecture design, modeling software architecture design, architectural styles/patterns, architectural requirements analysis, comparison and evaluation of architecture design methods, synthesis-based software architecture design, software productline architectures, domain modeling, domain engineering and application engineering, software architecture implementation, evaluating software architecture designs.

CS 3152 Introduction to Computer Game Development Cornell University



- A project-based course in which programmers and designers collaborate to make a computer game. This course investigates the theory and practice of developing computer games from a blend of technical, aesthetic, and cultural perspectives. Technical aspects of game architecture include software engineering, artificial intelligence, game physics, computer graphics, and networking. Aesthetic and cultural include art and modeling, sound and music, game balance, and player experience.
- **Syllabus**
- <u>Lectures</u>
- **Assignments**

CS 4152 Advanced Topics in Computer Game Development Cornell University



Project-based follow-up course to CS/INFO 3152. Students work in a multidisciplinary team to develop a game that incorporates innovative game technology. Advanced topics include 3D game development, mobile platforms, multiplayer gaming, and nontraditional input devices. There is a special emphasis on developing games that can be submitted to festivals and competitions, or that can be commercialized.

- <u>Syllabus</u>
- <u>Lectures</u>
- **Assignments**
- CS 4154 Analytics-driven Game Design Cornell University 📕 📝 管



A project-based course in which programmers and designers collaborate to design, implement, and release a video game online through popular game

portals. In this course, students will use the internet to gather data anonymously from players. Students will analyze this data in order to improve their game over multiple iterations. Technical aspects of this course include programming, database architecture, and statistical analysis.

- **Syllabus**
- Lectures
- **Assignments**
- CS 4812 Quantum Information Processing Cornell University | 📝 둘



Hardware that exploits quantum phenomena can dramatically alter the nature of computation. Though constructing a working quantum computer is a formidable technological challenge, there has been much recent experimental progress. In addition, the theory of quantum computation is of interest in itself, offering strikingly different perspectives on the nature of computation and information, as well as providing novel insights into the conceptual puzzles posed by the quantum theory. The course is intended both for physicists, unfamiliar with computational complexity theory or cryptography, and also for computer scientists and mathematicians, unfamiliar with quantum mechanics. The prerequisites are familiarity (and comfort) with finite dimensional vector spaces over the complex numbers, some standard group theory, and ability to count in binary.

- <u>Syllabus</u>
- Lectures
- CS 4860 Applied Logic Cornell University



In addition to basic first-order logic, when taught by Computer Science this course involves elements of Formal Methods and Automated Reasoning. Formal Methods is concerned with proving properties of algorithms, specifying programming tasks and synthesizing programs from proofs. We will use formal methods tools such as interactive proof assistants (see www.nuprl.org). We will also spend two weeks on constructive type theory, the language used by the Cog and Nuprl proof assistants.

- <u>Syllabus</u>
- Lectures
- **Assignments**
- CS 5150 Software Engineering Cornell University = 📝 答
 - Introduction to the practical problems of specifying, designing, building, testing, and delivering reliable software systems
 - Lectures
 - **Assignments**
- CS 5220 Applications of Parallel Computers Cornell University



How do we solve the large-scale problems of science quickly on modern computers? How do we measure the performance of new or existing simulation codes, and what things can we do to make them run faster? How can we best take advantage of features like multicore processors, vector units, and graphics co-

processors? These are the types of questions we will address in CS 5220, Applications of Parallel Computers. Topics include:

- Single-processor architecture, caches, and serial performance tuning
- Basics of parallel machine organization
- Distributed memory programming with MPI
- Shared memory programming with OpenMP
- Parallel patterns: data partitioning, synchronization, and load balancing
- Examples of parallel numerical algorithms
- Applications from science and engineering
- Lectures
- **Assignments**
- **CS 5540** Computational Techniques for Analyzing Clinical Data Cornell University 📕 🦻 😉
 - o CS5540 is a masters-level course that covers a wide range of clinical problems and their associated computational challenges. The practice of medicine is filled with digitally accessible information about patients, ranging from EKG readings to MRI images to electronic health records. This poses a huge opportunity for computer tools that make sense out of this data. Computation tools can be used to answer seemingly straightforward questions about a single patient's test results ("Does this patient have a normal heart rhythm?"), or to address vital questions about large populations ("Is there any clinical condition that affects the risks of Alzheimer"). In CS5540 we will look at many of the most important sources of clinical data and discuss the basic computational techniques used for their analysis, ranging in sophistication from current clinical practice to state-of-the-art research projects.
 - **Syllabus**
 - **Lectures**
 - **Assignments**
- CS 5724 Evolutionary Computation Cornell University

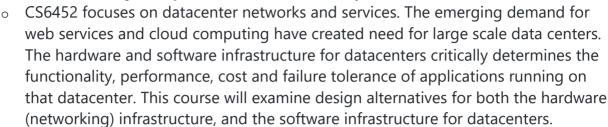


This course will cover advanced topics in evolutionary algorithms and their application to open-ended computational design. The field of evolutionary computation tries to address large-scale optimization and planning problems through stochastic population-based methods. It draws inspiration from evolutionary processes in nature and in engineering, and also serves as abstract models for these phenomena. Evolutionary processes are generally weak methods that require little information about the problem domain and hence can be applied across a wide variety of applications. They are especially useful for openended problem domains for which little formal knowledge exists and the number of parameters is undefined, such as for the general engineering design process. This course will provide insight to a variety of evolutionary computation paradigms, such as genetic algorithms, genetic programming, and evolutionary strategies, as well as governing dynamics of co-evolution, arms races and mediocre stable states. New methods involving symbiosis models and pattern recognition will also be presented. The material will be intertwined with

discussions of representations and results for design problems in a variety of problem domains including software, electronics, and mechanics.

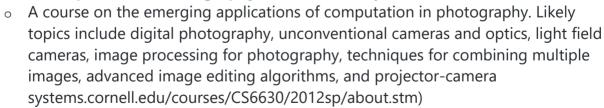
- Syllabus
- Lectures
- o **Assignments**

CS 6452 Evolutionary Computation Cornell University

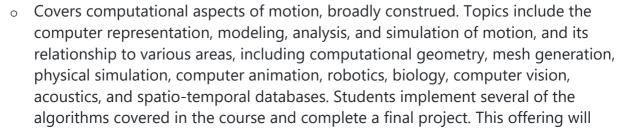


- Syllabus
- <u>Lectures</u>
- 🔸 CS 6630 Realistic Image Synthesis Cornell University 💻 🥟 管
 - CS6630 is an introduction to physics-based rendering at the graduate level. Starting from the fundamentals of light transport we will look at formulations of the Rendering Equation, and a series of Monte Carlo methods, from sequential sampling to multiple importance sampling to Markov Chains, for solving the equation to make pictures. We'll look at light reflection from surfaces and scattering in volumes, illumination from luminaries and environments, and diffusion models for translucent materials. We will build working implementations of many of the algorithms we study, and learn how to make sure they are actually working correctly. It's fun to watch integrals and probability distributions transform into photographs of a slightly too perfect synthetic world.
 - Syllabus
 - <u>Lectures</u>
 - o **Assignments**
 - Readings

• CS 6640 Computational Photography Cornell University



- o Lectures
- Assignments
- CS 6650 Computational Motion Cornell University



also explore the special role of motion processing in physically based sound rendering.

- CS 6840 Algorithmic Game Theory Cornell University \(\big| \) \(\big| \)
 - Algorithmic Game Theory combines algorithmic thinking with game-theoretic, or, more generally, economic concepts. The course will study a range of topics at this interface
 - o Syllabus
 - Lectures
 - Assignments
 - Readings
- CSE 154 Web Programming University of Washington
 - This course is an introduction to programming for the World Wide Web. Covers use of HTML, CSS, PHP, JavaScript, AJAX, and SQL.
 - Lectures
 - Assignments
- ESM 296-4F GIS & Spatial Analysis UC Santa Barbara 📕 🦻
 - Taught by James Frew, Ben Best, and Lisa Wedding
 - Focuses on specific computational languages (e.g., Python, R, shell) and tools (e.g., GDAL/OGR, InVEST, MGET, ModelBuilder) applied to the spatial analysis of environmental problems
 - GitHub (includes lecture materials and labs)
- 🔹 ICS 314 Software Engineering University of Hawaii 💻 📝 管
 - Taught by <u>Philip Johnson</u>
 - Introduction to software engineering using the <u>"Athletic Software Engineering"</u> <u>pedagogy</u>
 - o <u>Readings</u>
 - Experiences
 - Assessments
- IGME 582 Humanitarian Free & Open Source Software Development Rochester Institute of Technology
 - This course provides students with exposure to the design, creation and production of Open Source Software projects. Students will be introduced to the historic intersections of technology and intellectual property rights and will become familiar with Open Source development processes, tools and practices.
- <u>I485 / H400</u> Biologically Inspired Computation Indiana University
 - Course taught by <u>Luis Rocha</u> about the multi-disciplinary field algorithms inspired by naturally occurring phenomenon. This course provides introduces the following areas: L-systems, Cellular Automata, Emergence, Genetic Algorithms, Swarm Intelligence and Artificial Immune Systems. It's aim is to cover the fundamentals and enable readers to build up a proficiency in applying various algorithms to real-world problems.
 - <u>Lectures</u>
 - Assignments

Open Sourced Elective: Database and Rails Intro to Ruby on Rails University of Texas



- An introductory course in Ruby on Rails open sourced by University of Texas' CS Adjunct Professor, Richard Schneeman.
- Lectures
- **Assignments**
- Videos





- This is a graduate course in scientific computing created and taught by Oliver Serang in 2014, which covers topics in computer science and statistics with applications from biology. The course is designed top-down, starting with a problem and then deriving a variety of solutions from scratch.
- Topics include memoization, recurrence closed forms, string matching (sorting, hash tables, radix tries, and suffix tries), dynamic programming (e.g. Smith-Waterman and Needleman-Wunsch), Bayesian statistics (e.g. the envelope paradox), graphical models (HMMs, Viterbi, junction tree, belief propagation), FFT, and the probabilistic convolution tree.
- Lecture videos on Youtube and for direct download
- 14-740 Fundamentals of Computer Networks CMU 🖬 💻 隓 🦻
 - This is an introductory course on Networking for graduate students. It follows a top-down approach to teaching Computer Networks, so it starts with the Application layer which most of the students are familiar with and as the course unravels we learn more about transport, network and link layers of the protocol stack.
 - As far as prerequisites are concerned basic computer, programming and probability theory background is required.
 - The course site contains links to the lecture videos, reading material and assignments.

Awesome Artificial Intelligence (AI)

https://github.com/owainlewis/awesome-artificial-intelligence

A curated list of Artificial Intelligence (AI) courses, books, video lectures and papers.

Contributions most welcome.

Contents

- 1. Courses
- 2. Books
- 3. Programming
- 4. Philosophy
- 5. Free Content
- 6. Code
- 7. Videos

- 8. Learning
- 9. Organizations
- 10. Journals
- 11. Competitions
- 12. Movies
- 13. Misc

Courses

- MIT Artifical Intelligence Videos MIT AI Course
- Intro to Artificial Intelligence Learn the Fundamentals of Al. Course run by Peter Norvig
- <u>EdX Artificial Intelligence</u> The course will introduce the basic ideas and techniques underlying the design of intelligent computer systems
- <u>Artificial Intelligence For Robotics</u> This class will teach you basic methods in Artificial Intelligence, including: probabilistic inference, planning and search, localization, tracking and control, all with a focus on robotics
- <u>Machine Learning</u> Basic machine learning algorithms for supervised and unsupervised learning
- <u>Neural Networks For Machine Learning</u> Algorithmic and practical tricks for artifical neural networks.
- Deep Learning An Introductory course to the world of Deep Learning.
- <u>Stanford Statistical Learning</u> Introductory course on machine learning focusing on: linear and polynomial regression, logistic regression and linear discriminant analysis; cross-validation and the bootstrap, model selection and regularization methods (ridge and lasso); nonlinear models, splines and generalized additive models; tree-based methods, random forests and boosting; support-vector machines.
- <u>Knowledge Based Artificial Intelligence</u> Georgia Tech's course on Artificial Intelligence focussing on Symbolic Al.

Books

- Artificial Intelligence: A Modern Approach Stuart Russell & Peter Norvig
 - Also consider browsing the <u>list of recommended reading</u>, divided by each chapter in "Artificial Intelligence: A Modern Approach".
- <u>Paradigms Of Artificial Intelligence Programming: Case Studies in Common Lisp</u> -Paradigms of AI Programming is the first text to teach advanced Common Lisp techniques in the context of building major AI systems
- Reinforcement Learning: An Introduction This introductory textbook on reinforcement learning is targeted toward engineers and scientists in artificial intelligence, operations research, neural networks, and control systems, and we hope it will also be of interest to psychologists and neuroscientists.
- <u>The Cambridge Handbook Of Artificial Intelligence</u> Written for non-specialists, it covers
 the discipline's foundations, major theories, and principal research areas, plus related
 topics such as artificial life
- The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind - In this mind-expanding book, scientific pioneer Marvin Minsky continues his groundbreaking research, offering a fascinating new model for how our minds work
- Artificial Intelligence: A New Synthesis Beginning with elementary reactive agents,
 Nilsson gradually increases their cognitive horsepower to illustrate the most important and lasting ideas in Al
- On Intelligence Hawkins develops a powerful theory of how the human brain works, explaining why computers are not intelligent and how, based on this new theory, we can finally build intelligent machines. Also audio version available from audible.com
- How To Create A Mind Kurzweil discusses how the brain works, how the mind emerges, brain-computer interfaces, and the implications of vastly increasing the powers of our intelligence to address the world's problems

Programming

- <u>Prolog Programming For Artificial Intelligence</u> This best-selling guide to Prolog and Artificial Intelligence concentrates on the art of using the basic mechanisms of Prolog to solve interesting AI problems.
- Al Algorithms, Data Structures and Idioms in Prolog, Lisp and Java PDF here
- Python Tools for Machine Learning
- Python for Artificial Intelligence

Philosophy

• <u>Super Intelligence</u> - Superintelligence asks the questions: What happens when machines surpass humans in general intelligence. A really great book.

- Our Final Invention: Artificial Intelligence And The End Of The Human Era Our Final Invention explores the perils of the heedless pursuit of advanced Al. Until now, human intelligence has had no rival. Can we coexist with beings whose intelligence dwarfs our own? And will they allow us to?
- How to Create a Mind: The Secret of Human Thought Revealed Ray Kurzweil, director of engineering at Google, explored the process of reverse-engineering the brain to understand precisely how it works, then applies that knowledge to create vastly intelligent machines.
- <u>Minds, Brains, And Programs</u> The 1980 paper by philospher John Searle that contains the famous 'Chinese Room' thought experiment. Probably the most famous attack on the notion of a Strong Al possessing a 'mind' or a 'consciousness', and interesting reading for those interested in the intersection of Al and philosophy of mind.
- Gödel, Escher, Bach: An Eternal Golden Braid Written by Douglas Hofstadter and taglined "a metaphorical fugue on minds and machines in the spirit of Lewis Carroll", this wonderful journey into the the fundamental concepts of mathematics, symmetry and intelligence won a Pulitzer Price for Non-Fiction in 1979. A major theme throughout is the emergence of meaning from seemingly 'meaningless' elements, like 1's and 0's, arranged in special patterns.

Free Content

- <u>Foundations Of Computational Agents</u> This book is published by Cambridge University Press, 2010
- <u>The Quest For Artificial Intelligence</u> This book traces the history of the subject, from the early dreams of eighteenth-century (and earlier) pioneers to the more successful work of today's Al engineers.
- <u>Stanford CS229 Machine Learning</u> This course provides a broad introduction to machine learning and statistical pattern recognition.
- Computers and Thought: A practical Introduction to Artificial Intelligence The book covers computer simulation of human activities, such as problem solving and natural language understanding; computer vision; Al tools and techniques; an introduction to Al programming; symbolic and neural network models of cognition; the nature of mind and intelligence; and the social implications of Al and cognitive science.
- <u>Society of Mind</u> Marvin Minsky's seminal work on how our mind works. Lot of Symbolic Al concepts have been derived from this basis.
- <u>Artificial Intelligence and Molecular Biology</u> The current volume is an effort to bridge that range of exploration, from nucleotide to abstract concept, in contemporary AI/MB research.
- <u>Brief Introduction To Educational Implications Of Artificial Intelligence</u> This book is designed to help preservice and inservice teachers learn about some of the educational implications of current uses of Artificial Intelligence as an aid to solving problems and accomplishing tasks.
- <u>Encyclopedia: Computational intelligence</u> Scholarpedia is a peer-reviewed open-access encyclopedia written and maintained by scholarly experts from around the world.

• <u>Ethical Artificial Intelligence</u> - a book by Bill Hibbard that combines several peer reviewed papers and new material to analyze the issues of ethical artificial intelligence.

Code

- <u>AIMACode</u> Source code for "Artificial Intelligence: A Modern Approach" in Common Lisp, Java, Python. More to come.
- <u>FANN</u> Fast Artificial Neural Network Library, native for C

Videos

- A tutorial on Deep Learning
- Basics of Computational Reinforcement Learning
- Deep Reinforcement Learning
- Intelligent agents and paradigms for AI
- The Unreasonable Effectiveness Of Deep Learning The Director of Facebook's Al Research, Dr. Yann LeCun gives a talk on deep convolutional neural networks and their applications to machine learning and computer vision

Learning

- <u>Deep Learning. Methods And Applications</u> Free book from Microsoft Research
- <u>Neural Networks And Deep Learning</u> Neural networks and deep learning currently
 provide the best solutions to many problems in image recognition, speech recognition,
 and natural language processing. This book will teach you the core concepts behind
 neural networks and deep learning
- <u>Machine Learning: A Probabilistic Perspective</u> This textbook offers a comprehensive and self-contained introduction to the field of machine learning, based on a unified, probabilistic approach
- <u>Deep Learning</u> Yoshua Bengio, Ian Goodfellow and Aaron Courville put together this currently free (and draft version) book on deep learning. The book is kept up-to-date and covers a wide range of topics in depth (up to and including sequence-to-sequence learning).
- Getting Started with Deep Learning and Python
- Machine Learning Mastery
- Deep Learning.net Aggregation site for DL resources
- Awesome Machine Learning Like this Github, but ML-focused
- FastML
- Awesome Deep Learning Resources Rough list of learning resources for Deep Learning

Organizations

- <u>IEEE Computational Intelligence Society</u>
- Machine Intelligence Research Institute
- OpenAl
- Association For The Advancement of Artificial Intelligence

Journals

- Al & Society
- Annals of Mathematics and Artifical Intelligence
- Applicable Algebra in Engineering, Communication and Computing
- Applied Intelligence
- Artificial Intelligence Review
- Automated Software Engineering
- Autonomous Agents and Multi-Agent Systems
- Computational and Mathematical Organization Theory
- <u>Evolutionary Intelligence</u>
- Intelligent Industrial Systems
- Journal of Automated Reasoning
- Journal on Data Semantics
- Journal of Intelligent Information Systems
- Minds and Machines
- Progress in Artificial Intelligence
- Artificial Intelligence
- Journal of Artificial Intelligence Research
- Al Magazine
- EXPERT—IEEE Intelligent Systems
- Computational Intelligence
- International Journal of Intelligent Systems
- Applied Artificial Intelligence
- Knowledge Engineering Review
- Journal of Experimental and Theoretical Artificial Intelligence
- Artificial Intelligence for Engineering Design, Analysis and Manufacturing
- Al Communications
- International Journal on Artificial Intelligence Tools
- <u>Electronic Transactions on Artificial Intelligence</u>
- IEEE Transactions Automation Science and Engineering

Competitions

- Al Challenge
- Al Games
- Building JS robots

MIT Battlecode

Movies

- 2001: A Space Odyssey
- A.I. Artificial Intelligence
- Automata
- Blade Runner
- Chappie

- Ex Machina
- Her
- <u>I, Robot</u>
- <u>Prometheus</u>
- The Terminator
- Transcendence

Misc

- Open Cognition Project We're undertaking a serious effort to build a thinking machine
- <u>AlTopics</u> Large aggregation of Al resources
- <u>AIResources</u> Directory of open source software and open access data for the AI research community

Awesome Deep Learning

https://github.com/ChristosChristofidis/awesome-deep-learning

Table of Contents

- Free Online Books
- Courses
- Videos and Lectures
- Papers
- Tutorials
- Researchers

- WebSites
- Datasets
- Frameworks
- Miscellaneous
- Contributing

Free Online Books

- 1. <u>Deep Learning</u> by Yoshua Bengio, Ian Goodfellow and Aaron Courville (05/07/2015)
- 2. Neural Networks and Deep Learning by Michael Nielsen (Dec 2014)
- 3. Deep Learning by Microsoft Research (2013)
- 4. <u>Deep Learning Tutorial</u> by LISA lab, University of Montreal (Jan 6 2015)
- 5. neuraltalk by Andrej Karpathy: numpy-based RNN/LSTM implementation
- 6. An introduction to genetic algorithms
- 7. Artificial Intelligence: A Modern Approach
- 8. <u>Deep Learning in Neural Networks: An Overview</u>

Courses

- 1. Machine Learning Stanford by Andrew Ng in Coursera (2010-2014)
- 2. Machine Learning Caltech by Yaser Abu-Mostafa (2012-2014)
- 3. Machine Learning Carnegie Mellon by Tom Mitchell (Spring 2011)
- 4. Neural Networks for Machine Learning by Geoffrey Hinton in Coursera (2012)
- 5. Neural networks class by Hugo Larochelle from Université de Sherbrooke (2013)
- 6. <u>Deep Learning Course</u> by CILVR lab @ NYU (2014)
- 7. A.I Berkeley by Dan Klein and Pieter Abbeel (2013)
- 8. A.I MIT by Patrick Henry Winston (2010)
- 9. <u>Vision and learning computers and brains</u> by Shimon Ullman, Tomaso Poggio, Ethan Meyers @ MIT (2013)

- 10. <u>Convolutional Neural Networks for Visual Recognition Stanford</u> by Fei-Fei Li, Andrej Karpathy (2015)
- 11. <u>Convolutional Neural Networks for Visual Recognition Stanford</u> by Fei-Fei Li, Andrej Karpathy (2016)
- 12. Deep Learning for Natural Language Processing Stanford
- 13. Neural Networks usherbrooke
- 14. Machine Learning Oxford (2014-2015)
- 15. Deep Learning Nvidia (2015)
- Graduate Summer School: Deep Learning, Feature Learning by Geoffrey Hinton, Yoshua Bengio, Yann LeCun, Andrew Ng, Nando de Freitas and several others @ IPAM, UCLA (2012)
- 17. <u>Deep Learning Udacity/Google</u> by Vincent Vanhoucke and Arpan Chakraborty (2016)
- 18. <u>Deep Learning UWaterloo</u> by Prof. Ali Ghodsi at University of Waterloo (2015)
- 19. Statistical Machine Learning CMU by Prof. Larry Wasserman
- 20. <u>Deep Learning Course</u> by Yann LeCun (2016)
- 21. <u>Bay area DL school</u> by Andrew Ng, Yoshua Bengio, Samy Bengio, Andrej Karpathy, Richard Socher, Hugo Larochelle and many others @ Stanford, CA (2016)
- 22. Designing, Visualizing and Understanding Deep Neural Networks-UC Berkeley
- 23. <u>UVA Deep Learning Course</u> MSc in Artificial Intelligence for the University of Amsterdam.
- 24. MIT 6.S094: Deep Learning for Self-Driving Cars
- 25. MIT 6.S191: Introduction to Deep Learning
- 26. Berkeley CS 294: Deep Reinforcement Learning
- 27. Keras in Motion video course
- 28. Practical Deep Learning For Coders by Jeremy Howard Fast.ai

Videos and Lectures

- 1. How To Create A Mind By Ray Kurzweil
- 2. Deep Learning, Self-Taught Learning and Unsupervised Feature Learning By Andrew Ng
- 3. Recent Developments in Deep Learning By Geoff Hinton
- 4. The Unreasonable Effectiveness of Deep Learning by Yann LeCun
- 5. <u>Deep Learning of Representations</u> by Yoshua bengio
- 6. Principles of Hierarchical Temporal Memory by Jeff Hawkins
- 7. Machine Learning Discussion Group Deep Learning w/ Stanford Al Lab by Adam Coates
- 8. Making Sense of the World with Deep Learning By Adam Coates
- 9. <u>Demystifying Unsupervised Feature Learning</u> By Adam Coates
- 10. Visual Perception with Deep Learning By Yann LeCun
- 11. The Next Generation of Neural Networks By Geoffrey Hinton at GoogleTechTalks
- 12. <u>The wonderful and terrifying implications of computers that can learn</u> By Jeremy Howard at TEDxBrussels
- 13. <u>Unsupervised Deep Learning Stanford</u> by Andrew Ng in Stanford (2011)

- 14. Natural Language Processing By Chris Manning in Stanford
- 15. A beginners Guide to Deep Neural Networks By Natalie Hammel and Lorraine Yurshansky
- 16. <u>Deep Learning: Intelligence from Big Data</u> by Steve Jurvetson (and panel) at VLAB in Stanford.
- 17. <u>Introduction to Artificial Neural Networks and Deep Learning</u> by Leo Isikdogan at Motorola Mobility HQ
- 18. NIPS 2016 lecture and workshop videos NIPS 2016

Papers

You can also find the most cited deep learning papers from here

- 1. ImageNet Classification with Deep Convolutional Neural Networks
- 2. Using Very Deep Autoencoders for Content Based Image Retrieval
- 3. Learning Deep Architectures for Al
- 4. CMU's list of papers
- 5. Neural Networks for Named Entity Recognition zip
- 6. Training tricks by YB
- 7. Geoff Hinton's reading list (all papers)
- 8. Supervised Sequence Labelling with Recurrent Neural Networks
- 9. Statistical Language Models based on Neural Networks
- 10. Training Recurrent Neural Networks
- 11. Recursive Deep Learning for Natural Language Processing and Computer Vision
- 12. Bi-directional RNN
- 13. **LSTM**
- 14. GRU Gated Recurrent Unit
- 15. **GFRNN** . .
- 16. LSTM: A Search Space Odyssey
- 17. A Critical Review of Recurrent Neural Networks for Sequence Learning
- 18. Visualizing and Understanding Recurrent Networks
- 19. <u>Wojciech Zaremba, Ilya Sutskever, An Empirical Exploration of Recurrent Network</u>
 Architectures
- 20. Recurrent Neural Network based Language Model
- 21. Extensions of Recurrent Neural Network Language Model
- 22. Recurrent Neural Network based Language Modeling in Meeting Recognition
- 23. <u>Deep Neural Networks for Acoustic Modeling in Speech Recognition</u>
- 24. Speech Recognition with Deep Recurrent Neural Networks
- 25. Reinforcement Learning Neural Turing Machines
- 26. <u>Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation</u>
- 27. <u>Google Sequence to Sequence Learning with Neural Networks</u>

- 28. Memory Networks
- 29. Policy Learning with Continuous Memory States for Partially Observed Robotic Control
- 30. Microsoft Jointly Modeling Embedding and Translation to Bridge Video and Language
- 31. Neural Turing Machines
- 32. Ask Me Anything: Dynamic Memory Networks for Natural Language Processing
- 33. Mastering the Game of Go with Deep Neural Networks and Tree Search
- 34. Batch Normalization
- 35. Residual Learning
- 36. Image-to-Image Translation with Conditional Adversarial Networks
- 37. Berkeley Al Research (BAIR) Laboratory
- 38. MobileNets by Google
- 39. Cross Audio-Visual Recognition in the Wild Using Deep Learning

Tutorials

- 1. UFLDL Tutorial 1
- 2. UFLDL Tutorial 2
- 3. Deep Learning for NLP (without Magic)
- 4. A Deep Learning Tutorial: From Perceptrons to Deep Networks
- 5. Deep Learning from the Bottom up
- 6. Theano Tutorial
- 7. Neural Networks for Matlab
- 8. <u>Using convolutional neural nets to detect facial keypoints tutorial</u>
- 9. Torch7 Tutorials
- 10. The Best Machine Learning Tutorials On The Web
- 11. VGG Convolutional Neural Networks Practical
- 12. TensorFlow tutorials
- 13. More TensorFlow tutorials
- 14. TensorFlow Python Notebooks
- 15. Keras and Lasagne Deep Learning Tutorials
- 16. Classification on raw time series in TensorFlow with a LSTM RNN
- 17. Using convolutional neural nets to detect facial keypoints tutorial
- 18. TensorFlow-World

Researchers

- 1. Aaron Courville
- 2. Abdel-rahman Mohamed
- 3. Adam Coates
- 4. Alex Acero

- 5. Alex Krizhevsky
- 6. Alexander Ilin
- 7. Amos Storkey
- 8. Andrej Karpathy
- 9. Andrew M. Saxe
- 10. Andrew Ng

11. <u>Andrew W. Senior</u>	51. <u>Koray Kavukcuoglu</u>
12. <u>Andriy Mnih</u>	52. <u>KyungHyun Cho</u>
13. <u>Ayse Naz Erkan</u>	53. <u>Li Deng</u>
14. <u>Benjamin Schrauwen</u>	54. <u>Lucas Theis</u>
15. <u>Bernardete Ribeiro</u>	55. <u>Ludovic Arnold</u>
16. <u>Bo David Chen</u>	56. Marc'Aurelio Ranzato
17. <u>Boureau Y-Lan</u>	57. <u>Martin Längkvist</u>
18. <u>Brian Kingsbury</u>	58. Misha Denil
19. <u>Christopher Manning</u>	59. Mohammad Norouzi
20. <u>Clement Farabet</u>	60. <u>Nando de Freitas</u>
21. <u>Dan Claudiu Cireşan</u>	61. Navdeep Jaitly
22. <u>David Reichert</u>	62. <u>Nicolas Le Roux</u>
23. <u>Derek Rose</u>	63. Nitish Srivastava
24. <u>Dong Yu</u>	64. <u>Noel Lopes</u>
25. <u>Drausin Wulsin</u>	65. Oriol Vinyals
26. Erik M. Schmidt	66. Pascal Vincent
27. <u>Eugenio Culurciello</u>	67. <u>Patrick Nguyen</u>
28. <u>Frank Seide</u>	68. <u>Pedro Domingos</u>
29. <u>Galen Andrew</u>	69. <u>Peggy Series</u>
30. <u>Geoffrey Hinton</u>	70. <u>Pierre Sermanet</u>
31. <u>George Dahl</u>	71. <u>Piotr Mirowski</u>
32. <u>Graham Taylor</u>	72. <u>Quoc V. Le</u>
33. <u>Grégoire Montavon</u>	73. Reinhold Scherer
34. Guido Francisco Montúfar	74. Richard Socher
35. Guillaume Desjardins	75. Rob Fergus
36. <u>Hannes Schulz</u>	76. Robert Coop
37. <u>Hélène Paugam-Moisy</u>	77. Robert Gens
38. <u>Honglak Lee</u>	78. Roger Grosse
39. <u>Hugo Larochelle</u>	79. Ronan Collobert
40. <u>Ilya Sutskever</u>	80. Ruslan Salakhutdinov
41. <u>Itamar Arel</u>	81. Sebastian Gerwinn
42. <u>James Martens</u>	82. Stéphane Mallat
43. <u>Jason Morton</u>	83. <u>Sven Behnke</u>
44. <u>Jason Weston</u>	84. <u>Tapani Raiko</u>
45. <u>Jeff Dean</u>	85. <u>Tara Sainath</u>
46. <u>Jiquan Mgiam</u>	86. <u>Tijmen Tieleman</u>
47. <u>Joseph Turian</u>	87. <u>Tom Karnowski</u>
48. <u>Joshua Matthew Susskind</u>	88. <u>Tomáš Mikolov</u>
49. <u>Jürgen Schmidhuber</u>	89. <u>Ueli Meier</u>
50. <u>Justin A. Blanco</u>	90. <u>Vincent Vanhoucke</u>

- 91. Volodymyr Mnih
- 92. Yann LeCun
- 93. Yichuan Tang
- 94. Yoshua Bengio
- 95. Yotaro Kubo
- 96. Youzhi (Will) Zou
- 97. Fei-Fei Li
- 98. lan Goodfellow
- 99. Robert Laganière

WebSites

- 1. <u>deeplearning.net</u>
- 2. <u>deeplearning.stanford.edu</u>
- 3. <u>nlp.stanford.edu</u>
- 4. ai-junkie.com
- 5. cs.brown.edu/research/ai
- 6. eecs.umich.edu/ai
- 7. cs.utexas.edu/users/ai-lab
- 8. cs.washington.edu/research/ai

- 9. aiai.ed.ac.uk
- 10. www-aig.jpl.nasa.gov
- 11. csail.mit.edu
- 12. cgi.cse.unsw.edu.au/~aishare
- 13. cs.rochester.edu/research/ai
- 14. ai.sri.com
- 15. isi.edu/Al/isd.htm
- 16. nrl.navy.mil/itd/aic
- 17. hips.seas.harvard.edu
- 18. AI Weekly
- 19. stat.ucla.edu
- 20. <u>deeplearning.cs.toronto.edu</u>
- 21. jeffdonahue.com/lrcn/
- 22. visualqa.org
- 23. <u>www.mpi-inf.mpg.de/departments/computer-vision...</u>
- 24. Deep Learning News
- 25. <u>Machine Learning is Fun! Adam</u> <u>Geitgey's Blog</u>

Datasets

- 1. MNIST Handwritten digits
- 2. Google House Numbers from street view
- 3. CIFAR-10 and CIFAR-100
- 4. IMAGENET
- 5. <u>Tiny Images</u> 80 Million tiny images6.
- 6. Flickr Data 100 Million Yahoo dataset
- 7. Berkeley Segmentation Dataset 500
- 8. UC Irvine Machine Learning Repository
- 9. Flickr 8k
- 10. Flickr 30k
- 11. Microsoft COCO
- 12. <u>VQA</u>
- 13. Image QA
- 14. AT&T Laboratories Cambridge face database
- 15. AVHRR Pathfinder
- 16. <u>Air Freight</u> The Air Freight data set is a ray-traced image sequence along with ground truth segmentation based on textural characteristics. (455 images + GT, each 160x120 pixels). (Formats: PNG)

- 17. <u>Amsterdam Library of Object Images</u> ALOI is a color image collection of one-thousand small objects, recorded for scientific purposes. In order to capture the sensory variation in object recordings, we systematically varied viewing angle, illumination angle, and illumination color for each object, and additionally captured wide-baseline stereo images. We recorded over a hundred images of each object, yielding a total of 110,250 images for the collection. (Formats: png)
- 18. <u>Annotated face, hand, cardiac & meat images</u> Most images & annotations are supplemented by various ASM/AAM analyses using the AAM-API. (Formats: bmp,asf)
- 19. Image Analysis and Computer Graphics
- 20. <u>Brown University Stimuli</u> A variety of datasets including geons, objects, and "greebles". Good for testing recognition algorithms. (Formats: pict)
- 21. <u>CAVIAR video sequences of mall and public space behavior</u> 90K video frames in 90 sequences of various human activities, with XML ground truth of detection and behavior classification (Formats: MPEG2 & JPEG)
- 22. Machine Vision Unit
- 23. <u>CCITT Fax standard images</u> 8 images (Formats: gif)
- 24. <u>CMU CIL's Stereo Data with Ground Truth</u> 3 sets of 11 images, including color tiff images with spectroradiometry (Formats: gif, tiff)
- 25. <u>CMU PIE Database</u> A database of 41,368 face images of 68 people captured under 13 poses, 43 illuminations conditions, and with 4 different expressions.
- 26. <u>CMU VASC Image Database</u> Images, sequences, stereo pairs (thousands of images) (Formats: Sun Rasterimage)
- 27. <u>Caltech Image Database</u> about 20 images mostly top-down views of small objects and toys. (Formats: GIF)
- 28. <u>Columbia-Utrecht Reflectance and Texture Database</u> Texture and reflectance measurements for over 60 samples of 3D texture, observed with over 200 different combinations of viewing and illumination directions. (Formats: bmp)
- 29. <u>Computational Colour Constancy Data</u> A dataset oriented towards computational color constancy, but useful for computer vision in general. It includes synthetic data, camera sensor data, and over 700 images. (Formats: tiff)
- 30. Computational Vision Lab
- 31. <u>Content-based image retrieval database</u> 11 sets of color images for testing algorithms for content-based retrieval. Most sets have a description file with names of objects in each image. (Formats: jpg)
- 32. Efficient Content-based Retrieval Group
- 33. <u>Densely Sampled View Spheres</u> Densely sampled view spheres upper half of the view sphere of two toy objects with 2500 images each. (Formats: tiff)
- 34. Computer Science VII (Graphical Systems)
- 35. <u>Digital Embryos</u> Digital embryos are novel objects which may be used to develop and test object recognition systems. They have an organic appearance. (Formats: various formats are available on request)
- 36. <u>Univerity of Minnesota Vision Lab</u>
- 37. <u>El Salvador Atlas of Gastrointestinal VideoEndoscopy</u> Images and Videos of his-res of studies taken from Gastrointestinal Video endoscopy. (Formats: jpg, mpg, gif)

- 38. <u>FG-NET Facial Aging Database</u> Database contains 1002 face images showing subjects at different ages. (Formats: jpg)
- 39. <u>FVC2000 Fingerprint Databases</u> FVC2000 is the First International Competition for Fingerprint Verification Algorithms. Four fingerprint databases constitute the FVC2000 benchmark (3520 fingerprints in all).
- 40. Biometric Systems Lab University of Bologna
- 41. <u>Face and Gesture images and image sequences</u> Several image datasets of faces and gestures that are ground truth annotated for benchmarking
- 42. <u>German Fingerspelling Database</u> The database contains 35 gestures and consists of 1400 image sequences that contain gestures of 20 different persons recorded under non-uniform daylight lighting conditions. (Formats: mpg,jpg)
- 43. Language Processing and Pattern Recognition
- 44. <u>Groningen Natural Image Database</u> 4000+ 1536x1024 (16 bit) calibrated outdoor images (Formats: homebrew)
- 45. <u>ICG Testhouse sequence</u> 2 turntable sequences from ifferent viewing heights, 36 images each, resolution 1000x750, color (Formats: PPM)
- 46. Institute of Computer Graphics and Vision
- 47. <u>IEN Image Library</u> 1000+ images, mostly outdoor sequences (Formats: raw, ppm)
- 48. <u>INRIA's Syntim images database</u> 15 color image of simple objects (Formats: gif)
- **49. INRIA**
- 50. INRIA's Syntim stereo databases 34 calibrated color stereo pairs (Formats: gif)
- 51. <u>Image Analysis Laboratory</u> Images obtained from a variety of imaging modalities -- raw CFA images, range images and a host of "medical images". (Formats: homebrew)
- 52. Image Analysis Laboratory
- 53. <u>Image Database</u> An image database including some textures
- 54. <u>JAFFE Facial Expression Image Database</u> The JAFFE database consists of 213 images of Japanese female subjects posing 6 basic facial expressions as well as a neutral pose. Ratings on emotion adjectives are also available, free of charge, for research purposes. (Formats: TIFF Grayscale images.)
- 55. ATR Research, Kyoto, Japan
- 56. JISCT Stereo Evaluation 44 image pairs. These data have been used in an evaluation of stereo analysis, as described in the April 1993 ARPA Image Understanding Workshop paper ``The JISCT Stereo Evaluation'' by R.C.Bolles, H.H.Baker, and M.J.Hannah, 263--274 (Formats: SSI)
- 57. MIT Vision Texture Image archive (100+ images) (Formats: ppm)
- 58. MIT face images and more hundreds of images (Formats: homebrew)
- 59. <u>Machine Vision</u> Images from the textbook by Jain, Kasturi, Schunck (20+ images) (Formats: GIF TIFF)
- 60. <u>Mammography Image Databases</u> 100 or more images of mammograms with ground truth. Additional images available by request, and links to several other mammography databases are provided. (Formats: homebrew)
- 61. ftp://ftp.cps.msu.edu/pub/prip many images (Formats: unknown)

- 62. <u>Middlebury Stereo Data Sets with Ground Truth</u> Six multi-frame stereo data sets of scenes containing planar regions. Each data set contains 9 color images and subpixel-accuracy ground-truth data. (Formats: ppm)
- 63. Middlebury Stereo Vision Research Page Middlebury College
- 64. <u>Modis Airborne simulator, Gallery and data set</u> High Altitude Imagery from around the world for environmental modeling in support of NASA EOS program (Formats: JPG and HDF)
- 65. NIST Fingerprint and handwriting datasets thousands of images (Formats: unknown)
- 66. NIST Fingerprint data compressed multipart uuencoded tar file
- 67. <u>NLM HyperDoc Visible Human Project</u> Color, CAT and MRI image samples over 30 images (Formats: jpeg)
- 68. <u>National Design Repository</u> Over 55,000 3D CAD and solid models of (mostly) mechanical/machined engineerign designs. (Formats: gif,vrml,wrl,stp,sat)
- 69. Geometric & Intelligent Computing Laboratory
- 70. OSU (MSU) 3D Object Model Database several sets of 3D object models collected over several years to use in object recognition research (Formats: homebrew, vrml)
- 71. OSU (MSU/WSU) Range Image Database Hundreds of real and synthetic images (Formats: gif, homebrew)
- 72. <u>OSU/SAMPL Database: Range Images, 3D Models, Stills, Motion Sequences</u> Over 1000 range images, 3D object models, still images and motion sequences (Formats: gif, ppm, vrml, homebrew)
- 73. Signal Analysis and Machine Perception Laboratory
- 74. <u>Otago Optical Flow Evaluation Sequences</u> Synthetic and real sequences with machine-readable ground truth optical flow fields, plus tools to generate ground truth for new sequences. (Formats: ppm,tif,homebrew)
- 75. Vision Research Group
- 76. ftp://ftp.limsi.fr/pub/quenot/opflow/testdata/piv/ Real and synthetic image sequences used for testing a Particle Image Velocimetry application. These images may be used for the test of optical flow and image matching algorithms. (Formats: pgm (raw))
- 77. LIMSI-CNRS/CHM/IMM/vision
- 78. LIMSI-CNRS
- 79. <u>Photometric 3D Surface Texture Database</u> This is the first 3D texture database which provides both full real surface rotations and registered photometric stereo data (30 textures, 1680 images). (Formats: TIFF)
- 80. <u>SEQUENCES FOR OPTICAL FLOW ANALYSIS (SOFA)</u> 9 synthetic sequences designed for testing motion analysis applications, including full ground truth of motion and camera parameters. (Formats: gif)
- 81. Computer Vision Group
- 82. <u>Sequences for Flow Based Reconstruction</u> synthetic sequence for testing structure from motion algorithms (Formats: pgm)

- 83. <u>Stereo Images with Ground Truth Disparity and Occlusion</u> a small set of synthetic images of a hallway with varying amounts of noise added. Use these images to benchmark your stereo algorithm. (Formats: raw, viff (khoros), or tiff)
- 84. <u>Stuttgart Range Image Database</u> A collection of synthetic range images taken from high-resolution polygonal models available on the web (Formats: homebrew)
- 85. Department Image Understanding
- 86. The AR Face Database Contains over 4,000 color images corresponding to 126 people's faces (70 men and 56 women). Frontal views with variations in facial expressions, illumination, and occlusions. (Formats: RAW (RGB 24-bit))
- 87. Purdue Robot Vision Lab
- 88. <u>The MIT-CSAIL Database of Objects and Scenes</u> Database for testing multiclass object detection and scene recognition algorithms. Over 72,000 images with 2873 annotated frames. More than 50 annotated object classes. (Formats: jpg)
- 89. The RVL SPEC-DB (SPECularity DataBase) A collection of over 300 real images of 100 objects taken under three different illuminaiton conditions (Diffuse/Ambient/Directed). -- Use these images to test algorithms for detecting and compensating specular highlights in color images. (Formats: TIFF)
- 90. Robot Vision Laboratory
- 91. <u>The Xm2vts database</u> The XM2VTSDB contains four digital recordings of 295 people taken over a period of four months. This database contains both image and video data of faces.
- 92. Centre for Vision, Speech and Signal Processing
- 93. <u>Traffic Image Sequences and 'Marbled Block' Sequence</u> thousands of frames of digitized traffic image sequences as well as the 'Marbled Block' sequence (grayscale images) (Formats: GIF)
- 94. IAKS/KOGS
- 95. U Bern Face images hundreds of images (Formats: Sun rasterfile)
- 96. U Michigan textures (Formats: compressed raw)
- 97. <u>U Oulu wood and knots database</u> Includes classifications 1000+ color images (Formats: ppm)
- 98. <u>UCID an Uncompressed Colour Image Database</u> a benchmark database for image retrieval with predefined ground truth. (Formats: tiff)
- 99. <u>UMass Vision Image Archive</u> Large image database with aerial, space, stereo, medical images and more. (Formats: homebrew)
- 100. UNC's 3D image database many images (Formats: GIF)
- 101. <u>USF Range Image Data with Segmentation Ground Truth</u> 80 image sets (Formats: Sun rasterimage)
- 102. <u>University of Oulu Physics-based Face Database</u> contains color images of faces under different illuminants and camera calibration conditions as well as skin spectral reflectance measurements of each person.
- 103. Machine Vision and Media Processing Unit
- 104. <u>University of Oulu Texture Database</u> Database of 320 surface textures, each captured under three illuminants, six spatial resolutions and nine rotation angles. A set

- of test suites is also provided so that texture segmentation, classification, and retrieval algorithms can be tested in a standard manner. (Formats: bmp, ras, xv)
- 105. Machine Vision Group
- 106. Usenix face database - Thousands of face images from many different sites (circa 994)
- 107. View Sphere Database - Images of 8 objects seen from many different view points. The view sphere is sampled using a geodesic with 172 images/sphere. Two sets for training and testing are available. (Formats: ppm)
- 108. PRIMA, GRAVIR
- 109. Vision-list Imagery Archive - Many images, many formats
- Wiry Object Recognition Database Thousands of images of a cart, ladder, 110. stool, bicycle, chairs, and cluttered scenes with ground truth labelings of edges and regions. (Formats: jpg)
- 111. 3D Vision Group
- Yale Face Database 165 images (15 individuals) with different lighting, 112. expression, and occlusion configurations.
- 113. Yale Face Database B - 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses x 64 illumination conditions). (Formats: PGM)
- Center for Computational Vision and Control 114.
- DeepMind QA Corpus Textual QA corpus from CNN and DailyMail. More than 115 300K documents in total. Paper for reference.
- 116. YouTube-8M Dataset - YouTube-8M is a large-scale labeled video dataset that consists of 8 million YouTube video IDs and associated labels from a diverse vocabulary of 4800 visual entities.
- 117. Open Images dataset - Open Images is a dataset of ~9 million URLs to images that have been annotated with labels spanning over 6000 categories.

Frameworks

1.	<u>Caffe</u>	11. <u>Deepnet</u>	
2.	Torch7	12. <u>Deeppy</u>	
3.	<u>Theano</u>	13. <u>JavaNN</u>	
4.	<u>cuda-convnet</u>	14. <u>hebel</u>	
5.	<u>convetjs</u>	15. <u>Mocha.jl</u>	
6.	Ccv	16. OpenDL	
7.	<u>NuPIC</u>	17. <u>cuDNN</u>	
8.	DeepLearning4J	18. <u>MGL</u>	
9.	<u>Brain</u>	19. <u>Knet.jl</u>	
10. <u>DeepLearnToolbox</u>			

- 20. Nvidia DIGITS a web app based on Caffe
- 21. Neon Python based Deep Learning Framework
- 22. Keras Theano based Deep Learning Library

- 23. Chainer A flexible framework of neural networks for deep learning
- 24. RNNLM Toolkit
- 25. RNNLIB A recurrent neural network library
- 26. char-rnn
- 27. MatConvNet: CNNs for MATLAB
- 28. Minerva a fast and flexible tool for deep learning on multi-GPU
- 29. Brainstorm Fast, flexible and fun neural networks.
- 30. <u>Tensorflow Open source software library for numerical computation using data flow graphs</u>
- 31. DMTK Microsoft Distributed Machine Learning Tookit
- 32. Scikit Flow Simplified interface for TensorFlow (mimicking Scikit Learn)
- 33. MXnet Lightweight, Portable, Flexible Distributed/Mobile Deep Learning framework
- 34. Veles Samsung Distributed machine learning platform
- 35. Marvin A Minimalist GPU-only N-Dimensional ConvNets Framework
- 36. Apache SINGA A General Distributed Deep Learning Platform
- 37. DSSTNE Amazon's library for building Deep Learning models
- 38. SyntaxNet Google's syntactic parser A TensorFlow dependency library
- 39. mlpack A scalable Machine Learning library
- 40. Torchnet Torch based Deep Learning Library
- 41. Paddle PArallel Distributed Deep LEarning by Baidu
- 42. NeuPy Theano based Python library for ANN and Deep Learning
- 43. <u>Lasagne a lightweight library to build and train neural networks in Theano</u>
- 44. <u>nolearn wrappers and abstractions around existing neural network libraries, most notably Lasagne</u>
- 45. Sonnet a library for constructing neural networks by Google's DeepMind
- 46. <u>PyTorch Tensors and Dynamic neural networks in Python with strong GPU</u> acceleration
- 47. CNTK Microsoft Cognitive Toolkit

Miscellaneous

- 1. Google Plus Deep Learning Community
- 2. Caffe Webinar
- 3. 100 Best Github Resources in Github for DL
- 4. Word2Vec
- 5. Caffe DockerFile
- 6. TorontoDeepLEarning convnet
- 7. gfx.js
- 8. Torch7 Cheat sheet
- 9. <u>Misc from MIT's 'Advanced Natural Language Processing' course</u>
- 10. Misc from MIT's 'Machine Learning' course

- 11. Misc from MIT's 'Networks for Learning: Regression and Classification' course
- 12. Misc from MIT's 'Neural Coding and Perception of Sound' course
- 13. Implementing a Distributed Deep Learning Network over Spark
- 14. A chess AI that learns to play chess using deep learning.
- 15. <u>Reproducing the results of "Playing Atari with Deep Reinforcement Learning" by DeepMind</u>
- 16. Wiki2Vec. Getting Word2vec vectors for entities and word from Wikipedia Dumps
- 17. The original code from the DeepMind article + tweaks
- 18. Google deepdream Neural Network art
- 19. An efficient, batched LSTM.
- 20. A recurrent neural network designed to generate classical music.
- 21. Memory Networks Implementations Facebook
- 22. Face recognition with Google's FaceNet deep neural network.
- 23. Basic digit recognition neural network
- 24. Emotion Recognition API Demo Microsoft
- 25. Proof of concept for loading Caffe models in TensorFlow
- 26. YOLO: Real-Time Object Detection
- 27. <u>AlphaGo A replication of DeepMind's 2016 Nature publication, "Mastering the game of Go with deep neural networks and tree search"</u>
- 28. Machine Learning for Software Engineers
- 29. Machine Learning is Fun!
- 30. Siraj Raval's Deep Learning tutorials
- 31. <u>Dockerface</u> Easy to install and use deep learning Faster R-CNN face detection for images and video in a docker container.
- 32. <u>Awesome Deep Learning Music</u> Curated list of articles related to deep learning scientific research applied to music

Deep Learning Papers Reading Roadmap

https://github.com/songrotek/Deep-Learning-Papers-Reading-Roadmap

If you are a newcomer to the Deep Learning area, the first question you may have is "Which paper should I start reading from?"

Here is a reading roadmap of Deep Learning papers!

The roadmap is constructed in accordance with the following four guidelines:

- From outline to detail
- From old to state-of-the-art
- from generic to specific areas
- focus on state-of-the-art

You will find many papers that are quite new but really worth reading.

I would continue adding papers to this roadmap.

1 Deep Learning History and Basics

1.0 Book

1.1 Survey

[1] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." Nature 521.7553 (2015): 436-444. [pdf] (Three Giants' Survey) ☆☆☆☆☆

1.2 Deep Belief Network(DBN)(Milestone of Deep Learning Eve)

[2] Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets." Neural computation 18.7 (2006): 1527-1554. [pdf](Deep Learning Eve) $\swarrow \swarrow \swarrow \swarrow$

[3] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." Science 313.5786 (2006): 504-507. [pdf] (Milestone, Show the promise of deep learning) $\[\] \] \[\] \] \]$

1.3 ImageNet Evolution (Deep Learning broke out from here)

- [5] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014). [pdf] (VGGNet,Neural Networks become very deep!) \$\frac{1}{2} \frac{1}{2} \frac{1}

1.4 Speech Recognition Evolution

- [9] Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks." 2013 IEEE international conference on acoustics, speech and signal processing. IEEE, 2013. [pdf] (RNN) $^{\coloredge}$ $^{\coloredge}$ $^{\coloredge}$
- [10] Graves, Alex, and Navdeep Jaitly. "Towards End-To-End Speech Recognition with Recurrent Neural Networks." ICML. Vol. 14. 2014. [pdf] 💢 🏠 🛣
- [11] Sak, Haşim, et al. "Fast and accurate recurrent neural network acoustic models for speech recognition." arXiv preprint arXiv:1507.06947 (2015). [pdf] (Google Speech Recognition System) \(\frac{1}{2} \) \(\frac{1}{2} \)

[12] Amodei, Dario, et al. "Deep speech 2: End-to-end speech recognition in english and mandarin." arXiv preprint arXiv:1512.02595 (2015). [pdf] (Baidu Speech Recognition System)

[13] W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu, G. Zweig "Achieving Human Parity in Conversational Speech Recognition." arXiv preprint arXiv:1610.05256 (2016). [pdf] (State-of-the-art in speech recognition, Microsoft) $^{\downarrow}$ $^{\downarrow}$ $^{\downarrow}$ $^{\downarrow}$ $^{\downarrow}$

After reading above papers, you will have a basic understanding of the Deep Learning history, the basic architectures of Deep Learning model(including CNN, RNN, LSTM) and how deep learning can be applied to image and speech recognition issues. The following papers will take you in-depth understanding of the Deep Learning method, Deep Learning in different areas of application and the frontiers. I suggest that you can choose the following papers based on your interests and research direction. #2 Deep Learning Method

2.1 Model

[14] Hinton, Geoffrey E., et al. "Improving neural networks by preventing coadaptation of feature detectors." arXiv preprint arXiv:1207.0580 (2012). [pdf] (Dropout) $\swarrow \swarrow \swarrow \swarrow$

[15] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." Journal of Machine Learning Research 15.1 (2014): 1929-1958. [pdf] 公公公

[17] Ba, Jimmy Lei, Jamie Ryan Kiros, and Geoffrey E. Hinton. "Layer normalization." arXiv preprint arXiv:1607.06450 (2016). [pdf] (Update of Batch Normalization) ななななな

[18] Courbariaux, Matthieu, et al. "Binarized Neural Networks: Training Neural Networks with Weights and Activations Constrained to+ 1 or−1." [pdf] (New Model,Fast) ☆ ☆ ☆

[20] Chen, Tianqi, Ian Goodfellow, and Jonathon Shlens. "Net2net: Accelerating learning via knowledge transfer." arXiv preprint arXiv:1511.05641

[21] Wei, Tao, et al. "Network Morphism." arXiv preprint arXiv:1603.01670 (2016). [pdf] (Modify previously trained network to reduce training epochs) \(\frac{1}{2} \frac{1}{

2.2 Optimization

- [22] Sutskever, Ilya, et al. "On the importance of initialization and momentum in deep learning." ICML (3) 28 (2013): 1139-1147. [pdf] (Momentum optimizer) 🖒 🏠
- [24] Andrychowicz, Marcin, et al. "Learning to learn by gradient descent by gradient descent." arXiv preprint arXiv:1606.04474 (2016). [pdf] (Neural Optimizer,Amazing Work) ☆ ☆ ☆ ☆ ☆

[26] landola, Forrest N., et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 1MB model size." arXiv preprint arXiv:1602.07360 (2016). [pdf] (Also a new direction to optimize NN,DeePhi Tech Startup) \$\frac{1}{2} \frac{1}{2} \frac{1}{2}

2.3 Unsupervised Learning / Deep Generative Model

[27] Le, Quoc V. "Building high-level features using large scale unsupervised learning." 2013 IEEE international conference on acoustics, speech and signal processing. IEEE, 2013. [pdf] (Milestone, Andrew Ng, Google Brain Project, Cat) ☆ ☆ ☆ ☆

- [28] Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." arXiv preprint arXiv:1312.6114 (2013). [pdf](VAE) ☆☆☆☆

- [30] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015). [pdf] (DCGAN) \$\frac{1}{2} \frac{1}{2} \frac{
- [32] Oord, Aaron van den, Nal Kalchbrenner, and Koray Kavukcuoglu. "Pixel recurrent neural networks." arXiv preprint arXiv:1601.06759 (2016). [pdf] (PixelRNN) ☆ ☆ ☆ ☆
- [33] Oord, Aaron van den, et al. "Conditional image generation with PixelCNN decoders." arXiv preprint arXiv:1606.05328 (2016). [pdf] (PixelCNN) なななな

2.4 RNN / Sequence-to-Sequence Model

- [35] Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint arXiv:1406.1078 (2014). [pdf] (First Seq-to-Seq Paper) ☆ ☆ ☆ ☆
- [36] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems. 2014. [pdf] (Outstanding Work) $\swarrow \swarrow \swarrow \swarrow \swarrow \swarrow \swarrow$
- [38] Vinyals, Oriol, and Quoc Le. "A neural conversational model." arXiv preprint arXiv:1506.05869 (2015). [pdf] (Seq-to-Seq on Chatbot) ☆ ☆ ☆

2.5 Neural Turing Machine

[42] Sukhbaatar, Sainbayar, Jason Weston, and Rob Fergus. "**End-to-end memory networks**." Advances in neural information processing systems. 2015. [pdf] 公公公公

[43] Vinyals, Oriol, Meire Fortunato, and Navdeep Jaitly. "**Pointer networks**." Advances in Neural Information Processing Systems. 2015. [pdf] ☆ ☆ ☆ ☆

[44] Graves, Alex, et al. "Hybrid computing using a neural network with dynamic external memory." Nature (2016). [pdf](Milestone,combine above papers' ideas) なななななな

2.6 Deep Reinforcement Learning

[45] Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." arXiv preprint arXiv:1312.5602 (2013). [pdf]) (First Paper named deep reinforcement learning) ☆☆☆☆

[46] Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." Nature 518.7540 (2015): 529-533. [pdf] (Milestone) ☆☆☆☆☆

[47] Wang, Ziyu, Nando de Freitas, and Marc Lanctot. "Dueling network architectures for deep reinforcement learning." arXiv preprint arXiv:1511.06581 (2015). [pdf] (ICLR best paper,great idea) ☆☆☆☆

[48] Mnih, Volodymyr, et al. "Asynchronous methods for deep reinforcement learning." arXiv preprint arXiv:1602.01783 (2016). [pdf] (State-of-the-art method) ななななな

[49] Lillicrap, Timothy P., et al. "Continuous control with deep reinforcement learning." arXiv preprint arXiv:1509.02971 (2015). [pdf] (DDPG) $^{\ }$

[50] Gu, Shixiang, et al. "Continuous Deep Q-Learning with Model-based Acceleration." arXiv preprint arXiv:1603.00748 (2016). [pdf] (NAF) $\stackrel{\wedge}{\sim}$ $\stackrel{\wedge}{\sim}$ $\stackrel{\wedge}{\sim}$

2.7 Deep Transfer Learning / Lifelong Learning / especially for RL

[55] Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean. "Distilling the knowledge in a neural network." arXiv preprint arXiv:1503.02531 (2015). [pdf] (Godfather's Work) ☆☆☆☆

[56] Rusu, Andrei A., et al. "Policy distillation." arXiv preprint arXiv:1511.06295 (2015). [pdf] (RL domain) 公公公

[57] Parisotto, Emilio, Jimmy Lei Ba, and Ruslan Salakhutdinov. "Actor-mimic: Deep multitask and transfer reinforcement learning." arXiv preprint arXiv:1511.06342 (2015). [pdf] (RL domain) \(\frac{1}{2} \frac{1}{2} \frac{1}{2} \)

[58] Rusu, Andrei A., et al. "Progressive neural networks." arXiv preprint arXiv:1606.04671 (2016). [pdf] (Outstanding Work, A novel idea) ☆☆☆☆☆

2.8 One Shot Deep Learning

[59] Lake, Brenden M., Ruslan Salakhutdinov, and Joshua B. Tenenbaum. "Human-level concept learning through probabilistic program induction." Science 350.6266 (2015): 1332-1338. [pdf] (No Deep Learning,but worth reading) $^{\wedge}$ $^{\wedge}$ $^{\wedge}$ $^{\wedge}$ $^{\wedge}$ $^{\wedge}$

[60] Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov. "Siamese Neural Networks for One-shot Image Recognition." (2015) [pdf]

[63] Hariharan, Bharath, and Ross Girshick. "Low-shot visual object recognition." arXiv preprint arXiv:1606.02819 (2016). [pdf](A step to large data) 公公公公

3 Applications

3.1 NLP(Natural Language Processing)

- [1] Antoine Bordes, et al. "Joint Learning of Words and Meaning Representations for Open-Text Semantic Parsing." AISTATS(2012) [pdf]
- [2] Mikolov, et al. "Distributed representations of words and phrases and their compositionality." ANIPS(2013): 3111-3119 [pdf] (word2vec) ななな
- [4] Ankit Kumar, et al. ""Ask Me Anything: Dynamic Memory Networks for Natural Language Processing." arXiv preprint arXiv:1506.07285(2015) [pdf]
- [5] Yoon Kim, et al. "Character-Aware Neural Language Models." NIPS(2015) arXiv preprint arXiv:1508.06615(2015) [pdf] 公 公 公
- [6] Jason Weston, et al. "Towards Al-Complete Question Answering: A Set of Prerequisite Toy Tasks." arXiv preprint arXiv:1502.05698(2015) [pdf] (bAbl tasks) \(\frac{1}{2} \f
- [7] Karl Moritz Hermann, et al. "Teaching Machines to Read and Comprehend." arXiv preprint arXiv:1506.03340(2015) [pdf](CNN/DailyMail cloze style questions) \$\frac{1}{2} \frac{1}{2} \fra
- [8] Alexis Conneau, et al. "Very Deep Convolutional Networks for Natural Language Processing." arXiv preprint arXiv:1606.01781(2016) [pdf] (state-of-the-art in text classification) \$\frac{1}{2} \frac{1}{2} \frac{1}{2}\$

3.2 Object Detection

- [4] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE International Conference on Computer Vision. 2015. [pdf] なななな

- [5] Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015. [pdf] 公公公公

- [8] Dai, Jifeng, et al. "R-FCN: Object Detection via Region-based Fully Convolutional Networks." arXiv preprint arXiv:1605.06409 (2016). [pdf]
- [9] He, Gkioxari, et al. "**Mask R-CNN**" arXiv preprint arXiv:1703.06870 (2017). [pdf] なななな

3.3 Visual Tracking

- [1] Wang, Naiyan, and Dit-Yan Yeung. "Learning a deep compact image representation for visual tracking." Advances in neural information processing systems. 2013. [pdf] (First Paper to do visual tracking using Deep Learning, DLT Tracker) \$\frac{1}{2} \frac{1}{2} \frac{1}
- [2] Wang, Naiyan, et al. "Transferring rich feature hierarchies for robust visual tracking." arXiv preprint arXiv:1501.04587 (2015). [pdf] (SO-DLT) ☆ ☆ ☆ ☆
- [3] Wang, Lijun, et al. "Visual tracking with fully convolutional networks." Proceedings of the IEEE International Conference on Computer Vision. 2015. [pdf] (FCNT) ☆☆☆☆
- [4] Held, David, Sebastian Thrun, and Silvio Savarese. "Learning to Track at 100 FPS with Deep Regression Networks." arXiv preprint arXiv:1604.01802 (2016). [pdf] (GOTURN,Really fast as a deep learning method,but still far behind un-deep-learning methods) $\stackrel{\wedge}{\sim} \stackrel{\wedge}{\sim} \stackrel{\wedge}{\sim} \stackrel{\wedge}{\sim}$
- [5] Bertinetto, Luca, et al. "Fully-Convolutional Siamese Networks for Object Tracking." arXiv preprint arXiv:1606.09549 (2016). [pdf] (SiameseFC,New state-of-the-art for real-time object tracking) \$\frac{1}{2} \frac{1}{2} \frac{1}{2}
- [6] Martin Danelljan, Andreas Robinson, Fahad Khan, Michael Felsberg. "Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking." ECCV (2016) [pdf] (C-COT) \$\frac{1}{2} \sqrt{1} \sqrt{2} \sqrt{2}\$

[7] Nam, Hyeonseob, Mooyeol Baek, and Bohyung Han. "Modeling and Propagating CNNs in a Tree Structure for Visual Tracking." arXiv preprint arXiv:1608.07242 (2016). [pdf] (VOT2016 Winner,TCNN) $\stackrel{\wedge}{\sim} \stackrel{\wedge}{\sim} \stackrel{\wedge}{\sim} \stackrel{\wedge}{\sim}$

3.4 Image Caption

- [1] Farhadi,Ali,etal. "Every picture tells a story: Generating sentences from images". In Computer VisionECCV 2010. Springer Berlin Heidelberg:15-29, 2010. [pdf] 公分分
- [2] Kulkarni, Girish, et al. "Baby talk: Understanding and generating image descriptions". In Proceedings of the 24th CVPR, 2011. [pdf] なななない
- [3] Vinyals, Oriol, et al. "**Show and tell: A neural image caption generator**". In arXiv preprint arXiv:1411.4555, 2014. [pdf] 公 公
- [4] Donahue, Jeff, et al. "Long-term recurrent convolutional networks for visual recognition and description". In arXiv preprint arXiv:1411.4389 ,2014. [pdf]
- [5] Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating image descriptions". In arXiv preprint arXiv:1412.2306, 2014. [pdf] ななななな
- [6] Karpathy, Andrej, Armand Joulin, and Fei Fei F. Li. "Deep fragment embeddings for bidirectional image sentence mapping". In Advances in neural information processing systems, 2014. [pdf] $\swarrow \swarrow \swarrow \swarrow$
- [7] Fang, Hao, et al. "From captions to visual concepts and back". In arXiv preprint arXiv:1411.4952, 2014. [pdf] 公分公公公
- [8] Chen, Xinlei, and C. Lawrence Zitnick. "Learning a recurrent visual representation for image caption generation". In arXiv preprint arXiv:1411.5654, 2014. [pdf] 公 公 公 公
- [9] Mao, Junhua, et al. "Deep captioning with multimodal recurrent neural networks (m-rnn)". In arXiv preprint arXiv:1412.6632, 2014. [pdf] \(\frac{1}{2} \) \(\frac{1}{2} \)
- [10] Xu, Kelvin, et al. "Show, attend and tell: Neural image caption generation with visual attention". In arXiv preprint arXiv:1502.03044, 2015. [pdf] 💢 💢 🧘 🛣

3.5 Machine Translation

Some milestone papers are listed in RNN / Seq-to-Seq topic.

[1] Luong, Minh-Thang, et al. "Addressing the rare word problem in neural machine translation." arXiv preprint arXiv:1410.8206 (2014). [pdf] \(\frac{1}{2} \) \(\frac{1}{2} \) \(\frac{1}{2} \)

- [2] Sennrich, et al. "Neural Machine Translation of Rare Words with Subword Units". In arXiv preprint arXiv:1508.07909, 2015. [pdf] ☆ ☆ ☆
- [4] Chung, et al. "A Character-Level Decoder without Explicit Segmentation for Neural Machine Translation". In arXiv preprint arXiv:1603.06147, 2016. [pdf] \(\frac{1}{2} \)
- [5] Lee, et al. "Fully Character-Level Neural Machine Translation without Explicit Segmentation". In arXiv preprint arXiv:1610.03017, 2016. [pdf] ☆ ☆ ☆ ☆ ☆

3.6 Robotics

- [2] Levine, Sergey, et al. "End-to-end training of deep visuomotor policies." Journal of Machine Learning Research 17.39 (2016): 1-40. [pdf] 公公公公公

- [5] Zhu, Yuke, et al. "Target-driven Visual Navigation in Indoor Scenes using Deep Reinforcement Learning." arXiv preprint arXiv:1609.05143 (2016). [pdf] ☆ ☆ ☆ ☆
- [6] Yahya, Ali, et al. "Collective Robot Reinforcement Learning with Distributed Asynchronous Guided Policy Search." arXiv preprint arXiv:1610.00673 (2016). [pdf] ☆☆☆☆☆
- [8] A Rusu, M Vecerik, Thomas Rothörl, N Heess, R Pascanu, R Hadsell."Sim-to-Real Robot Learning from Pixels with Progressive Nets." arXiv preprint arXiv:1610.04286 (2016). [pdf] 公公公公

[9] Mirowski, Piotr, et al. "Learning to navigate in complex environments." arXiv preprint arXiv:1611.03673 (2016). [pdf] 公 ななな

3.7 Art

- [3] Zhu, Jun-Yan, et al. "Generative Visual Manipulation on the Natural Image Manifold." European Conference on Computer Vision. Springer International Publishing, 2016. [pdf] (iGAN) なななな
- [4] Champandard, Alex J. "Semantic Style Transfer and Turning Two-Bit Doodles into Fine Artworks." arXiv preprint arXiv:1603.01768 (2016). [pdf] (Neural Doodle) なななな
- [5] Zhang, Richard, Phillip Isola, and Alexei A. Efros. "Colorful Image Colorization." arXiv preprint arXiv:1603.08511 (2016). [pdf] 公公公公
- [6] Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." arXiv preprint arXiv:1603.08155 (2016). [pdf] 公公公公
- [8] Gatys, Leon and Ecker, et al. "Controlling Perceptual Factors in Neural Style Transfer." arXiv preprint arXiv:1611.07865 (2016). [pdf] (control style transfer over spatial location, colour information and across spatial scale) \(\frac{1}{2} \frac{1}{2}

3.8 Object Segmentation

- [5] Dai, J., He, K., Sun, J. "Instance-sensitive Fully Convolutional Networks." arXiv preprint arXiv:1603.08678 (2016). [pdf] ☆ ☆ ☆

Machine Learning & Deep Learning Tutorials

https://github.com/ujjwalkarn/Machine-Learning-Tutorials

- This repository contains a topic-wise curated list of Machine Learning and Deep Learning tutorials, articles and other resources. Other awesome lists can be found in this <u>list</u>.
- If you want to contribute to this list, please read <u>Contributing Guidelines</u>.
- Curated list of R tutorials for Data Science, NLP and Machine Learning.
- Curated list of Python tutorials for Data Science, NLP and Machine Learning.

Contents

- Miscellaneous
- Interview Resources
- Artificial Intelligence
- Genetic Algorithms
- Statistics
- Useful Blogs
- Resources on Quora
- Resources on Kaggle
- Cheat Sheets
- Classification
- Linear Regression
- Logistic Regression
- Model Validation using Resampling
 - Cross Validation
 - Bootstraping
- Deep Learning
 - o Frameworks
 - Feed Forward Networks
 - Recurrent Neural Nets, LSTM, GRU

- Restricted Boltzmann Machine, DBNs
- Autoencoders
- Convolutional Neural Nets
- Natural Language Processing
 - o Topic Modeling, LDA
 - o Word2Vec
- Computer Vision
- Support Vector Machine
- Reinforcement Learning
- Decision Trees
- Random Forest / Bagging
- Boosting
- Ensembles
- Stacking Models
- VC Dimension
- Bayesian Machine Learning
- Semi Supervised Learning
- Optimizations
- Other Useful Tutorials

Miscellaneous

Machine Learning for Software Engineers

- Dive into Machine Learning
- A curated list of awesome Machine Learning frameworks, libraries and software
- A curated list of awesome data visualization libraries and resources.
- An awesome Data Science repository to learn and apply for real world problems
- The Open Source Data Science Masters
- Machine Learning FAQs on Cross Validated
- <u>List of Machine Learning University Courses</u>
- Machine Learning algorithms that you should always have a strong understanding of
- <u>Difference between Linearly Independent, Orthogonal, and Uncorrelated Variables</u>
- <u>List of Machine Learning Concepts</u>
- Slides on Several Machine Learning Topics
- MIT Machine Learning Lecture Slides
- Comparison Supervised Learning Algorithms
- <u>Learning Data Science Fundamentals</u>
- Machine Learning mistakes to avoid
- Statistical Machine Learning Course
- TheAnalyticsEdge edX Notes and Codes
- In-depth introduction to machine learning in 15 hours of expert videos
- Have Fun With Machine Learning
- Twitter's Most Shared #machineLearning Content From The Past 7 Days

Interview Resources

- 41 Essential Machine Learning Interview Questions (with answers)
- How can a computer science graduate student prepare himself for data scientist interviews?
- How do I learn Machine Learning?

- FAQs about Data Science Interviews
- What are the key skills of a data scientist?

Artificial Intelligence

- Awesome Artificial Intelligence (GitHub Repo)
- UC Berkeley CS188 Intro to AI, Lecture Videos, 2
- MIT 6.034 Artificial Intelligence Lecture Videos, Complete Course
- edX course | Klein & Abbeel
- Udacity Course | Norvig & Thrun
- TED talks on Al

Genetic Algorithms

- Genetic Algorithms Wikipedia Page
- <u>Simple Implementation of Genetic Algorithms in Python (Part 1)</u>, <u>Part 2</u>
- Genetic Algorithms vs Artificial Neural Networks
- Genetic Algorithms Explained in Plain English
- Genetic Programming
 - Genetic Programming in Python (GitHub)
 - Genetic Alogorithms vs Genetic Programming (Quora), StackOverflow

Statistics

- <u>Stat Trek Website</u> A dedicated website to teach yourselves Statistics
- <u>Learn Statistics Using Python</u> Learn Statistics using an application-centric programming approach
- Statistics for Hackers | Slides | @jakevdp Slides by Jake VanderPlas
- Online Statistics Book An Interactive Multimedia Course for Studying Statistics
- What is a Sampling Distribution?

- Tutorials
 - AP Statistics Tutorial
 - Statistics and Probability Tutorial
 - Matrix Algebra Tutorial
- What is an Unbiased Estimator?
- Goodness of Fit Explained
- What are QQ Plots?
- OpenIntro Statistics Free PDF textbook

Useful Blogs

- Edwin Chen's Blog A blog about Math, stats, ML, crowdsourcing, data science
- The Data School Blog Data science for beginners!
- ML Wave A blog for Learning Machine Learning
- Andrej Karpathy A blog about Deep Learning and Data Science in general
- Colah's Blog Awesome Neural Networks Blog
- Alex Minnaar's Blog A blog about Machine Learning and Software Engineering
- Statistically Significant Andrew Landgraf's Data Science Blog
- Simply Statistics A blog by three biostatistics professors
- Yanir Seroussi's Blog A blog about Data Science and beyond
- <u>fastML</u> Machine learning made easy
- <u>Trevor Stephens Blog</u> Trevor Stephens Personal Page
- no free hunch | kaggle The Kaggle Blog about all things Data Science
- A Quantitative Journey | outlace learning quantitative applications
- r4stats analyze the world of data science, and to help people learn to use R
- <u>Variance Explained</u> David Robinson's Blog
- Al Junkie a blog about Artificial Intellingence

- Deep Learning Blog by Tim Dettmers Making deep learning accessible
- <u>J Alammar's Blog</u>- Blog posts about Machine Learning and Neural Nets
- Adam Geitgey Easiest Introduction to machine learning
- Ethen's Notebook Collection Continuously updated machine learning documentations (mainly in Python3). Contents include educational implementation of machine learning algorithms from scratch and open-source library usage

Resources on Quora

- Most Viewed Machine Learning writers
- Data Science Topic on Quora
- William Chen's Answers
- Michael Hochster's Answers
- Ricardo Vladimiro's Answers
- Storytelling with Statistics
- Data Science FAQs on Quora
- Machine Learning FAQs on Quora

Kaggle Competitions WriteUp

- How to almost win Kaggle Competitions
- Convolution Neural Networks for EEG detection
- Facebook Recruiting III Explained
- Predicting CTR with Online ML
- How to Rank 10% in Your First Kaggle Competition

Cheat Sheets

- Probability Cheat Sheet, Source
- Machine Learning Cheat Sheet

Classification

- <u>Does Balancing Classes Improve Classifier Performance?</u>
- What is Deviance?
- When to choose which machine learning classifier?
- What are the advantages of different classification algorithms?
- ROC and AUC Explained (related video)
- An introduction to ROC analysis
- Simple guide to confusion matrix terminology

Linear Regression

- General
 - Assumptions of Linear Regression, Stack Exchange
 - o <u>Linear Regression Comprehensive Resource</u>
 - Applying and Interpreting Linear Regression
 - What does having constant variance in a linear regression model mean?
 - o Difference between linear regression on y with x and x with y
 - <u>Is linear regression valid when the dependant variable is not normally</u> distributed?
- Multicollinearity and VIF
 - Dummy Variable Trap | Multicollinearity
 - Dealing with multicollinearity using VIFs
- Residual Analysis
 - Interpreting plot.lm() in R
 - o How to interpret a QQ plot?
 - o <u>Interpreting Residuals vs Fitted Plot</u>
- Outliers

- How should outliers be dealt with?
- Elastic Net
 - o Regularization and Variable Selection via the Elastic Net

Logistic Regression

- Logistic Regression Wiki
- Geometric Intuition of Logistic Regression
- Obtaining predicted categories (choosing threshold)
- Residuals in logistic regression
- <u>Difference between logit and probit models</u>, <u>Logistic Regression Wiki</u>, <u>Probit</u>
 Model Wiki
- Pseudo R2 for Logistic Regression, How to calculate, Other Details
- Guide to an in-depth understanding of logistic regression

Model Validation using Resampling

- Resampling Explained
- Partioning data set in R
- Implementing hold-out Validaion in R, 2
- Cross Validation
 - How to use cross-validation in predictive modeling
 - Training with Full dataset after CV?
 - Which CV method is best?
 - Variance Estimates in k-fold CV
 - Is CV a subsitute for Validation Set?
 - Choice of k in k-fold CV
 - CV for ensemble learning
 - o k-fold CV in R

- Good Resources
- Overfitting and Cross Validation
 - Preventing Overfitting the Cross Validation Data | Andrew Ng
 - Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation
 - CV for detecting and preventing Overfitting
 - How does CV overcome the Overfitting Problem

Bootstrapping

- o Why Bootstrapping Works?
- Good Animation
- Example of Bootstapping
- Understanding Bootstapping for Validation and Model Selection
- Cross Validation vs Bootstrap to estimate prediction error, Crossvalidation vs .632 bootstrapping to evaluate classification performance

Deep Learning

- A curated list of awesome Deep Learning tutorials, projects and communities
- Lots of Deep Learning Resources
- Interesting Deep Learning and NLP Projects (Stanford), Website
- Core Concepts of Deep Learning
- <u>Understanding Natural Language with Deep Neural Networks Using Torch</u>
- Stanford Deep Learning Tutorial
- Deep Learning FAQs on Quora
- Google+ Deep Learning Page
- Recent Reddit AMAs related to Deep Learning, Another AMA
- Where to Learn Deep Learning?
- Deep Learning nvidia concepts

- Introduction to Deep Learning Using Python (GitHub), Good Introduction Slides
- Video Lectures Oxford 2015, Video Lectures Summer School Montreal
- Deep Learning Software List
- Hacker's guide to Neural Nets
- Top arxiv Deep Learning Papers explained
- Geoff Hinton Youtube Vidoes on Deep Learning
- Awesome Deep Learning Reading List
- Deep Learning Comprehensive Website, Software
- deeplearning Tutorials
- AWESOME! Deep Learning Tutorial
- <u>Deep Learning Basics</u>
- Stanford Tutorials
- Train, Validation & Test in Artificial Neural Networks
- Artificial Neural Networks Tutorials
- Neural Networks FAQs on Stack Overflow
- Deep Learning Tutorials on deeplearning.net
- Neural Networks and Deep Learning Online Book
- Neural Machine Translation
 - Introduction to Neural Machine Translation with GPUs (part 1), Part
 2, Part 3
 - Deep Speech: Accurate Speech Recognition with GPU-Accelerated Deep Learning
- Deep Learning Frameworks
 - Torch vs. Theano
 - o dl4j vs. torch7 vs. theano
 - o Deep Learning Libraries by Language
 - o <u>Theano</u>

- Website
- Theano Introduction
- Theano Tutorial
- Good Theano Tutorial
- Logistic Regression using Theano for classifying digits
- MLP using Theano
- CNN using Theano
- RNNs using Theano
- LSTM for Sentiment Analysis in Theano
- RBM using Theano
- DBNs using Theano
- All Codes
- Deep Learning Implementation Tutorials Keras and Lasagne

o <u>Torch</u>

- Torch ML Tutorial, Code
- Intro to Torch
- Learning Torch GitHub Repo
- Awesome-Torch (Repository on GitHub)
- Machine Learning using Torch Oxford Univ, Code
- Torch Internals Overview
- Torch Cheatsheet
- Understanding Natural Language with Deep Neural Networks
 Using Torch

Caffe

- Deep Learning for Computer Vision with Caffe and cuDNN
- TensorFlow

- Website
- TensorFlow Examples for Beginners
- Stanford Tensorflow for Deep Learning Research Course
 - GitHub Repo
- Simplified Scikit-learn Style Interface to TensorFlow
- Learning TensorFlow GitHub Repo
- Benchmark TensorFlow GitHub
- Awesome TensorFlow List
- TensorFlow Book
- Android TensorFlow Machine Learning Example
 - GitHub Repo
- Creating Custom Model For Android Using TensorFlow
 - GitHub Repo
- Feed Forward Networks
 - A Quick Introduction to Neural Networks
 - Implementing a Neural Network from scratch, Code
 - Speeding up your Neural Network with Theano and the gpu, Code
 - Basic ANN Theory
 - Role of Bias in Neural Networks
 - Choosing number of hidden layers and nodes, 2,3
 - Backpropagation in Matrix Form
 - ANN implemented in C++ | Al Junkie
 - Simple Implementation
 - NN for Beginners
 - Regression and Classification with NNs (Slides)
 - Another Intro

- Recurrent and LSTM Networks
 - o awesome-rnn: list of resources (GitHub Repo)
 - o Recurrent Neural Net Tutorial Part 1, Part 2, Part 3, Code
 - NLP RNN Representations
 - o The Unreasonable effectiveness of RNNs, Torch Code, Python Code
 - Intro to RNN, LSTM
 - An application of RNN
 - o Optimizing RNN Performance
 - o Simple RNN
 - Auto-Generating Clickbait with RNN
 - Sequence Learning using RNN (Slides)
 - Machine Translation using RNN (Paper)
 - Music generation using RNNs (Keras)
 - Using RNN to create on-the-fly dialogue (Keras)
 - Long Short Term Memory (LSTM)
 - Understanding LSTM Networks
 - LSTM explained
 - Beginner's Guide to LSTM
 - Implementing LSTM from scratch, Python/Theano code
 - Torch Code for character-level language models using LSTM
 - LSTM for Kaggle EEG Detection competition (Torch Code)
 - LSTM for Sentiment Analysis in Theano
 - Deep Learning for Visual Q&A | LSTM | CNN, Code
 - Computer Responds to email using LSTM | Google
 - LSTM dramatically improves Google Voice Search, Another Article
 - Understanding Natural Language with LSTM Using Torch

- Torch code for Visual Question Answering using a CNN+LSTM model
- LSTM for Human Activity Recognition
- Gated Recurrent Units (GRU)
 - LSTM vs GRU
- <u>Time series forecasting with Sequence-to-Sequence (seq2seq) rnn</u> models
- Recursive Neural Network (not Recurrent)
 - Recursive Neural Tensor Network (RNTN)
 - word2vec, DBN, RNTN for Sentiment Analysis
- Restricted Boltzmann Machine
 - Beginner's Guide about RBMs
 - Another Good Tutorial
 - Introduction to RBMs
 - Hinton's Guide to Training RBMs
 - o RBMs in R
 - Deep Belief Networks Tutorial
 - word2vec, DBN, RNTN for Sentiment Analysis
- Autoencoders: Unsupervised (applies BackProp after setting target = input)
 - Andrew Ng Sparse Autoencoders pdf
 - <u>Deep Autoencoders Tutorial</u>
 - <u>Denoising Autoencoders</u>, <u>Theano Code</u>
 - o Stacked Denoising Autoencoders
- Convolutional Neural Networks
 - o An Intuitive Explanation of Convolutional Neural Networks
 - Awesome Deep Vision: List of Resources (GitHub)
 - Intro to CNNs

- Understanding CNN for NLP
- o Stanford Notes, Codes, GitHub
- JavaScript Library (Browser Based) for CNNs
- Using CNNs to detect facial keypoints
- Deep learning to classify business photos at Yelp
- Interview with Yann LeCun | Kaggle
- Visualising and Understanding CNNs

Natural Language Processing

- A curated list of speech and natural language processing resources
- Understanding Natural Language with Deep Neural Networks Using Torch
- tf-idf explained
- Interesting Deep Learning NLP Projects Stanford, Website
- NLP from Scratch | Google Paper
- Graph Based Semi Supervised Learning for NLP
- Bag of Words
 - Classification text with Bag of Words
- Topic Modeling
 - o LDA, LSA, Probabilistic LSA
 - What is a good explanation of Latent Dirichlet Allocation?
 - Awesome LDA Explanation!. Another good explanation
 - The LDA Buffet- Intuitive Explanation
 - <u>Difference between LSI and LDA</u>
 - Original LDA Paper
 - alpha and beta in LDA
 - Intuitive explanation of the Dirichlet distribution

- Topic modeling made just simple enough
- o Online LDA, Online LDA with Spark
- LDA in Scala, Part 2
- Segmentation of Twitter Timelines via Topic Modeling
- Topic Modeling of Twitter Followers

word2vec

- Google word2vec
- Bag of Words Model Wiki
- word2vec Tutorial
- A closer look at Skip Gram Modeling
- Skip Gram Model Tutorial, CBoW Model
- Word Vectors Kaggle Tutorial Python, Part 2
- Making sense of word2vec
- word2vec explained on deeplearning4j
- o Quora word2vec
- o Other Quora Resources, 2, 3
- word2vec, DBN, RNTN for Sentiment Analysis

Text Clustering

- How string clustering works
- <u>Levenshtein distance for measuring the difference between two sequences</u>
- o Text clustering with Levenshtein distances
- Text Classification
 - Classification Text with Bag of Words
- Language learning with NLP and reinforcement learning
- Kaggle Tutorial Bag of Words and Word vectors, Part 2, Part 3

- What would Shakespeare say (NLP Tutorial)
- A closer look at Skip Gram Modeling

Computer Vision

- Awesome computer vision (github)
- Awesome deep vision (github)

Support Vector Machine

- Highest Voted Questions about SVMs on Cross Validated
- Help me Understand SVMs!
- SVM in Layman's terms
- How does SVM Work | Comparisons
- A tutorial on SVMs
- Practical Guide to SVC, Slides
- Introductory Overview of SVMs
- Comparisons
 - SVMs > ANNs, ANNs > SVMs, Another Comparison
 - Trees > SVMs
 - Kernel Logistic Regression vs SVM
 - Logistic Regression vs SVM, 2, 3
- Optimization Algorithms in Support Vector Machines
- Variable Importance from SVM
- Software
 - o <u>LIBSVM</u>
 - Intro to SVM in R
- Kernels
 - o What are Kernels in ML and SVM?

- Intuition Behind Gaussian Kernel in SVMs?
- Probabilities post SVM
 - o Platt's Probabilistic Outputs for SVM
 - Platt Calibration Wiki
 - Why use Platts Scaling
 - Classifier Classification with Platt's Scaling

Reinforcement Learning

- Awesome Reinforcement Learning (GitHub)
- RL Tutorial Part 1, Part 2

Decision Trees

- Wikipedia Page Lots of Good Info
- FAQs about Decision Trees
- Brief Tour of Trees and Forests
- Tree Based Models in R
- How Decision Trees work?
- Weak side of Decision Trees
- Thorough Explanation and different algorithms
- What is entropy and information gain in the context of building decision trees?
- Slides Related to Decision Trees
- How do decision tree learning algorithms deal with missing values?
- <u>Using Surrogates to Improve Datasets with Missing Values</u>
- Good Article
- Are decision trees almost always binary trees?
- Pruning Decision Trees, Grafting of Decision Trees
- What is Deviance in context of Decision Trees?

- <u>Discover structure behind data with decision trees</u> Grow and plot a decision tree to automatically figure out hidden rules in your data
- Comparison of Different Algorithms
 - CART vs CTREE
 - Comparison of complexity or performance
 - CHAID vs CART, CART vs CHAID
 - Good Article on comparison

CART

- Recursive Partitioning Wikipedia
- CART Explained
- How to measure/rank "variable importance" when using CART?
- Pruning a Tree in R
- Does rpart use multivariate splits by default?
- FAQs about Recursive Partitioning

CTREE

- party package in R
- Show volumne in each node using ctree in R
- How to extract tree structure from ctree function?

CHAID

- Wikipedia Artice on CHAID
- Basic Introduction to CHAID
- Good Tutorial on CHAID

MARS

- Wikipedia Article on MARS
- Probabilistic Decision Trees
 - Bayesian Learning in Probabilistic Decision Trees

Probabilistic Trees Research Paper

Random Forest / Bagging

- Awesome Random Forest (GitHub)**
- How to tune RF parameters in practice?
- Measures of variable importance in random forests
- Compare R-squared from two different Random Forest models
- OOB Estimate Explained | RF vs LDA
- Evaluating Random Forests for Survival Analysis Using Prediction Error Curve
- Why doesn't Random Forest handle missing values in predictors?
- How to build random forests in R with missing (NA) values?
- FAQs about Random Forest, More FAQs
- Obtaining knowledge from a random forest
- Some Questions for R implementation, 2, 3

Boosting

- Boosting for Better Predictions
- Boosting Wikipedia Page
- Introduction to Boosted Trees | Tianqi Chen
- Gradient Boosting Machine
 - Gradiet Boosting Wiki
 - Guidelines for GBM parameters in R, Strategy to set parameters
 - Meaning of Interaction Depth, 2
 - o Role of n.minobsinnode parameter of GBM in R
 - o GBM in R
 - FAQs about GBM
 - GBM vs xgboost

- xgboost
 - xgboost tuning kaggle
 - xgboost vs gbm
 - xgboost survey
 - Practical XGBoost in Python online course (free)
- AdaBoost
 - AdaBoost Wiki, Python Code
 - AdaBoost Sparse Input Support
 - adaBag R package
 - <u>Tutorial</u>

Ensembles

- Wikipedia Article on Ensemble Learning
- Kaggle Ensembling Guide
- The Power of Simple Ensembles
- Ensemble Learning Intro
- Ensemble Learning Paper
- Ensembling models with R, Ensembling Regression Models in R, Intro to Ensembles in R
- Ensembling Models with caret
- Bagging vs Boosting vs Stacking
- Good Resources | Kaggle Africa Soil Property Prediction
- Boosting vs Bagging
- Resources for learning how to implement ensemble methods
- How are classifications merged in an ensemble classifier?

Stacking Models

- Stacking, Blending and Stacked Generalization
- Stacked Generalization (Stacking)
- Stacked Generalization: when does it work?
- Stacked Generalization Paper

Vapnik-Chervonenkis Dimension

- Wikipedia article on VC Dimension
- Intuitive Explanantion of VC Dimension
- Video explaining VC Dimension
- Introduction to VC Dimension
- FAQs about VC Dimension
- Do ensemble techniques increase VC-dimension?

Bayesian Machine Learning

- Bayesian Methods for Hackers (using pyMC)
- Should all Machine Learning be Bayesian?
- Tutorial on Bayesian Optimisation for Machine Learning
- Bayesian Reasoning and Deep Learning, Slides
- Bayesian Statistics Made Simple
- Kalman & Bayesian Filters in Python
- Markov Chain Wikipedia Page

Semi Supervised Learning

- Wikipedia article on Semi Supervised Learning
- <u>Tutorial on Semi Supervised Learning</u>
- Graph Based Semi Supervised Learning for NLP
- <u>Taxonomy</u>

- Video Tutorial Weka
- Unsupervised, Supervised and Semi Supervised learning
- Research Papers 1, 2, 3

Optimization

- Mean Variance Portfolio Optimization with R and Quadratic Programming
- Algorithms for Sparse Optimization and Machine Learning
- Optimization Algorithms in Machine Learning, Video Lecture
- Optimization Algorithms for Data Analysis
- Video Lectures on Optimization
- Optimization Algorithms in Support Vector Machines
- The Interplay of Optimization and Machine Learning Research
- Hyperopt tutorial for Optimizing Neural Networks' Hyperparameters

Other Tutorials

- For a collection of Data Science Tutorials using R, please refer to this list.
- For a collection of Data Science Tutorials using Python, please refer to this list.

Data science Python notebooks

Data science Python notebooks: Deep learning (TensorFlow, Theano, Caffe, Keras), scikit-learn, Kaggle, big data (Spark, Hadoop MapReduce, HDFS), matplotlib, pandas, NumPy, SciPy, Python essentials, AWS, and various command lines.

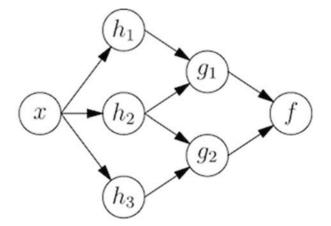
data-science-ipython-notebooks

https://github.com/donnemartin/data-science-ipython-notebooks

Index

- deep-learning
 - tensorflow
 - o theano
 - o keras
 - o caffe
- scikit-learn
- statistical-inference-scipy
- pandas
- matplotlib
- numpy
- python-data
- kaggle-and-business-analyses

- spark
- mapreduce-python
- amazon web services
- command lines
- misc
- notebook-installation
- credits
- contributing
- contact-info
- license



deep-learning

IPython Notebook(s) demonstrating deep learning functionality.



tensor-flow-tutorials

Additional TensorFlow tutorials:

- <u>pkmital/tensorflow_tutorials</u>
- nlintz/TensorFlow-Tutorials
- <u>alrojo/tensorflow-tutorial</u>
- BinRoot/TensorFlow-Book

Notebook	Description
tsf-basics	Learn basic operations in TensorFlow, a library for various kinds of perceptual and language understanding tasks from Google.
tsf-linear	Implement linear regression in TensorFlow.
tsf-logistic	Implement logistic regression in TensorFlow.
<u>tsf-nn</u>	Implement nearest neighboars in TensorFlow.
<u>tsf-alex</u>	Implement AlexNet in TensorFlow.
<u>tsf-cnn</u>	Implement convolutional neural networks in TensorFlow.
tsf-mlp	Implement multilayer perceptrons in TensorFlow.
<u>tsf-rnn</u>	Implement recurrent neural networks in TensorFlow.
tsf-gpu	Learn about basic multi-GPU computation in TensorFlow.
tsf-gviz	Learn about graph visualization in TensorFlow.

Notebook	Description
<u>tsf-lviz</u>	Learn about loss visualization in TensorFlow.

tensor-flow-exercises

Notebook	Description
tsf-not-mnist	Learn simple data curation by creating a pickle with formatted datasets for training, development and testing in TensorFlow.
tsf-fully- connected	Progressively train deeper and more accurate models using logistic regression and neural networks in TensorFlow.
tsf- regularization	Explore regularization techniques by training fully connected networks to classify notMNIST characters in TensorFlow.
tsf-convolutions	Create convolutional neural networks in TensorFlow.
tsf-word2vec	Train a skip-gram model over Text8 data in TensorFlow.
<u>tsf-lstm</u>	Train a LSTM character model over Text8 data in TensorFlow.



theano-tutorials

Notebook	Description
theano- intro	Intro to Theano, which allows you to define, optimize, and evaluate mathematical expressions involving multi-dimensional arrays efficiently. It can use GPUs and perform efficient symbolic differentiation.
theano- scan	Learn scans, a mechanism to perform loops in a Theano graph.

Notebook	Description
theano- logistic	Implement logistic regression in Theano.
theano-rnn	Implement recurrent neural networks in Theano.
theano-mlp	Implement multilayer perceptrons in Theano.



keras-tutorials

Notebook	Description
keras	Keras is an open source neural network library written in Python. It is capable of running on top of either Tensorflow or Theano.
<u>setup</u>	Learn about the tutorial goals and how to set up your Keras environment.
intro-deep- learning-ann	Get an intro to deep learning with Keras and Artificial Neural Networks (ANN).
<u>theano</u>	Learn about Theano by working with weights matrices and gradients.
<u>keras-otto</u>	Learn about Keras by looking at the Kaggle Otto challenge.
ann-mnist	Review a simple implementation of ANN for MNIST using Keras.
conv-nets	Learn about Convolutional Neural Networks (CNNs) with Keras.
conv-net-1	Recognize handwritten digits from MNIST using Keras - Part 1.

Notebook	Description
conv-net-2	Recognize handwritten digits from MNIST using Keras - Part 2.
<u>keras-models</u>	Use pre-trained models such as VGG16, VGG19, ResNet50, and Inception v3 with Keras.
<u>auto-encoders</u>	Learn about Autoencoders with Keras.
<u>rnn-lstm</u>	Learn about Recurrent Neural Networks (RNNs) with Keras.
<u>lstm-sentence-</u> gen	Learn about RNNs using Long Short Term Memory (LSTM) networks with Keras.

deep-learning-misc

Notebook	Description
<u>deep-</u> <u>dream</u>	Caffe-based computer vision program which uses a convolutional neural network to find and enhance patterns in images.



scikit-learn

IPython Notebook(s) demonstrating scikit-learn functionality.

Notebook	Description
intro	Intro notebook to scikit-learn. Scikit-learn adds Python support for large, multi-dimensional arrays and matrices, along with a large library of high-level mathematical functions to operate on these arrays.

Notebook	Description
<u>knn</u>	Implement k-nearest neighbors in scikit-learn.
linear-reg	Implement linear regression in scikit-learn.
<u>svm</u>	Implement support vector machine classifiers with and without kernels in scikit-learn.
random- forest	Implement random forest classifiers and regressors in scikit-learn.
k-means	Implement k-means clustering in scikit-learn.
pca	Implement principal component analysis in scikit-learn.
<u>gmm</u>	Implement Gaussian mixture models in scikit-learn.
validation	Implement validation and model selection in scikit-learn.



statistical-inference-scipy

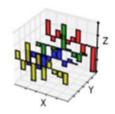
IPython Notebook(s) demonstrating statistical inference with SciPy functionality.

Notebook	Description
scipy	SciPy is a collection of mathematical algorithms and convenience functions built on the Numpy extension of Python. It adds significant power to the interactive Python session by providing the user with high-level commands and classes for manipulating and visualizing data.
effect-size	Explore statistics that quantify effect size by analyzing the difference in height between men and women. Uses data from the Behavioral Risk Factor Surveillance System (BRFSS) to estimate the mean and standard deviation of height for adult women and men in the United States.
sampling	Explore random sampling by analyzing the average weight of men and women in the United States using BRFSS data.
hypothesis	Explore hypothesis testing by analyzing the difference of first-born babies compared with others.









pandas

IPython Notebook(s) demonstrating pandas functionality.

Notebook	Description
pandas	Software library written for data manipulation and analysis in Python. Offers data structures and operations for manipulating numerical tables and time series.
github-data- wrangling	Learn how to load, clean, merge, and feature engineer by analyzing GitHub data from the Viz repo.

Notebook	Description
Introduction-to- Pandas	Introduction to Pandas.
Introducing- Pandas-Objects	Learn about Pandas objects.
Data Indexing and Selection	Learn about data indexing and selection in Pandas.
Operations-in- Pandas	Learn about operating on data in Pandas.
Missing-Values	Learn about handling missing data in Pandas.
<u>Hierarchical-</u> <u>Indexing</u>	Learn about hierarchical indexing in Pandas.
Concat-And- Append	Learn about combining datasets: concat and append in Pandas.
Merge-and-Join	Learn about combining datasets: merge and join in Pandas.
Aggregation-and- Grouping	Learn about aggregation and grouping in Pandas.
<u>Pivot-Tables</u>	Learn about pivot tables in Pandas.
Working-With- Strings	Learn about vectorized string operations in Pandas.
Working-with- Time-Series	Learn about working with time series in pandas.
Performance-Eval- and-Query	Learn about high-performance Pandas: eval() and query() in Pandas.



matplotlib

IPython Notebook(s) demonstrating matplotlib functionality.

Notebook	Description
matplotlib	Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.
matplotlib-applied	Apply matplotlib visualizations to Kaggle competitions for exploratory data analysis. Learn how to create bar plots, histograms, subplot2grid, normalized plots, scatter plots, subplots, and kernel density estimation plots.
Introduction-To- Matplotlib	Introduction to Matplotlib.
Simple-Line-Plots	Learn about simple line plots in Matplotlib.
Simple-Scatter- Plots	Learn about simple scatter plots in Matplotlib.
Errorbars.ipynb	Learn about visualizing errors in Matplotlib.
<u>Density-and-</u> <u>Contour-Plots</u>	Learn about density and contour plots in Matplotlib.
<u>Histograms-and-</u> <u>Binnings</u>	Learn about histograms, binnings, and density in Matplotlib.
Customizing- Legends	Learn about customizing plot legends in Matplotlib.

Notebook	Description
Customizing- Colorbars	Learn about customizing colorbars in Matplotlib.
Multiple-Subplots	Learn about multiple subplots in Matplotlib.
<u>Text-and-</u> <u>Annotation</u>	Learn about text and annotation in Matplotlib.
<u>Customizing-Ticks</u>	Learn about customizing ticks in Matplotlib.
<u>Settings-and-</u> <u>Stylesheets</u>	Learn about customizing Matplotlib: configurations and stylesheets.
Three- Dimensional- Plotting	Learn about three-dimensional plotting in Matplotlib.
Geographic-Data- With-Basemap	Learn about geographic data with basemap in Matplotlib.
<u>Visualization-With-</u> <u>Seaborn</u>	Learn about visualization with Seaborn.



numpy

IPython Notebook(s) demonstrating NumPy functionality.

Notebook	Description
numpy	Adds Python support for large, multi-dimensional arrays and matrices, along with a large library of high-level mathematical functions to operate on these arrays.
Introduction-to- NumPy	Introduction to NumPy.
<u>Understanding-Data-</u> <u>Types</u>	Learn about data types in Python.
The-Basics-Of- NumPy-Arrays	Learn about the basics of NumPy arrays.
Computation-on- arrays-ufuncs	Learn about computations on NumPy arrays: universal functions.
Computation-on- arrays-aggregates	Learn about aggregations: min, max, and everything in between in NumPy.
Computation-on- arrays-broadcasting	Learn about computation on arrays: broadcasting in NumPy.
Boolean-Arrays-and- Masks	Learn about comparisons, masks, and boolean logic in NumPy.
Fancy-Indexing	Learn about fancy indexing in NumPy.
Sorting	Learn about sorting arrays in NumPy.
Structured-Data- NumPy	Learn about structured data: NumPy's structured arrays.



python-data

IPython Notebook(s) demonstrating Python functionality geared towards data analysis.

Notebook	Description
<u>data</u> <u>structures</u>	Learn Python basics with tuples, lists, dicts, sets.
data structure utilities	Learn Python operations such as slice, range, xrange, bisect, sort, sorted, reversed, enumerate, zip, list comprehensions.
functions	Learn about more advanced Python features: Functions as objects, lambda functions, closures, *args, **kwargs currying, generators, generator expressions, itertools.
<u>datetime</u>	Learn how to work with Python dates and times: datetime, strftime, strptime, timedelta.
logging	Learn about Python logging with RotatingFileHandler and TimedRotatingFileHandler.
pdb	Learn how to debug in Python with the interactive source code debugger.
unit tests	Learn how to test in Python with Nose unit tests.



kaggle-and-business-analyses

IPython Notebook(s) used in <u>kaggle</u> competitions and business analyses.

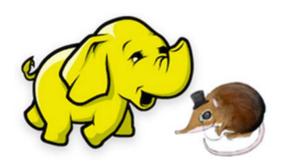
Notebook	Description
titanic	Predict survival on the Titanic. Learn data cleaning, exploratory data analysis, and machine learning.
<u>churn-</u> <u>analysis</u>	Predict customer churn. Exercise logistic regression, gradient boosting classifers, support vector machines, random forests, and k-nearest-neighbors. Includes discussions of confusion matrices, ROC plots, feature importances, prediction probabilities, and calibration/descrimination.



spark

IPython Notebook(s) demonstrating spark and HDFS functionality.

Notebook	Description
<u>spark</u>	In-memory cluster computing framework, up to 100 times faster for certain applications and is well suited for machine learning algorithms.
<u>hdfs</u>	Reliably stores very large files across machines in a large cluster.



mapreduce-python

IPython Notebook(s) demonstrating Hadoop MapReduce with mrjob functionality.

Notebook	Description
mapreduce- python	Runs MapReduce jobs in Python, executing jobs locally or on Hadoop clusters. Demonstrates Hadoop Streaming in Python code with unit test and mrjob config file to analyze Amazon S3 bucket logs on Elastic MapReduce. Disco is another python-based alternative.



aws

IPython Notebook(s) demonstrating Amazon Web Services (AWS) and AWS tools functionality.

Also check out:

- <u>SAWS</u>: A Supercharged AWS command line interface (CLI).
- <u>Awesome AWS</u>: A curated list of libraries, open source repos, guides, blogs, and other resources.

Notebook	Description
<u>boto</u>	Official AWS SDK for Python.
s3cmd	Interacts with S3 through the command line.
s3distcp	Combines smaller files and aggregates them together by taking in a pattern and target file. S3DistCp can also be used to transfer large volumes of data from S3 to your Hadoop cluster.

Notebook	Description
<u>s3-parallel-</u> <u>put</u>	Uploads multiple files to S3 in parallel.
<u>redshift</u>	Acts as a fast data warehouse built on top of technology from massive parallel processing (MPP).
<u>kinesis</u>	Streams data in real time with the ability to process thousands of data streams per second.
lambda	Runs code in response to events, automatically managing compute resources.



commands

IPython Notebook(s) demonstrating various command lines for Linux, Git, etc.

Notebook	Description
<u>linux</u>	Unix-like and mostly POSIX-compliant computer operating system. Disk usage, splitting files, grep, sed, curl, viewing running processes, terminal syntax highlighting, and Vim.
<u>anaconda</u>	Distribution of the Python programming language for large-scale data processing, predictive analytics, and scientific computing, that aims to simplify package management and deployment.
ipython notebook	Web-based interactive computational environment where you can combine code execution, text, mathematics, plots and rich media into a single document.

Notebook	Description
git	Distributed revision control system with an emphasis on speed, data integrity, and support for distributed, non-linear workflows.
ruby	Used to interact with the AWS command line and for Jekyll, a blog framework that can be hosted on GitHub Pages.
<u>jekyll</u>	Simple, blog-aware, static site generator for personal, project, or organization sites. Renders Markdown or Textile and Liquid templates, and produces a complete, static website ready to be served by Apache HTTP Server, Nginx or another web server.
pelican	Python-based alternative to Jekyll.
<u>django</u>	High-level Python Web framework that encourages rapid development and clean, pragmatic design. It can be useful to share reports/analyses and for blogging. Lighter-weight alternatives include Pyramid , Flask , Tornado , and Bottle .

misc

IPython Notebook(s) demonstrating miscellaneous functionality.

Notebook	Description
regex	Regular expression cheat sheet useful in data wrangling.
<u>algorithmia</u>	Algorithmia is a marketplace for algorithms. This notebook showcases 4 different algorithms: Face Detection, Content Summarizer, Latent Dirichlet Allocation and Optical Character Recognition.

notebook-installation

anaconda

Anaconda is a free distribution of the Python programming language for large-scale data processing, predictive analytics, and scientific computing that aims to simplify package management and deployment.

Follow instructions to install Anaconda or the more lightweight miniconda.

dev-setup

For detailed instructions, scripts, and tools to set up your development environment for data analysis, check out the <u>dev-setup</u> repo.

running-notebooks

To view interactive content or to modify elements within the IPython notebooks, you must first clone or download the repository then run the notebook. More information on IPython Notebooks can be found here.

```
$ git clone https://github.com/donnemartin/data-science-ipython-notebooks.git
$ cd data-science-ipython-notebooks
```

\$ jupyter notebook

Notebooks tested with Python 2.7.x.

credits

- Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython by Wes McKinney
- PyCon 2015 Scikit-learn Tutorial by Jake VanderPlas
- <u>Python Data Science Handbook</u> by Jake VanderPlas
- Parallel Machine Learning with scikit-learn and IPython by Olivier Grisel
- <u>Statistical Interference Using Computational Methods in Python</u> by Allen Downey
- <u>TensorFlow Examples</u> by Aymeric Damien
- <u>TensorFlow Tutorials</u> by Parag K Mital
- TensorFlow Tutorials by Nathan Lintz
- TensorFlow Tutorials by Alexander R Johansen
- <u>TensorFlow Book</u> by Nishant Shukla
- <u>Summer School 2015</u> by mila-udem
- Keras tutorials by Valerio Maggio
- Kaggle
- Yhat Blog

Awesome Machine Learning

https://github.com/josephmisiti/awesome-machine-learning

A curated list of awesome machine learning frameworks, libraries and software (by language). Inspired by awesome-php.

If you want to contribute to this list (please do), send me a pull request or contact me @josephmisiti Also, a listed repository should be deprecated if:

- Repository's owner explicitly say that "this library is not maintained".
- Not committed for long time (2~3 years).

Further resources:

- For a list of free machine learning books available for download, go here.
- For a list of (mostly) free machine learning courses available online, go here.
- For a list of blogs on data science and machine learning, go <u>here</u>.
- For a list of free-to-attend meetups and local events, go <u>here</u>.

Table of Contents

- APL
 - General-Purpose Machine Learning
- C
- o General-Purpose Machine Learning
- o Computer Vision
- C++
 - Computer Vision
 - General-Purpose Machine Learning
 - Natural Language Processing
 - Speech Recognition
 - Sequence Analysis
 - Gesture Recognition
- Common Lisp
 - General-Purpose Machine Learning
- Clojure
 - Natural Language Processing
 - o General-Purpose Machine Learning
 - Data Analysis / Data Visualization
- Crystal
 - General-Purpose Machine Learning

- Elixir
 - General-Purpose Machine Learning
 - Natural Language Processing
- Erlang
 - General-Purpose Machine Learning
- Go
 - o Natural Language Processing
 - o General-Purpose Machine Learning
 - Data Analysis / Data Visualization
- Haskell
 - General-Purpose Machine Learning
- Java
 - Natural Language Processing
 - General-Purpose Machine Learning
 - o Data Analysis / Data Visualization
 - Deep Learning
- Javascript
 - o Natural Language Processing
 - Data Analysis / Data Visualization
 - General-Purpose Machine Learning
 - o <u>Misc</u>
- Julia
 - General-Purpose Machine Learning
 - Natural Language Processing
 - Data Analysis / Data Visualization
 - Misc Stuff / Presentations
- <u>Lua</u>
 - General-Purpose Machine Learning
 - Demos and Scripts
- Matlab
 - Computer Vision
 - Natural Language Processing
 - General-Purpose Machine Learning
 - Data Analysis / Data Visualization
- .NET
 - Computer Vision
 - Natural Language Processing
 - General-Purpose Machine Learning
 - Data Analysis / Data Visualization
- Objective C
 - General-Purpose Machine Learning
- OCaml

- General-Purpose Machine Learning
 Perl
 Data Analysis / Data Visualization
- Perl 6
- PHP
 - Natural Language Processing
 - o General-Purpose Machine Learning

- Python
 - o Computer Vision
 - Natural Language Processing
 - General-Purpose Machine Learning
 - o <u>Data Analysis / Data Visualization</u>
 - Misc Scripts / iPython Notebooks / Codebases
 - Kaggle Competition Source Code
 - Neural Networks
 - o Reinforcement Learning
- Ruby
 - Natural Language Processing
 - o General-Purpose Machine Learning
 - Data Analysis / Data Visualization
 - Misc
- Rust
 - General-Purpose Machine Learning
- R
- General-Purpose Machine Learning
- Data Analysis / Data Visualization
- SAS
 - General-Purpose Machine Learning
 - o Data Analysis / Data Visualization
 - High Performance Machine Learning (MPP)
 - Natural Language Processing
 - Demos and Scripts
- Scala
 - Natural Language Processing
 - Data Analysis / Data Visualization
 - General-Purpose Machine Learning
- Swift
 - General-Purpose Machine Learning
- TensorFlow
 - General-Purpose Machine Learning
- Credits

APL

General-Purpose Machine Learning

<u>naive-apl</u> - Naive Bayesian Classifier implementation in APL

C

General-Purpose Machine Learning

- <u>Darknet</u> Darknet is an open source neural network framework written in C and CUDA. It is fast, easy to install, and supports CPU and GPU computation.
- <u>Recommender</u> A C library for product recommendations/suggestions using collaborative filtering (CF).
- <u>Hybrid Recommender System</u> A hybrid recommender system based upon scikit-learn algorithms.

Computer Vision

- <u>CCV</u> C-based/Cached/Core Computer Vision Library, A Modern Computer Vision Library
- <u>VLFeat</u> VLFeat is an open and portable library of computer vision algorithms, which has Matlab toolbox

Speech Recognition

• <u>HTK</u> -The Hidden Markov Model Toolkit (HTK) is a portable toolkit for building and manipulating hidden Markov models.

C++

Computer Vision

- <u>DLib</u> DLib has C++ and Python interfaces for face detection and training general object detectors.
- <u>EBLearn</u> Eblearn is an object-oriented C++ library that implements various machine learning models
- OpenCV OpenCV has C++, C, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS.
- <u>VIGRA</u> VIGRA is a generic cross-platform C++ computer vision and machine learning library for volumes of arbitrary dimensionality with Python bindings.

- BanditLib A simple Multi-armed Bandit library.
- <u>Caffe</u> A deep learning framework developed with cleanliness, readability, and speed in mind. [DEEP LEARNING]
- <u>CNTK</u> The Computational Network Toolkit (CNTK) by Microsoft Research, is a unified deep-learning toolkit that describes neural networks as a series of computational steps via a directed graph.
- <u>CUDA</u> This is a fast C++/CUDA implementation of convolutional [DEEP LEARNING]
- <u>CXXNET</u> Yet another deep learning framework with less than 1000 lines core code [DEEP LEARNING]
- <u>DeepDetect</u> A machine learning API and server written in C++11. It makes state of the art machine learning easy to work with and integrate into existing applications.
- <u>Distributed Machine learning Tool Kit (DMTK)</u> A distributed machine learning (parameter server) framework by Microsoft. Enables training models on large data sets across multiple machines. Current tools bundled with it include: LightLDA and Distributed (Multisense) Word Embedding.
- <u>DLib</u> A suite of ML tools designed to be easy to imbed in other applications
- <u>DSSTNE</u> A software library created by Amazon for training and deploying deep neural networks using GPUs which emphasizes speed and scale over experimental flexibility.
- <u>DyNet</u> A dynamic neural network library working well with networks that have dynamic structures that change for every training instance. Written in C++ with bindings in Python.
- encoq-cpp
- <u>Fido</u> A highly-modular C++ machine learning library for embedded electronics and robotics.
- <u>igraph</u> General purpose graph library
- Intel(R) DAAL A high performance software library developed by Intel and optimized for Intel's architectures. Library provides algorithmic building blocks for all stages of data analytics and allows to process data in batch, online and distributed modes.
- <u>LightGBM</u> Microsoft's fast, distributed, high performance gradient boosting (GBDT, GBRT, GBM or MART) framework based on decision tree algorithms, used for ranking, classification and many other machine learning tasks.
- <u>libfm</u> A generic approach that allows to mimic most factorization models by feature engineering.
- MLDB The Machine Learning Database is a database designed for machine learning.
 Send it commands over a RESTful API to store data, explore it using SQL, then train machine learning models and expose them as APIs.
- <u>mlpack</u> A scalable C++ machine learning library
- <u>proNet-core</u> A general-purpose network embedding framework: pair-wise representations optimization Network Edit
- ROOT A modular scientific software framework. It provides all the functionalities needed to deal with big data processing, statistical analysis, visualization and storage.

- <u>shark</u> A fast, modular, feature-rich open-source C++ machine learning library.
- Shogun The Shogun Machine Learning Toolbox
- sofia-ml Suite of fast incremental algorithms.
- <u>Stan</u> A probabilistic programming language implementing full Bayesian statistical inference with Hamiltonian Monte Carlo sampling
- <u>Timbl</u> A software package/C++ library implementing several memory-based learning algorithms, among which IB1-IG, an implementation of k-nearest neighbor classification, and IGTree, a decision-tree approximation of IB1-IG. Commonly used for NLP.
- Vowpal Wabbit (VW) A fast out-of-core learning system.
- <u>Warp-CTC</u> A fast parallel implementation of Connectionist Temporal Classification (CTC), on both CPU and GPU.
- XGBoost A parallelized optimized general purpose gradient boosting library.
- <u>LKYDeepNN</u> A header-only C++11 Neural Network library. Low dependency, native traditional chinese document.

Natural Language Processing

- <u>BLLIP Parser</u> BLLIP Natural Language Parser (also known as the Charniak-Johnson parser)
- <u>colibri-core</u> C++ library, command line tools, and Python binding for extracting and working with basic linguistic constructions such as n-grams and skipgrams in a quick and memory-efficient way.
- <u>CRF++</u> Open source implementation of Conditional Random Fields (CRFs) for segmenting/labeling sequential data & other Natural Language Processing tasks.
- <u>CRFsuite</u> CRFsuite is an implementation of Conditional Random Fields (CRFs) for labeling sequential data.
- <u>frog</u> Memory-based NLP suite developed for Dutch: PoS tagger, lemmatiser, dependency parser, NER, shallow parser, morphological analyzer.
- <u>libfolia</u> C++ library for the <u>FoLiA format</u>
- <u>MeTA</u> <u>MeTA</u> : <u>ModErn Text Analysis</u> is a C++ Data Sciences Toolkit that facilitates mining big text data.
- <u>MIT Information Extraction Toolkit</u> C, C++, and Python tools for named entity recognition and relation extraction
- <u>ucto</u> Unicode-aware regular-expression based tokenizer for various languages. Tool and C++ library. Supports FoLiA format.

Speech Recognition

• <u>Kaldi</u> - Kaldi is a toolkit for speech recognition written in C++ and licensed under the Apache License v2.0. Kaldi is intended for use by speech recognition researchers.

Sequence Analysis

• <u>ToPS</u> - This is an objected-oriented framework that facilitates the integration of probabilistic models for sequences over a user defined alphabet.

Gesture Detection

grt - The Gesture Recognition Toolkit (GRT) is a cross-platform, open-source, C++
machine learning library designed for real-time gesture recognition.

Common Lisp

General-Purpose Machine Learning

- mgl Neural networks (boltzmann machines, feed-forward and recurrent nets),
 Gaussian Processes
- <u>mgl-gpr</u> Evolutionary algorithms
- <u>cl-libsvm</u> Wrapper for the libsvm support vector machine library

Clojure

Natural Language Processing

- <u>Clojure-openNLP</u> Natural Language Processing in Clojure (opennlp)
- <u>Infections-clj</u> Rails-like inflection library for Clojure and ClojureScript

- <u>Touchstone</u> Clojure A/B testing library
- <u>Clojush</u> The Push programming language and the PushGP genetic programming system implemented in Clojure
- Infer Inference and machine learning in Clojure
- <u>Clj-ML</u> A machine learning library for Clojure built on top of Weka and friends
- <u>DL4CLJ</u> Clojure wrapper for Deeplearning4j
- Encog Clojure wrapper for Encog (v3) (Machine-Learning framework that specializes in neural-nets)
- Fungp A genetic programming library for Clojure
- <u>Statistiker</u> Basic Machine Learning algorithms in Clojure.
- <u>clortex</u> General Machine Learning library using Numenta's Cortical Learning Algorithm
- comportex Functionally composable Machine Learning library using Numenta's Cortical Learning Algorithm
- <u>cortex</u> Neural networks, regression and feature learning in Clojure.

• <u>lambda-ml</u> - Simple, concise implementations of machine learning techniques and utilities in Clojure.

Data Analysis / Data Visualization

- <u>Incanter</u> Incanter is a Clojure-based, R-like platform for statistical computing and graphics.
- PiqPen Map-Reduce for Clojure.
- Envision Clojure Data Visualisation library, based on Statistiker and D3

Crystal

General-Purpose Machine Learning

- machine Simple machine learning algorithm.
- <u>crystal-fann</u> FANN (Fast Artifical Neural Network) binding.

Elixir

General-Purpose Machine Learning

<u>Simple Bayes</u> - A Simple Bayes / Naive Bayes implementation in Elixir.

Natural Language Processing

• <u>Stemmer</u> - An English (Porter2) stemming implementation in Elixir.

Erlang

General-Purpose Machine Learning

- <u>Disco</u> Map Reduce in Erlang
- Yanni ANN neural networks using Erlangs leightweight processes

Go

Natural Language Processing

- go-porterstemmer A native Go clean room implementation of the Porter Stemming algorithm.
- <u>paicehusk</u> Golang implementation of the Paice/Husk Stemming Algorithm.
- snowball Snowball Stemmer for Go.

- <u>qo-ngram</u> In-memory n-gram index with compression.
- word-embedding Word Embeddings: the full implementation of word2vec, GloVe in Go.

General-Purpose Machine Learning

- gago Multi-population, flexible, parallel genetic algorithm.
- Go Learn Machine Learning for Go
- go-pr Pattern recognition package in Go lang.
- <u>go-ml</u> Linear / Logistic regression, Neural Networks, Collaborative Filtering and Gaussian Multivariate Distribution
- <u>bayesian</u> Naive Bayesian Classification for Golang.
- go-galib Genetic Algorithms library written in Go / Golang
- <u>Cloudforest</u> Ensembles of decision trees in go/Golang.
- gobrain Neural Networks written in go
- GoNN GoNN is an implementation of Neural Network in Go Language, which includes BPNN, RBF, PCN
- <u>MXNet</u> Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Go, Javascript and more.
- go-mxnet-predictor Go binding for MXNet c_predict_api to do inference with pretrained model
- <u>neat</u> Plug-and-play, parallel Go framework for NeuroEvolution of Augmenting Topologies (NEAT)

Data Analysis / Data Visualization

- go-graph Graph library for Go/Golang language.
- <u>SVGo</u> The Go Language library for SVG generation
- <u>RF</u> Random forests implementation in Go

Haskell

- <u>haskell-ml</u> Haskell implementations of various ML algorithms.
- <u>HLearn</u> a suite of libraries for interpreting machine learning models according to their algebraic structure.
- <u>hnn</u> Haskell Neural Network library.
- <u>hopfield-networks</u> Hopfield Networks for unsupervised learning in Haskell.
- <u>caffegraph</u> A DSL for deep neural networks
- <u>LambdaNet</u> Configurable Neural Networks in Haskell

Java

Natural Language Processing

- <u>Cortical.io</u> Retina: an API performing complex NLP operations (disambiguation, classification, streaming text filtering, etc...) as quickly and intuitively as the brain.
- IRIS Cortical.io's FREE NLP, Retina API Analysis Tool (written in JavaFX!) See the Tutorial Video
- <u>CoreNLP</u> Stanford CoreNLP provides a set of natural language analysis tools which can take raw English language text input and give the base forms of words
- <u>Stanford Parser</u> A natural language parser is a program that works out the grammatical structure of sentences
- <u>Stanford POS Tagger</u> A Part-Of-Speech Tagger (POS Tagger
- <u>Stanford Name Entity Recognizer</u> Stanford NER is a Java implementation of a Named Entity Recognizer.
- <u>Stanford Word Segmenter</u> Tokenization of raw text is a standard pre-processing step for many NLP tasks.
- <u>Tregex, Tsurgeon and Semgrex</u> Tregex is a utility for matching patterns in trees, based on tree relationships and regular expression matches on nodes (the name is short for "tree regular expressions").
- <u>Stanford Phrasal: A Phrase-Based Translation System</u>
- <u>Stanford English Tokenizer</u> Stanford Phrasal is a state-of-the-art statistical phrase-based machine translation system, written in Java.
- <u>Stanford Tokens Regex</u> A tokenizer divides text into a sequence of tokens, which roughly correspond to "words"
- <u>Stanford Temporal Tagger</u> SUTime is a library for recognizing and normalizing time expressions.
- <u>Stanford SPIED</u> Learning entities from unlabeled text starting with seed sets using patterns in an iterative fashion
- <u>Stanford Topic Modeling Toolbox</u> Topic modeling tools to social scientists and others who wish to perform analysis on datasets
- <u>Twitter Text Java</u> A Java implementation of Twitter's text processing library
- MALLET A Java-based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text.
- OpenNLP a machine learning based toolkit for the processing of natural language text.
- <u>LingPipe</u> A tool kit for processing text using computational linguistics.
- <u>ClearTK</u> ClearTK provides a framework for developing statistical natural language processing (NLP) components in Java and is built on top of Apache UIMA.
- Apache cTAKES Apache clinical Text Analysis and Knowledge Extraction System (cTAKES) is an open-source natural language processing system for information extraction from electronic medical record clinical free-text.

- <u>ClearNLP</u> The ClearNLP project provides software and resources for natural language processing. The project started at the Center for Computational Language and EducAtion Research, and is currently developed by the Center for Language and Information Research at Emory University. This project is under the Apache 2 license.
- <u>CogcompNLP</u> This project collects a number of core libraries for Natural Language
 Processing (NLP) developed in the University of Illinois' Cognitive Computation Group,
 for example illinois-core-utilities which provides a set of NLP-friendly data
 structures and a number of NLP-related utilities that support writing NLP applications,
 running experiments, etc, illinois-edison a library for feature extraction from illinoiscore-utilities data structures and many other packages.

- <u>aerosolve</u> A machine learning library by Airbnb designed from the ground up to be human friendly.
- AMIDST Toolbox A Java Toolbox for Scalable Probabilistic Machine Learning.
- <u>Datumbox</u> Machine Learning framework for rapid development of Machine Learning and Statistical applications
- <u>ELKI</u> Java toolkit for data mining. (unsupervised: clustering, outlier detection etc.)
- Encog An advanced neural network and machine learning framework. Encog contains classes to create a wide variety of networks, as well as support classes to normalize and process data for these neural networks. Encog trains using multithreaded resilient propagation. Encog can also make use of a GPU to further speed processing time. A GUI based workbench is also provided to help model and train neural networks.
- FlinkML in Apache Flink Distributed machine learning library in Flink
- <u>H2O</u> ML engine that supports distributed learning on Hadoop, Spark or your laptop via APIs in R, Python, Scala, REST/JSON.
- htm.java General Machine Learning library using Numenta's Cortical Learning Algorithm
- java-deeplearning Distributed Deep Learning Platform for Java, Clojure, Scala
- Mahout Distributed machine learning
- Meka An open source implementation of methods for multi-label classification and evaluation (extension to Weka).
- MLlib in Apache Spark Distributed machine learning library in Spark
- <u>Hydrosphere Mist</u> a service for deployment Apache Spark MLLib machine learning models as realtime, batch or reactive web services.
- Neuroph Neuroph is lightweight Java neural network framework
- ORYX Lambda Architecture Framework using Apache Spark and Apache Kafka with a specialization for real-time large-scale machine learning.
- <u>Samoa</u> SAMOA is a framework that includes distributed machine learning for data streams with an interface to plug-in different stream processing platforms.
- RankLib RankLib is a library of learning to rank algorithms
- rapaio statistics, data mining and machine learning toolbox in Java
- RapidMiner RapidMiner integration into Java code

- <u>Stanford Classifier</u> A classifier is a machine learning tool that will take data items and place them into one of k classes.
- <u>SmileMiner</u> Statistical Machine Intelligence & Learning Engine
- <u>SystemML</u> flexible, scalable machine learning (ML) language.
- WalnutiQ object oriented model of the human brain
- Weka Weka is a collection of machine learning algorithms for data mining tasks
- <u>LBJava</u> Learning Based Java is a modeling language for the rapid development of software systems, offers a convenient, declarative syntax for classifier and constraint definition directly in terms of the objects in the programmer's application.

Speech Recognition

 <u>CMU Sphinx</u> - Open Source Toolkit For Speech Recognition purely based on Java speech recognition library.

Data Analysis / Data Visualization

- Flink Open source platform for distributed stream and batch data processing.
- Hadoop Hadoop/HDFS
- Onyx Distributed, masterless, high performance, fault tolerant data processing.
 Written entirely in Clojure.
- Spark Spark is a fast and general engine for large-scale data processing.
- Storm Storm is a distributed realtime computation system.
- Impala Real-time Query for Hadoop
- <u>DataMelt</u> Mathematics software for numeric computation, statistics, symbolic calculations, data analysis and data visualization.
- Dr. Michael Thomas Flanagan's Java Scientific Library

Deep Learning

<u>Deeplearning4j</u> - Scalable deep learning for industry with parallel GPUs

Javascript

Natural Language Processing

- Twitter-text A JavaScript implementation of Twitter's text processing library
- <u>natural</u> General natural language facilities for node
- Knwl.js A Natural Language Processor in JS
- <u>Retext</u> Extensible system for analyzing and manipulating natural language
- <u>TextProcessing</u> Sentiment analysis, stemming and lemmatization, part-of-speech tagging and chunking, phrase extraction and named entity recognition.
- NLP Compromise Natural Language processing in the browser

Data Analysis / Data Visualization

- D3.is
- High Charts
- NVD3.js
- dc.js
- chartis
- dimple
- amCharts
- D3xter Straight forward plotting built on D3
- <u>statkit</u> Statistics kit for JavaScript
- <u>datakit</u> A lightweight framework for data analysis in JavaScript
- <u>science.js</u> Scientific and statistical computing in JavaScript.
- Z3d Easily make interactive 3d plots built on Three.js
- <u>Sigma.js</u> JavaScript library dedicated to graph drawing.
- <u>C3.js</u>- customizable library based on D3.js for easy chart drawing.
- Datamaps- Customizable SVG map/geo visualizations using D3.js.
- ZingChart- library written on Vanilla JS for big data visualization.
- <u>cheminfo</u> Platform for data visualization and analysis, using the <u>visualizer</u> project.
- Learn JS Data
- AnyChart
- FusionCharts

- <u>Convnet.js</u> ConvNetJS is a Javascript library for training Deep Learning models[DEEP LEARNING]
- <u>Clusterfck</u> Agglomerative hierarchical clustering implemented in Javascript for Node.js and the browser
- <u>Clustering.js</u> Clustering algorithms implemented in Javascript for Node.js and the browser
- <u>Decision Trees</u> NodeJS Implementation of Decision Tree using ID3 Algorithm
- DN2A Digital Neural Networks Architecture
- figue K-means, fuzzy c-means and agglomerative clustering
- <u>Gaussian Mixture Model</u> Unsupervised machine learning with multivariate Gaussian mixture model
- <u>Node-fann</u> FANN (Fast Artificial Neural Network Library) bindings for Node.js
- <u>Kmeans.js</u> Simple Javascript implementation of the k-means algorithm, for node.js and the browser
- LDA.js LDA topic modeling for Node.js
- Learning.js Javascript implementation of logistic regression/c4.5 decision tree
- Machine Learning Machine learning library for Node.js

- <u>machineJS</u> Automated machine learning, data formatting, ensembling, and hyperparameter optimization for competitions and exploration- just give it a .csv file!
- mil-tokyo List of several machine learning libraries
- Node-SVM Support Vector Machine for Node.js
- <u>Brain</u> Neural networks in JavaScript [**Deprecated**]
- Brain.js Neural networks in JavaScript continued community fork of Brain
- <u>Bayesian-Bandit</u> Bayesian bandit implementation for Node and the browser.
- Synaptic Architecture-free neural network library for Node.js and the browser
- <u>kNear</u> JavaScript implementation of the k nearest neighbors algorithm for supervised learning
- <u>NeuralN</u> C++ Neural Network library for Node.js. It has advantage on large dataset and multi-threaded training.
- <u>kalman</u> Kalman filter for Javascript.
- <u>shaman</u> Node.js library with support for both simple and multiple linear regression.
- <u>ml.js</u> Machine learning and numerical analysis tools for Node.js and the Browser!
- Pavlov.js Reinforcement learning using Markov Decision Processes
- <u>MXNet</u> Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Go, Javascript and more.
- <u>deeplearnis</u> A hardware-accelerated machine intelligence library for the web

Misc

- <u>sylvester</u> Vector and Matrix math for JavaScript.
- <u>simple-statistics</u> A JavaScript implementation of descriptive, regression, and inference statistics. Implemented in literate JavaScript with no dependencies, designed to work in all modern browsers (including IE) as well as in Node.js.
- <u>regression-js</u> A javascript library containing a collection of least squares fitting methods for finding a trend in a set of data.
- <u>Lyric</u> Linear Regression library.
- <u>GreatCircle</u> Library for calculating great circle distance.

Julia

- MachineLearning Julia Machine Learning library
- MLBase A set of functions to support the development of machine learning algorithms
- <u>PGM</u> A Julia framework for probabilistic graphical models.
- <u>DA</u> Julia package for Regularized Discriminant Analysis

- <u>Regression</u> Algorithms for regression analysis (e.g. linear regression and logistic regression)
- Local Regression Local regression, so smooooth!
- Naive Bayes Simple Naive Bayes implementation in Julia
- Mixed Models A Julia package for fitting (statistical) mixed-effects models
- <u>Simple MCMC</u> basic mcmc sampler implemented in Julia
- Distance Julia module for Distance evaluation
- <u>Decision Tree</u> Decision Tree Classifier and Regressor
- Neural A neural network in Julia
- MCMC MCMC tools for Julia
- Mamba Markov chain Monte Carlo (MCMC) for Bayesian analysis in Julia
- GLM Generalized linear models in Julia
- Gaussian Processes Julia package for Gaussian processes
- Online Learning
- GLMNet Julia wrapper for fitting Lasso/ElasticNet GLM models using glmnet
- <u>Clustering</u> Basic functions for clustering data: k-means, dp-means, etc.
- SVM SVM's for Julia
- Kernel Density Kernel density estimators for julia
- <u>Dimensionality Reduction</u> Methods for dimensionality reduction
- NMF A Julia package for non-negative matrix factorization
- ANN Julia artificial neural networks
- Mocha Deep Learning framework for Julia inspired by Caffe
- XGBoost eXtreme Gradient Boosting Package in Julia
- <u>ManifoldLearning</u> A Julia package for manifold learning and nonlinear dimensionality reduction
- MXNet Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Go, Javascript and more.
- Merlin Flexible Deep Learning Framework in Julia
- <u>ROCAnalysis</u> Receiver Operating Characteristics and functions for evaluation probabilistic binary classifiers
- GaussianMixtures Large scale Gaussian Mixture Models
- <u>ScikitLearn</u> Julia implementation of the scikit-learn API
- Knet Koç University Deep Learning Framework

Natural Language Processing

- Topic Models TopicModels for Julia
- Text Analysis Julia package for text analysis

Data Analysis / Data Visualization

- Graph Layout Graph layout algorithms in pure Julia
- <u>LightGraphs</u> Graph modeling and analysis
- <u>Data Frames Meta</u> Metaprogramming tools for DataFrames
- <u>Julia Data</u> library for working with tabular data in Julia
- Data Read Read files from Stata, SAS, and SPSS
- <u>Hypothesis Tests</u> Hypothesis tests for Julia
- Gadfly Crafty statistical graphics for Julia.
- Stats Statistical tests for Julia
- RDataSets Julia package for loading many of the data sets available in R
- <u>DataFrames</u> library for working with tabular data in Julia
- <u>Distributions</u> A Julia package for probability distributions and associated functions.
- <u>Data Arrays</u> Data structures that allow missing values
- Time Series Time series toolkit for Julia
- <u>Sampling</u> Basic sampling algorithms for Julia

Misc Stuff / Presentations

- <u>DSP</u> Digital Signal Processing (filtering, periodograms, spectrograms, window functions).
- JuliaCon Presentations Presentations for JuliaCon
- <u>SignalProcessing</u> Signal Processing tools for Julia
- Images An image library for Julia

Lua

General-Purpose Machine Learning

Torch7

- cephes Cephes mathematical functions library, wrapped for Torch. Provides and wraps the 180+ special mathematical functions from the Cephes mathematical library, developed by Stephen L. Moshier. It is used, among many other places, at the heart of SciPy.
- <u>autograd</u> Autograd automatically differentiates native Torch code. Inspired by the original Python version.
- o graph Graph package for Torch
- o randomkit Numpy's randomkit, wrapped for Torch
- signal A signal processing toolbox for Torch-7. FFT, DCT, Hilbert, cepstrums, stft
- <u>nn</u> Neural Network package for Torch

- torchnet framework for torch which provides a set of abstractions aiming at encouraging code re-use as well as encouraging modular programming
- nngraph This package provides graphical computation for nn library in Torch7.
- nnx A completely unstable and experimental package that extends Torch's builtin nn library
- <u>rnn</u> A Recurrent Neural Network library that extends Torch's nn. RNNs, LSTMs, GRUs, BRNNs, BLSTMs, etc.
- o dpnn Many useful features that aren't part of the main nn package.
- dp A deep learning library designed for streamlining research and development using the Torch7 distribution. It emphasizes flexibility through the elegant use of object-oriented design patterns.
- optim An optimization library for Torch. SGD, Adagrad, Conjugate-Gradient, LBFGS, RProp and more.
- unsup A package for unsupervised learning in Torch. Provides modules that are compatible with nn (LinearPsd, ConvPsd, AutoEncoder, ...), and selfcontained algorithms (k-means, PCA).
- manifold A package to manipulate manifolds
- o svm Torch-SVM library
- o <u>lbfgs</u> FFI Wrapper for liblbfgs
- o <u>vowpalwabbit</u> An old vowpalwabbit interface to torch.
- OpenGM OpenGM is a C++ library for graphical modeling, and inference. The Lua bindings provide a simple way of describing graphs, from Lua, and then optimizing them with OpenGM.
- o sphagetti Spaghetti (sparse linear) module for torch7 by @MichaelMathieu
- <u>LuaSHKit</u> A lua wrapper around the Locality sensitive hashing library SHKit
- kernel smoothing KNN, kernel-weighted average, local linear regression smoothers
- <u>cutorch</u> Torch CUDA Implementation
- <u>cunn</u> Torch CUDA Neural Network Implementation
- imgraph An image/graph library for Torch. This package provides routines to construct graphs on images, segment them, build trees out of them, and convert them back to images.
- videograph A video/graph library for Torch. This package provides routines to construct graphs on videos, segment them, build trees out of them, and convert them back to videos.
- <u>saliency</u> code and tools around integral images. A library for finding interest points based on fast integral histograms.
- <u>stitch</u> allows us to use hugin to stitch images and apply same stitching to a video sequence
- o sfm A bundle adjustment/structure from motion package
- <u>fex</u> A package for feature extraction in Torch. Provides SIFT and dSIFT modules.

- o OverFeat A state-of-the-art generic dense feature extractor
- Numeric Lua
- <u>Lunatic Python</u>
- SciLua
- Lua Numerical Algorithms
- Lunum

Demos and Scripts

- Core torch7 demos repository.
 - o linear-regression, logistic-regression
 - face detector (training and detection as separate demos)
 - o mst-based-segmenter
 - train-a-digit-classifier
 - train-autoencoder
 - o optical flow demo
 - o train-on-housenumbers
 - o train-on-cifar
 - tracking with deep nets
 - kinect demo
 - filter-bank visualization
 - saliency-networks
- Training a Convnet for the Galaxy-Zoo Kaggle challenge(CUDA demo)
- Music Tagging Music Tagging scripts for torch7
- <u>torch-datasets</u> Scripts to load several popular datasets including:
 - o BSR 500
 - o CIFAR-10
 - o COIL
 - Street View House Numbers
 - MNIST
 - NORB
- Atari2600 Scripts to generate a dataset with static frames from the Arcade Learning Environment

Matlab

Computer Vision

- <u>Contourlets</u> MATLAB source code that implements the contourlet transform and its utility functions.
- Shearlets MATLAB code for shearlet transform

- <u>Curvelets</u> The Curvelet transform is a higher dimensional generalization of the Wavelet transform designed to represent images at different scales and different angles.
- Bandlets MATLAB code for bandlet transform
- mexopencv Collection and a development kit of MATLAB mex functions for OpenCV library

Natural Language Processing

NLP - An NLP library for Matlab

General-Purpose Machine Learning

- <u>Training a deep autoencoder or a classifier on MNIST digits</u> Training a deep autoencoder or a classifier on MNIST digits[DEEP LEARNING]
- <u>Convolutional-Recursive Deep Learning for 3D Object Classification</u> Convolutional-Recursive Deep Learning for 3D Object Classification[DEEP LEARNING]
- <u>t-Distributed Stochastic Neighbor Embedding</u> t-Distributed Stochastic Neighbor Embedding (t-SNE) is a (prize-winning) technique for dimensionality reduction that is particularly well suited for the visualization of high-dimensional datasets.
- <u>Spider</u> The spider is intended to be a complete object orientated environment for machine learning in Matlab.
- <u>LibSVM</u> A Library for Support Vector Machines
- <u>LibLinear</u> A Library for Large Linear Classification
- Machine Learning Module Class on machine w/ PDF, lectures, code
- <u>Caffe</u> A deep learning framework developed with cleanliness, readability, and speed in mind.
- <u>Pattern Recognition Toolbox</u> A complete object-oriented environment for machine learning in Matlab.
- <u>Pattern Recognition and Machine Learning</u> This package contains the matlab implementation of the algorithms described in the book Pattern Recognition and Machine Learning by C. Bishop.
- Optunity A library dedicated to automated hyperparameter optimization with a simple, lightweight API to facilitate drop-in replacement of grid search. Optunity is written in Python but interfaces seamlessly with MATLAB.

Data Analysis / Data Visualization

- <u>matlab_gbl</u> MatlabBGL is a Matlab package for working with graphs.
- gamic Efficient pure-Matlab implementations of graph algorithms to complement MatlabBGL's mex functions.

.NET

Computer Vision

- OpenCVDotNet A wrapper for the OpenCV project to be used with .NET applications.
- <u>Emgu CV</u> Cross platform wrapper of OpenCV which can be compiled in Mono to e run on Windows, Linus, Mac OS X, iOS, and Android.
- <u>AForge.NET</u> Open source C# framework for developers and researchers in the fields of Computer Vision and Artificial Intelligence. Development has now shifted to GitHub.
- Accord.NET Together with AForge.NET, this library can provide image processing and computer vision algorithms to Windows, Windows RT and Windows Phone. Some components are also available for Java and Android.

Natural Language Processing

 <u>Stanford.NLP for .NET</u> - A full port of Stanford NLP packages to .NET and also available precompiled as a NuGet package.

- <u>Accord-Framework</u> -The Accord.NET Framework is a complete framework for building machine learning, computer vision, computer audition, signal processing and statistical applications.
- <u>Accord.MachineLearning</u> Support Vector Machines, Decision Trees, Naive Bayesian models, K-means, Gaussian Mixture models and general algorithms such as Ransac, Cross-validation and Grid-Search for machine-learning applications. This package is part of the Accord.NET Framework.
- <u>DiffSharp</u> An automatic differentiation (AD) library providing exact and efficient derivatives (gradients, Hessians, Jacobians, directional derivatives, and matrix-free Hessian- and Jacobian-vector products) for machine learning and optimization applications. Operations can be nested to any level, meaning that you can compute exact higher-order derivatives and differentiate functions that are internally making use of differentiation, for applications such as hyperparameter optimization.
- <u>Vulpes</u> Deep belief and deep learning implementation written in F# and leverages CUDA GPU execution with Alea.cuBase.
- Encog An advanced neural network and machine learning framework. Encog contains
 classes to create a wide variety of networks, as well as support classes to normalize
 and process data for these neural networks. Encog trains using multithreaded resilient
 propagation. Encog can also make use of a GPU to further speed processing time. A
 GUI based workbench is also provided to help model and train neural networks.
- <u>Neural Network Designer</u> DBMS management system and designer for neural networks. The designer application is developed using WPF, and is a user interface which allows you to design your neural network, query the network, create and configure chat bots that are capable of asking questions and learning from your feed back. The chat bots can even scrape the internet for information to return in their output as well as to use for learning.
- <u>Infer.NET</u> Infer.NET is a framework for running Bayesian inference in graphical models. One can use Infer.NET to solve many different kinds of machine learning

problems, from standard problems like classification, recommendation or clustering through to customised solutions to domain-specific problems. Infer.NET has been used in a wide variety of domains including information retrieval, bioinformatics, epidemiology, vision, and many others.

Data Analysis / Data Visualization

- <u>numl</u> numl is a machine learning library intended to ease the use of using standard modeling techniques for both prediction and clustering.
- Math.NET Numerics Numerical foundation of the Math.NET project, aiming to provide methods and algorithms for numerical computations in science, engineering and every day use. Supports .Net 4.0, .Net 3.5 and Mono on Windows, Linux and Mac; Silverlight 5, WindowsPhone/SL 8, WindowsPhone 8.1 and Windows 8 with PCL Portable Profiles 47 and 344; Android/iOS with Xamarin.
- Sho Sho is an interactive environment for data analysis and scientific computing that lets you seamlessly connect scripts (in IronPython) with compiled code (in .NET) to enable fast and flexible prototyping. The environment includes powerful and efficient libraries for linear algebra as well as data visualization that can be used from any .NET language, as well as a feature-rich interactive shell for rapid development.

Objective C

- YCML A Machine Learning framework for Objective-C and Swift (OS X / iOS).
- MLPNeuralNet Fast multilayer perceptron neural network library for iOS and Mac OS
 X. MLPNeuralNet predicts new examples by trained neural network. It is built on top of
 the Apple's Accelerate Framework, using vectorized operations and hardware
 acceleration if available.
- MAChineLearning An Objective-C multilayer perceptron library, with full support for training through backpropagation. Implemented using vDSP and vecLib, it's 20 times faster than its Java equivalent. Includes sample code for use from Swift.
- <u>BPN-NeuralNetwork</u> It implemented 3 layers neural network (Input Layer, Hidden Layer and Output Layer) and it named Back Propagation Neural Network (BPN). This network can be used in products recommendation, user behavior analysis, data mining and data analysis.
- <u>Multi-Perceptron-NeuralNetwork</u> it implemented multi-perceptrons neural network (ニューラルネットワーク) based on Back Propagation Neural Network (BPN) and designed unlimited-hidden-layers.
- <u>KRHebbian-Algorithm</u> It is a non-supervisor and self-learning algorithm (adjust the weights) in neural network of Machine Learning.
- <u>KRKmeans-Algorithm</u> It implemented K-Means the clustering and classification algorithm. It could be used in data mining and image compression.

 KRFuzzyCMeans-Algorithm - It implemented Fuzzy C-Means (FCM) the fuzzy clustering / classification algorithm on Machine Learning. It could be used in data mining and image compression.

OCaml

General-Purpose Machine Learning

- Oml A general statistics and machine learning library.
- GPR Efficient Gaussian Process Regression in OCaml.
- <u>Libra-Tk</u> Algorithms for learning and inference with discrete probabilistic models.
- <u>TensorFlow</u> OCaml bindings for TensorFlow.

Perl

Data Analysis / Data Visualization

• <u>Perl Data Language</u>, a pluggable architecture for data and image processing, which can be used for machine learning.

General-Purpose Machine Learning

- MXnet for Deep Learning, in Perl, also released in CPAN.
- Paws::MachineLearning, using AWS machine learning platform from Perl.
- <u>Algorithm::SVMLight</u>, implementation of Support Vector Machines with SVMLight under it.
- Several machine learning and artificial intelligence models are included in the AI namespace. For instance, you can find Naïve Bayes.

Perl 6

- Support Vector Machines
- Naïve Bayes

Data Analysis / Data Visualization

• <u>Perl Data Language</u>, a pluggable architecture for data and image processing, which can be <u>used for machine learning</u>.

General-Purpose Machine Learning

PHP

Natural Language Processing

• <u>jieba-php</u> - Chinese Words Segmentation Utilities.

General-Purpose Machine Learning

- <u>PHP-ML</u> Machine Learning library for PHP. Algorithms, Cross Validation, Neural Network, Preprocessing, Feature Extraction and much more in one library.
- <u>PredictionBuilder</u> A library for machine learning that builds predictions using a linear regression.

Python

Computer Vision

- <u>Scikit-Image</u> A collection of algorithms for image processing in Python.
- <u>SimpleCV</u> An open source computer vision framework that gives access to several high-powered computer vision libraries, such as OpenCV. Written on Python and runs on Mac, Windows, and Ubuntu Linux.
- <u>Vigranumpy</u> Python bindings for the VIGRA C++ computer vision library.
- OpenFace Free and open source face recognition with deep neural networks.
- PCV Open source Python module for computer vision
- <u>face recognition</u> Face recognition library that recognize and manipulate faces from Python or from the command line
- <u>dockerface</u> Easy to install and use deep learning Faster R-CNN face detection for images and video in a docker container.

Natural Language Processing

- NLTK A leading platform for building Python programs to work with human language data.
- <u>Pattern</u> A web mining module for the Python programming language. It has tools for natural language processing, machine learning, among others.
- Quepy A python framework to transform natural language questions to queries in a database query language
- <u>TextBlob</u> Providing a consistent API for diving into common natural language processing (NLP) tasks. Stands on the giant shoulders of NLTK and Pattern, and plays nicely with both.

- <u>YAlign</u> A sentence aligner, a friendly tool for extracting parallel sentences from comparable corpora.
- <u>jieba</u> Chinese Words Segmentation Utilities.
- SnowNLP A library for processing Chinese text.
- spammy A library for email Spam filtering built on top of nltk
- <u>loso</u> Another Chinese segmentation library.
- genius A Chinese segment base on Conditional Random Field.
- <u>KoNLPy</u> A Python package for Korean natural language processing.
- <u>nut</u> Natural language Understanding Toolkit
- Rosetta Text processing tools and wrappers (e.g. Vowpal Wabbit)
- <u>BLLIP Parser</u> Python bindings for the BLLIP Natural Language Parser (also known as the Charniak-Johnson parser)
- <u>PyNLPI</u> Python Natural Language Processing Library. General purpose NLP library for Python. Also contains some specific modules for parsing common NLP formats, most notably for <u>FoLiA</u>, but also ARPA language models, Moses phrasetables, GIZA++ alignments.
- <u>python-ucto</u> Python binding to ucto (a unicode-aware rule-based tokenizer for various languages)
- <u>python-frog</u> Python binding to Frog, an NLP suite for Dutch. (pos tagging, lemmatisation, dependency parsing, NER)
- <u>python-zpar</u> Python bindings for <u>ZPar</u>, a statistical part-of-speech-tagger, constiuency parser, and dependency parser for English.
- <u>colibri-core</u> Python binding to C++ library for extracting and working with with basic linguistic constructions such as n-grams and skipgrams in a quick and memory-efficient way.
- spaCy Industrial strength NLP with Python and Cython.
- <u>PyStanfordDependencies</u> Python interface for converting Penn Treebank trees to Stanford Dependencies.
- <u>Distance</u> Levenshtein and Hamming distance computation
- Fuzzy Wuzzy Fuzzy String Matching in Python
- jellyfish a python library for doing approximate and phonetic matching of strings.
- <u>editdistance</u> fast implementation of edit distance
- <u>textacy</u> higher-level NLP built on Spacy
- <u>stanford-corenlp-python</u> Python wrapper for <u>Stanford CoreNLP</u>
- <u>CLTK</u> The Classical Language Toolkit
- <u>rasa nlu</u> turn natural language into structured data
- <u>yase</u> Transcode sentence (or other sequence) to list of word vector
- Polyglot Multilingual text (NLP) processing toolkit
- <u>DrQA</u> Reading Wikipedia to answer open-domain questions

- <u>auto_ml</u> Automated machine learning for production and analytics. Lets you focus on the fun parts of ML, while outputting production-ready code, and detailed analytics of your dataset and results. Includes support for NLP, XGBoost, LightGBM, and soon, deep learning.
- <u>machine learning</u> automated build consisting of a <u>web-interface</u>, and set
 of <u>programmatic-interface</u> API, for support vector machines. Corresponding dataset(s)
 are stored into a SQL database, then generated model(s) used for prediction(s), are
 stored into a NoSQL datastore.
- XGBoost Python bindings for eXtreme Gradient Boosting (Tree) Library
- <u>Bayesian Methods for Hackers</u> Book/iPython notebooks on Probabilistic Programming in Python
- <u>Featureforge</u> A set of tools for creating and testing machine learning features, with a scikit-learn compatible API
- MLlib in Apache Spark Distributed machine learning library in Spark
- <u>Hydrosphere Mist</u> a service for deployment Apache Spark MLLib machine learning models as realtime, batch or reactive web services.
- scikit-learn A Python module for machine learning built on top of SciPy.
- <u>metric-learn</u> A Python module for metric learning.
- <u>SimpleAl</u> Python implementation of many of the artificial intelligence algorithms described on the book "Artificial Intelligence, a Modern Approach". It focuses on providing an easy to use, well documented and tested library.
- <u>astroML</u> Machine Learning and Data Mining for Astronomy.
- <u>graphlab-create</u> A library with various machine learning models (regression, clustering, recommender systems, graph analytics, etc.) implemented on top of a disk-backed DataFrame.
- <u>BigML</u> A library that contacts external servers.
- pattern Web mining module for Python.
- NuPIC Numenta Platform for Intelligent Computing.
- <u>Pylearn2</u> A Machine Learning library based on <u>Theano</u>.
- <u>keras</u> Modular neural network library based on <u>Theano</u>.
- <u>Lasagne</u> Lightweight library to build and train neural networks in Theano.
- <u>hebel</u> GPU-Accelerated Deep Learning Library in Python.
- Chainer Flexible neural network framework
- prophet Fast and automated time series forecasting framework by Facebook.
- <u>gensim</u> Topic Modelling for Humans.
- <u>topik</u> Topic modelling toolkit
- <u>PyBrain</u> Another Python Machine Learning Library.
- <u>Brainstorm</u> Fast, flexible and fun neural networks. This is the successor of PyBrain.
- <u>Crab</u> A flexible, fast recommender engine.
- python-recsys A Python library for implementing a Recommender System.

- thinking bayes Book on Bayesian Analysis
- <u>Image-to-Image Translation with Conditional Adversarial Networks</u> Implementation of image to image (pix2pix) translation from the paper by <u>isola et al.</u>[DEEP LEARNING]
- Restricted Boltzmann Machines -Restricted Boltzmann Machines in Python. [DEEP LEARNING]
- <u>Bolt</u> Bolt Online Learning Toolbox
- <u>CoverTree</u> Python implementation of cover trees, near-drop-in replacement for scipy.spatial.kdtree
- <u>nilearn</u> Machine learning for NeuroImaging in Python
- neuropredict Aimed at novice machine learners and non-expert programmers, this
 package offers easy (no coding needed) and comprehensive machine learning
 (evaluation and full report of predictive performance WITHOUT requiring you to code)
 in Python for NeuroImaging and any other type of features. This is aimed at absorbing
 the much of the ML workflow, unlike other packages like nilearn and pymvpa, which
 require you to learn their API and code to produce anything useful.
- <u>imbalanced-learn</u> Python module to perform under sampling and over sampling with various techniques.
- Shogun The Shogun Machine Learning Toolbox
- Pyevolve Genetic algorithm framework.
- <u>Caffe</u> A deep learning framework developed with cleanliness, readability, and speed in mind.
- <u>breze</u> Theano based library for deep and recurrent neural networks
- <u>pyhsmm</u> library for approximate unsupervised inference in Bayesian Hidden Markov Models (HMMs) and explicit-duration Hidden semi-Markov Models (HSMMs), focusing on the Bayesian Nonparametric extensions, the HDP-HMM and HDP-HSMM, mostly with weak-limit approximations.
- <u>mrjob</u> A library to let Python program run on Hadoop.
- <u>SKLL</u> A wrapper around scikit-learn that makes it simpler to conduct experiments.
- <u>neurolab</u> <u>https://github.com/zueve/neurolab</u>
- Spearmint Spearmint is a package to perform Bayesian optimization according to the algorithms outlined in the paper: Practical Bayesian Optimization of Machine Learning Algorithms. Jasper Snoek, Hugo Larochelle and Ryan P. Adams. Advances in Neural Information Processing Systems, 2012.
- <u>Pebl</u> Python Environment for Bayesian Learning
- <u>Theano</u> Optimizing GPU-meta-programming code generating array oriented optimizing math compiler in Python
- <u>TensorFlow</u> Open source software library for numerical computation using data flow graphs
- <u>yahmm</u> Hidden Markov Models for Python, implemented in Cython for speed and efficiency.
- <u>python-timbl</u> A Python extension module wrapping the full TiMBL C++ programming interface. Timbl is an elaborate k-Nearest Neighbours machine learning toolkit.

- <u>deap</u> Evolutionary algorithm framework.
- <u>pydeep</u> Deep Learning In Python
- <u>mlxtend</u> A library consisting of useful tools for data science and machine learning tasks.
- <u>neon</u> Nervana's <u>high-performance</u> Python-based Deep Learning framework [DEEP LEARNING]
- Optunity A library dedicated to automated hyperparameter optimization with a simple, lightweight API to facilitate drop-in replacement of grid search.
- <u>Neural Networks and Deep Learning</u> Code samples for my book "Neural Networks and Deep Learning" [DEEP LEARNING]
- <u>Annoy</u> Approximate nearest neighbours implementation
- <u>skflow</u> Simplified interface for TensorFlow, mimicking Scikit Learn.
- <u>TPOT</u> Tool that automatically creates and optimizes machine learning pipelines using genetic programming. Consider it your personal data science assistant, automating a tedious part of machine learning.
- pgmpy A python library for working with Probabilistic Graphical Models.
- <u>DIGITS</u> The Deep Learning GPU Training System (DIGITS) is a web application for training deep learning models.
- Orange Open source data visualization and data analysis for novices and experts.
- MXNet Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Go, Javascript and more.
- <u>milk</u> Machine learning toolkit focused on supervised classification.
- <u>TFLearn</u> Deep learning library featuring a higher-level API for TensorFlow.
- REP an IPython-based environment for conducting data-driven research in a consistent and reproducible way. REP is not trying to substitute scikit-learn, but extends it and provides better user experience.
- rgf python Python bindings for Regularized Greedy Forest (Tree) Library.
- <u>skbayes</u> Python package for Bayesian Machine Learning with scikit-learn API
- <u>fuku-ml</u> Simple machine learning library, including Perceptron, Regression, Support Vector Machine, Decision Tree and more, it's easy to use and easy to learn for beginners.
- Xcessiv A web-based application for quick, scalable, and automated hyperparameter tuning and stacked ensembling
- <u>PyTorch</u> Tensors and Dynamic neural networks in Python with strong GPU acceleration
- ML-From-Scratch Implementations of Machine Learning models from scratch in Python with a focus on transparency. Aims to showcase the nuts and bolts of ML in an accessible way.

- <u>SciPy</u> A Python-based ecosystem of open-source software for mathematics, science, and engineering.
- NumPy A fundamental package for scientific computing with Python.
- <u>Numba</u> Python JIT (just in time) complier to LLVM aimed at scientific Python by the developers of Cython and NumPy.
- <u>NetworkX</u> A high-productivity software for complex networks.
- <u>igraph</u> binding to igraph library General purpose graph library
- <u>Pandas</u> A library providing high-performance, easy-to-use data structures and data analysis tools.
- Open Mining Business Intelligence (BI) in Python (Pandas web interface)
- PyMC Markov Chain Monte Carlo sampling toolkit.
- <u>zipline</u> A Pythonic algorithmic trading library.
- <u>PyDy</u> Short for Python Dynamics, used to assist with workflow in the modeling of dynamic motion based around NumPy, SciPy, IPython, and matplotlib.
- SymPy A Python library for symbolic mathematics.
- <u>statsmodels</u> Statistical modeling and econometrics in Python.
- <u>astropy</u> A community Python library for Astronomy.
- <u>matplotlib</u> A Python 2D plotting library.
- <u>bokeh</u> Interactive Web Plotting for Python.
- <u>plotly</u> Collaborative web plotting for Python and matplotlib.
- <u>vincent</u> A Python to Vega translator.
- <u>d3py</u> A plotting library for Python, based on <u>D3.js</u>.
- <u>PyDexter</u> Simple plotting for Python. Wrapper for D3xterjs; easily render charts inbrowser.
- ggplot Same API as ggplot2 for R.
- <u>ggfortify</u> Unified interface to ggplot2 popular R packages.
- <u>Kartograph.py</u> Rendering beautiful SVG maps in Python.
- <u>pygal</u> A Python SVG Charts Creator.
- <u>PyQtGraph</u> A pure-python graphics and GUI library built on PyQt4 / PySide and NumPy.
- pycascading
- <u>Petrel</u> Tools for writing, submitting, debugging, and monitoring Storm topologies in pure Python.
- Blaze NumPy and Pandas interface to Big Data.
- <u>emcee</u> The Python ensemble sampling toolkit for affine-invariant MCMC.
- windML A Python Framework for Wind Energy Analysis and Prediction
- vispy GPU-based high-performance interactive OpenGL 2D/3D data visualization library
- <u>cerebro2</u> A web-based visualization and debugging platform for NuPIC.

- <u>NuPIC Studio</u> An all-in-one NuPIC Hierarchical Temporal Memory visualization and debugging super-tool!
- SparklingPandas Pandas on PySpark (POPS)
- Seaborn A python visualization library based on matplotlib
- <u>baplot</u> An API for plotting in Jupyter (IPython)
- <u>pastalog</u> Simple, realtime visualization of neural network training performance.
- <u>caravel</u> A data exploration platform designed to be visual, intuitive, and interactive.
- <u>Dora</u> Tools for exploratory data analysis in Python.
- Ruffus Computation Pipeline library for python.
- <u>SOMPY</u> Self Organizing Map written in Python (Uses neural networks for data analysis).
- <u>somoclu</u> Massively parallel self-organizing maps: accelerate training on multicore CPUs, GPUs, and clusters, has python API.
- HDBScan implementation of the hdbscan algorithm in Python used for clustering
- <u>visualize ML</u> A python package for data exploration and data analysis.
- <u>scikit-plot</u> A visualization library for quick and easy generation of common plots in data analysis and machine learning.
- <u>Bowtie</u> A dashboard library for interactive visualizations using flask socketio and react.

Misc Scripts / iPython Notebooks / Codebases

- <u>BioPy</u> Biologically-Inspired and Machine Learning Algorithms in Python.
- pattern classification
- thinking stats 2
- <u>hyperopt</u>
- <u>numpic</u>
- <u>2012-paper-diginorm</u>
- A gallery of interesting IPython notebooks
- ipython-notebooks
- <u>data-science-ipython-notebooks</u> Continually updated Data Science Python Notebooks: Spark, Hadoop MapReduce, HDFS, AWS, Kaggle, scikit-learn, matplotlib, pandas, NumPy, SciPy, and various command lines.
- decision-weights
- <u>Sarah Palin LDA</u> Topic Modeling the Sarah Palin emails.
- <u>Diffusion Segmentation</u> A collection of image segmentation algorithms based on diffusion methods
- Scipy Tutorials SciPy tutorials. This is outdated, check out scipy-lecture-notes
- Crab A recommendation engine library for Python
- <u>BayesPy</u> Bayesian Inference Tools in Python
- <u>scikit-learn tutorials</u> Series of notebooks for learning scikit-learn
- <u>sentiment-analyzer</u> Tweets Sentiment Analyzer

- <u>sentiment_classifier</u> Sentiment classifier using word sense disambiguation.
- <u>group-lasso</u> Some experiments with the coordinate descent algorithm used in the (Sparse) Group Lasso model
- <u>jProcessing</u> Kanji / Hiragana / Katakana to Romaji Converter. Edict Dictionary & parallel sentences Search. Sentence Similarity between two JP Sentences. Sentiment Analysis of Japanese Text. Run Cabocha(ISO--8859-1 configured) in Python.
- <u>mne-python-notebooks</u> IPython notebooks for EEG/MEG data processing using mne-python
- <u>Neon Course</u> IPython notebooks for a complete course around understanding Nervana's Neon
- pandas cookbook Recipes for using Python's pandas library
- <u>climin</u> Optimization library focused on machine learning, pythonic implementations of gradient descent, LBFGS, rmsprop, adadelta and others
- Allen Downey's Data Science Course Code for Data Science at Olin College, Spring 2014.
- Allen Downey's Think Bayes Code Code repository for Think Bayes.
- Allen Downey's Think Complexity Code Code for Allen Downey's book Think Complexity.
- <u>Allen Downey's Think OS Code</u> Text and supporting code for Think OS: A Brief Introduction to Operating Systems.
- <u>Python Programming for the Humanities</u> Course for Python programming for the Humanities, assuming no prior knowledge. Heavy focus on text processing / NLP.
- GreatCircle Library for calculating great circle distance.
- Optunity examples Examples demonstrating how to use Optunity in synergy with machine learning libraries.
- <u>Dive into Machine Learning with Python Jupyter notebook and scikit-learn</u> "I learned Python by hacking first, and getting serious *later*. I wanted to do this with Machine Learning. If this is your style, join me in getting a bit ahead of yourself."
- <u>TDB</u> TensorDebugger (TDB) is a visual debugger for deep learning. It features interactive, node-by-node debugging and visualization for TensorFlow.
- Suiron Machine Learning for RC Cars.
- <u>Introduction to machine learning with scikit-learn</u> IPython notebooks from Data School's video tutorials on scikit-learn.
- <u>Practical XGBoost in Python</u> comprehensive online course about using XGBoost in Python

Neural Networks

- NeuralTalk NeuralTalk is a Python+numpy project for learning Multimodal Recurrent Neural Networks that describe images with sentences.
- <u>Neuron</u> Neuron is simple class for time series predictions. It's utilize LNU (Linear Neural Unit), QNU (Quadratic Neural Unit), RBF (Radial Basis Function), MLP (Multi Layer Perceptron), MLP-ELM (Multi Layer Perceptron - Extreme Learning Machine) neural networks learned with Gradient descent or LeLevenberg–Marquardt algorithm.

• <u>Data Driven Code</u> - Very simple implementation of neural networks for dummies in python without using any libraries, with detailed comments.

Kaggle Competition Source Code

- wiki challenge An implementation of Dell Zhang's solution to Wikipedia's Participation Challenge on Kaggle
- <u>kaggle insults</u> Kaggle Submission for "Detecting Insults in Social Commentary"
- <u>kaggle_acquire-valued-shoppers-challenge</u> Code for the Kaggle acquire valued shoppers challenge
- <u>kaggle-cifar</u> Code for the CIFAR-10 competition at Kaggle, uses cuda-convnet
- <u>kaggle-blackbox</u> Deep learning made easy
- <u>kaggle-accelerometer</u> Code for Accelerometer Biometric Competition at Kaggle
- <u>kaggle-advertised-salaries</u> Predicting job salaries from ads a Kaggle competition
- kaggle amazon Amazon access control challenge
- <u>kaggle-bestbuy_big</u> Code for the Best Buy competition at Kaggle
- <u>kaggle-bestbuy_small</u>
- <u>Kaggle Dogs vs. Cats</u> Code for Kaggle Dogs vs. Cats competition
- <u>Kaggle Galaxy Challenge</u> Winning solution for the Galaxy Challenge on Kaggle
- <u>Kaggle Gender</u> A Kaggle competition: discriminate gender based on handwriting
- Kaggle Merck Merck challenge at Kaggle
- <u>Kaggle Stackoverflow</u> Predicting closed questions on Stack Overflow
- <u>kaggle_acquire-valued-shoppers-challenge</u> Code for the Kaggle acquire valued shoppers challenge
- wine-quality Predicting wine quality

Reinforcement Learning

- <u>DeepMind Lab</u> DeepMind Lab is a 3D learning environment based on id Software's
 Quake III Arena via ioquake3 and other open source software. Its primary purpose is to
 act as a testbed for research in artificial intelligence, especially deep reinforcement
 learning.
- Gym OpenAl Gym is a toolkit for developing and comparing reinforcement learning algorithms.
- <u>Universe</u> Universe is a software platform for measuring and training an Al's general intelligence across the world's supply of games, websites and other applications.
- <u>ViZDoom</u> ViZDoom allows developing AI bots that play Doom using only the visual information (the screen buffer). It is primarily intended for research in machine visual learning, and deep reinforcement learning, in particular.

Ruby

Natural Language Processing

- Awesome NLP with Ruby Curated link list for practical natural language processing in Ruby.
- <u>Treat</u> Text REtrieval and Annotation Toolkit, definitely the most comprehensive toolkit I've encountered so far for Ruby
- <u>Ruby Linguistics</u> Linguistics is a framework for building linguistic utilities for Ruby objects in any language. It includes a generic language-independent front end, a module for mapping language codes into language names, and a module which contains various English-language utilities.
- <u>Stemmer</u> Expose libstemmer_c to Ruby
- Ruby Wordnet This library is a Ruby interface to WordNet
- Raspel raspell is an interface binding for ruby
- <u>UEA Stemmer</u> Ruby port of UEALite Stemmer a conservative stemmer for search and indexing
- <u>Twitter-text-rb</u> A library that does auto linking and extraction of usernames, lists and hashtags in tweets

General-Purpose Machine Learning

- <u>Awesome Machine Learning with Ruby</u> Curated list of ML related resources for Ruby
- Ruby Machine Learning Some Machine Learning algorithms, implemented in Ruby
- Machine Learning Ruby
- <u>jRuby Mahout</u> JRuby Mahout is a gem that unleashes the power of Apache Mahout in the world of JRuby.
- <u>CardMagic-Classifier</u> A general classifier module to allow Bayesian and other types of classifications.
- <u>rb-libsvm</u> Ruby language bindings for LIBSVM which is a Library for Support Vector Machines
- Random Forester Creates Random Forest classifiers from PMML files

Data Analysis / Data Visualization

- rsruby Ruby R bridge
- <u>data-visualization-ruby</u> Source code and supporting content for my Ruby Manor presentation on Data Visualisation with Ruby
- ruby-plot gnuplot wrapper for ruby, especially for plotting roc curves into svg files
- plot-rb A plotting library in Ruby built on top of Vega and D3.
- scruffy A beautiful graphing toolkit for Ruby
- SciRuby
- Glean A data management tool for humans
- Bioruby
- Arel

- Big Data For Chimps
- <u>Listof</u> Community based data collection, packed in gem. Get list of pretty much anything (stop words, countries, non words) in txt, json or hash. <u>Demo/Search for a list</u>

Rust

General-Purpose Machine Learning

- deeplearn-rs deeplearn-rs provides simple networks that use matrix multiplication, addition, and ReLU under the MIT license.
- <u>rustlearn</u> a machine learning framework featuring logistic regression, support vector machines, decision trees and random forests.
- <u>rusty-machine</u> a pure-rust machine learning library.
- <u>leaf</u> open source framework for machine intelligence, sharing concepts from TensorFlow and Caffe. Available under the MIT license. [Deprecated]
- <u>RustNN</u> RustNN is a feedforward neural network library.

R

General-Purpose Machine Learning

- <u>ahaz</u> ahaz: Regularization for semiparametric additive hazards regression
- <u>arules</u> arules: Mining Association Rules and Frequent Itemsets
- <u>biglasso</u> biglasso: Extending Lasso Model Fitting to Big Data in R
- <u>bigrf</u> bigrf: Big Random Forests: Classification and Regression Forests for Large Data Sets
- $\underline{\text{bigRR}}$ $\underline{\text{bigRR}}$: Generalized Ridge Regression (with special advantage for p >> n cases)
- <u>bmrm</u> bmrm: Bundle Methods for Regularized Risk Minimization Package
- Boruta Boruta: A wrapper algorithm for all-relevant feature selection
- <u>bst</u> bst: Gradient Boosting
- C50 C50: C5.0 Decision Trees and Rule-Based Models
- <u>caret</u> Classification and Regression Training: Unified interface to ~150 ML algorithms in R.
- <u>caretEnsemble</u> caretEnsemble: Framework for fitting multiple caret models as well as creating ensembles of such models.
- Clever Algorithms For Machine Learning
- <u>CORElearn</u> CORElearn: Classification, regression, feature evaluation and ordinal evaluation
- <u>CoxBoost</u> CoxBoost: Cox models by likelihood based boosting for a single survival endpoint or competing risks

- <u>Cubist</u> Cubist: Rule- and Instance-Based Regression Modeling
- e1071 e1071: Misc Functions of the Department of Statistics (e1071), TU Wien
- <u>earth</u> earth: Multivariate Adaptive Regression Spline Models
- <u>elasticnet</u> elasticnet: Elastic-Net for Sparse Estimation and Sparse PCA
- <u>ElemStatLearn</u> ElemStatLearn: Data sets, functions and examples from the book: "The Elements of Statistical Learning, Data Mining, Inference, and Prediction" by Trevor Hastie, Robert Tibshirani and Jerome Friedman Prediction" by Trevor Hastie, Robert Tibshirani and Jerome Friedman
- evtree evtree: Evolutionary Learning of Globally Optimal Trees
- <u>forecast</u> forecast: Timeseries forecasting using ARIMA, ETS, STLM, TBATS, and neural network models
- <u>forecastHybrid</u> forecastHybrid: Automatic ensemble and cross validation of ARIMA, ETS, STLM, TBATS, and neural network models from the "forecast" package
- <u>fpc</u> fpc: Flexible procedures for clustering
- <u>frbs</u> frbs: Fuzzy Rule-based Systems for Classification and Regression Tasks
- GAMBoost GAMBoost: Generalized linear and additive models by likelihood based boosting
- gamboostLSS gamboostLSS: Boosting Methods for GAMLSS
- <u>gbm</u> gbm: Generalized Boosted Regression Models
- glmnet glmnet: Lasso and elastic-net regularized generalized linear models
- glmpath glmpath: L1 Regularization Path for Generalized Linear Models and Cox Proportional Hazards Model
- GMMBoost GMMBoost: Likelihood-based Boosting for Generalized mixed models
- grplasso grplasso: Fitting user specified models with Group Lasso penalty
- grpreg grpreg: Regularization paths for regression models with grouped covariates
- <u>h2o</u> A framework for fast, parallel, and distributed machine learning algorithms at scale -- Deeplearning, Random forests, GBM, KMeans, PCA, GLM
- <u>hda</u> hda: Heteroscedastic Discriminant Analysis
- Introduction to Statistical Learning
- <u>ipred</u> ipred: Improved Predictors
- <u>kernlab</u> kernlab: Kernel-based Machine Learning Lab
- <u>klaR</u> klaR: Classification and visualization
- <u>lars</u> lars: Least Angle Regression, Lasso and Forward Stagewise
- lasso2 lasso2: L1 constrained estimation aka 'lasso'
- <u>LiblineaR</u> LiblineaR: Linear Predictive Models Based On The Liblinear C/C++ Library
- <u>LogicReg</u> LogicReg: Logic Regression
- Machine Learning For Hackers
- maptree maptree: Mapping, pruning, and graphing tree models
- <u>mboost</u> mboost: Model-Based Boosting
- medley medley: Blending regression models, using a greedy stepwise approach
- mlr mlr: Machine Learning in R

- <u>mvpart</u> mvpart: Multivariate partitioning
- ncvreg ncvreg: Regularization paths for SCAD- and MCP-penalized regression models
- <u>nnet</u> nnet: Feed-forward Neural Networks and Multinomial Log-Linear Models
- <u>oblique.tree</u> oblique.tree: Oblique Trees for Classification Data
- pamr pamr: Pam: prediction analysis for microarrays
- party party: A Laboratory for Recursive Partytioning
- partykit partykit: A Toolkit for Recursive Partytioning
- <u>penalized</u> penalized: L1 (lasso and fused lasso) and L2 (ridge) penalized estimation in GLMs and in the Cox model
- <u>penalizedLDA</u> penalizedLDA: Penalized classification using Fisher's linear discriminant
- <u>penalizedSVM</u> penalizedSVM: Feature Selection SVM using penalty functions
- quantregForest quantregForest: Quantile Regression Forests
- <u>randomForest</u> randomForest: Breiman and Cutler's random forests for classification and regression
- <u>randomForestSRC</u> randomForestSRC: Random Forests for Survival, Regression and Classification (RF-SRC)
- rattle rattle: Graphical user interface for data mining in R
- rda rda: Shrunken Centroids Regularized Discriminant Analysis
- <u>rdetools</u> rdetools: Relevant Dimension Estimation (RDE) in Feature Spaces
- <u>REEMtree</u> REEMtree: Regression Trees with Random Effects for Longitudinal (Panel)
 Data
- relaxo relaxo: Relaxed Lasso
- rgenoud rgenoud: R version of GENetic Optimization Using Derivatives
- rgp rgp: R genetic programming framework
- <u>Rmalschains</u> Rmalschains: Continuous Optimization using Memetic Algorithms with Local Search Chains (MA-LS-Chains) in R
- <u>rminer</u> rminer: Simpler use of data mining methods (e.g. NN and SVM) in classification and regression
- ROCR ROCR: Visualizing the performance of scoring classifiers
- RoughSets RoughSets: Data Analysis Using Rough Set and Fuzzy Rough Set Theories
- rpart rpart: Recursive Partitioning and Regression Trees
- RPMM RPMM: Recursively Partitioned Mixture Model
- <u>RSNNS</u> RSNNS: Neural Networks in R using the Stuttgart Neural Network Simulator (SNNS)
- RWeka RWeka: R/Weka interface
- RXshrink RXshrink: Maximum Likelihood Shrinkage via Generalized Ridge or Least Angle Regression
- <u>sda</u> sda: Shrinkage Discriminant Analysis and CAT Score Variable Selection
- <u>SDDA</u> SDDA: Stepwise Diagonal Discriminant Analysis
- <u>SuperLearner</u> and <u>subsemble</u> Multi-algorithm ensemble learning packages.

- svmpath svmpath: svmpath: the SVM Path algorithm
- tqp tqp: Bayesian treed Gaussian process models
- tree: Classification and regression trees
- <u>varSelRF</u> varSelRF: Variable selection using random forests
- XGBoost.R R binding for eXtreme Gradient Boosting (Tree) Library
- Optunity A library dedicated to automated hyperparameter optimization with a simple, lightweight API to facilitate drop-in replacement of grid search. Optunity is written in Python but interfaces seamlessly to R.
- <u>igraph</u> binding to igraph library General purpose graph library
- <u>MXNet</u> Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Go, Javascript and more.
- <u>TDSP-Utilities</u> Two data science utilities in R from Microsoft: 1) Interactive Data Exploration, Analysis, and Reporting (IDEAR); 2) Automated Modeling and Reporting (AMR).

Data Analysis / Data Visualization

<u>qqplot2</u> - A data visualization package based on the grammar of graphics.

SAS

General-Purpose Machine Learning

- <u>Enterprise Miner</u> Data mining and machine learning that creates deployable models using a GUI or code.
- <u>Factory Miner</u> Automatically creates deployable machine learning models across numerous market or customer segments using a GUI.

Data Analysis / Data Visualization

- <u>SAS/STAT</u> For conducting advanced statistical analysis.
- <u>University Edition</u> FREE! Includes all SAS packages necessary for data analysis and visualization, and includes online SAS courses.

High Performance Machine Learning

- <u>High Performance Data Mining</u> Data mining and machine learning that creates deployable models using a GUI or code in an MPP environment, including Hadoop.
- <u>High Performance Text Mining</u> Text mining using a GUI or code in an MPP environment, including Hadoop.

Natural Language Processing

- Contextual Analysis Add structure to unstructured text using a GUI.
- Sentiment Analysis Extract sentiment from text using a GUI.
- Text Miner Text mining using a GUI or code.

Demos and Scripts

- ML_Tables Concise cheat sheets containing machine learning best practices.
- <u>enlighten-apply</u> Example code and materials that illustrate applications of SAS machine learning techniques.
- <u>enlighten-integration</u> Example code and materials that illustrate techniques for integrating SAS with other analytics technologies in Java, PMML, Python and R.
- <u>enlighten-deep</u> Example code and materials that illustrate using neural networks with several hidden layers in SAS.
- <u>dm-flow</u> Library of SAS Enterprise Miner process flow diagrams to help you learn by example about specific data mining topics.

Scala

Natural Language Processing

- <u>ScalaNLP</u> ScalaNLP is a suite of machine learning and numerical computing libraries.
- <u>Breeze</u> Breeze is a numerical processing library for Scala.
- <u>Chalk</u> Chalk is a natural language processing library.
- FACTORIE FACTORIE is a toolkit for deployable probabilistic modeling, implemented as a software library in Scala. It provides its users with a succinct language for creating relational factor graphs, estimating parameters and performing inference.
- Montague Montague is a semantic parsing library for Scala with an easy-to-use DSL.

Data Analysis / Data Visualization

- MLlib in Apache Spark Distributed machine learning library in Spark
- <u>Hydrosphere Mist</u> a service for deployment Apache Spark MLLib machine learning models as realtime, batch or reactive web services.
- Scalding A Scala API for Cascading
- <u>Summing Bird</u> Streaming MapReduce with Scalding and Storm
- Algebird Abstract Algebra for Scala
- xerial Data management utilities for Scala
- <u>PredictionIO</u> PredictionIO, a machine learning server for software developers and data engineers.
- <u>BIDMat</u> CPU and GPU-accelerated matrix library intended to support large-scale exploratory data analysis.

- Flink Open source platform for distributed stream and batch data processing.
- Spark Notebook Interactive and Reactive Data Science using Scala and Spark.

General-Purpose Machine Learning

- <u>DeepLearning.scala</u> Creating statically typed dynamic neural networks from objectoriented & functional programming constructs.
- <u>Conjecture</u> Scalable Machine Learning in Scalding
- <u>brushfire</u> Distributed decision tree ensemble learning in Scala
- ganitha scalding powered machine learning
- <u>adam</u> A genomics processing engine and specialized file format built using Apache Avro, Apache Spark and Parquet. Apache 2 licensed.
- <u>bioscala</u> Bioinformatics for the Scala programming language
- <u>BIDMach</u> CPU and GPU-accelerated Machine Learning Library.
- Figaro a Scala library for constructing probabilistic models.
- H2O Sparkling Water H2O and Spark interoperability.
- <u>FlinkML in Apache Flink</u> Distributed machine learning library in Flink
- <u>DynaML</u> Scala Library/REPL for Machine Learning Research
- <u>Saul</u> Flexible Declarative Learning-Based Programming.
- <u>SwiftLearner</u> Simply written algorithms to help study ML or write your own implementations.
- Smile Statistical Machine Intelligence and Learning Engine

Swift

General-Purpose Machine Learning

- <u>Bender</u> Fast Neural Networks framework built on top of Metal. Supports TensorFlow models.
- <u>Swift Al</u> Highly optimized artificial intelligence and machine learning library written in Swift.
- BrainCore The iOS and OS X neural network framework
- <u>swix</u> A bare bones library that includes a general matrix language and wraps some OpenCV for iOS development.
- <u>DeepLearningKit</u> an Open Source Deep Learning Framework for Apple's iOS, OS X and tvOS. It currently allows using deep convolutional neural network models trained in Caffe on Apple operating systems.
- <u>AlToolbox</u> A toolbox framework of Al modules written in Swift: Graphs/Trees, Linear Regression, Support Vector Machines, Neural Networks, PCA, KMeans, Genetic Algorithms, MDP, Mixture of Gaussians.
- <u>MLKit</u> A simple Machine Learning Framework written in Swift. Currently features Simple Linear Regression, Polynomial Regression, and Ridge Regression.

- <u>Swift Brain</u> The first neural network / machine learning library written in Swift. This is a project for AI algorithms in Swift for iOS and OS X development. This project includes algorithms focused on Bayes theorem, neural networks, SVMs, Matrices, etc..
- <u>Perfect TensorFlow</u> Swift Language Bindings of TensorFlow. Using native TensorFlow models on both macOS / Linux.
- Awesome CoreML A curated list of pretrained CoreML models
- Awesome Core ML Models A curated list of machine learning models in CoreML format.

TensorFlow

General-Purpose Machine Learning

Awesome TensorFlow - A list of all things related to TensorFlow

Awesome Public Datasets

https://github.com/caesar0301/awesome-public-datasets

<u>This list of a topic-centric public data sources</u> in high quality. They are collected and tidied from blogs, answers, and user responses. Most of the data sets listed below are free, however, some are not. Other amazingly awesome lists can be found in the <u>awesome-awesomeness</u> and <u>sindresorhus's awesome</u> list.

Table of Contents

- Agriculture
- Biology
- Climate/Weather
- Complex Networks
- Computer Networks
- Data Challenges
- Earth Science
- Economics
- Education
- Energy
- Finance
- GIS
- Government
- Healthcare
- Image Processing

- Machine Learning
- Museums
- Natural Language
- Neuroscience
- Physics
- <u>Psychology/Cognition</u>
- Public Domains
- Search Engines
- Social Networks
- Social Sciences
- Software
- Sports
- Time Series
- Transportation
- Complementary Collections

Agriculture

- U.S. Department of Agriculture's PLANTS Database
- U.S. Department of Agriculture's Nutrient Database

Biology

- 1000 Genomes
- American Gut (Microbiome Project)
- Broad Bioimage Benchmark Collection (BBBC)
- Broad Cancer Cell Line Encyclopedia (CCLE)
- Cell Image Library
- Complete Genomics Public Data
- EBI ArrayExpress

- EBI Protein Data Bank in Europe
- Electron Microscopy Pilot Image Archive (EMPIAR)
- ENCODE project
- Ensembl Genomes
- Gene Expression Omnibus (GEO)
- Gene Ontology (GO)
- Global Biotic Interactions (GloBI)
- Harvard Medical School (HMS) LINCS Project
- Human Genome Diversity Project
- Human Microbiome Project (HMP)
- ICOS PSP Benchmark
- International HapMap Project
- Journal of Cell Biology DataViewer
- MIT Cancer Genomics Data
- NCBI Proteins
- NCBI Taxonomy
- NCI Genomic Data Commons
- NIH Microarray data or FTP (see FTP link on RAW)
- OpenSNP genotypes data
- Pathquid Protein-Protein Interactions Catalog
- Protein Data Bank
- Psychiatric Genomics Consortium
- PubChem Project
- PubGene (now Coremine Medical)
- Sanger Catalogue of Somatic Mutations in Cancer (COSMIC)
- Sanger Genomics of Drug Sensitivity in Cancer Project (GDSC)
- Sequence Read Archive(SRA)
- Stanford Microarray Data
- Stowers Institute Original Data Repository
- Systems Science of Biological Dynamics (SSBD) Database
- The Cancer Genome Atlas (TCGA), available via Broad GDAC
- The Catalogue of Life
- The Personal Genome Project or PGP
- UCSC Public Data
- UniGene
- Universal Protein Resource (UnitProt)

Climate/Weather

Actuaries Climate Index

- Australian Weather
- Aviation Weather Center Consistent, timely and accurate weather information for the world airspace system
- Brazilian Weather Historical data (In Portuguese)
- Canadian Meteorological Centre
- Climate Data from UEA (updated monthly)
- European Climate Assessment & Dataset
- Global Climate Data Since 1929
- NASA Global Imagery Browse Services
- NOAA Bering Sea Climate
- NOAA Climate Datasets
- NOAA Realtime Weather Models
- NOAA SURFRAD Meteorology and Radiation Datasets
- The World Bank Open Data Resources for Climate Change
- UEA Climatic Research Unit
- WorldClim Global Climate Data
- WU Historical Weather Worldwide

Complex Networks

- AMiner Citation Network Dataset
- CrossRef DOI URLs
- DBLP Citation dataset
- DIMACS Road Networks Collection
- NBER Patent Citations
- Network Repository with Interactive Exploratory Analysis Tools
- NIST complex networks data collection
- Protein-protein interaction network
- PyPI and Maven Dependency Network
- Scopus Citation Database
- Small Network Data
- Stanford GraphBase (Steven Skiena)
- Stanford Large Network Dataset Collection
- Stanford Longitudinal Network Data Sources
- The Koblenz Network Collection
- The Laboratory for Web Algorithmics (UNIMI)
- The Nexus Network Repository
- UCI Network Data Repository
- UFL sparse matrix collection
- WSU Graph Database

Computer Networks

- 3.5B Web Pages from CommonCrawl 2012
- 53.5B Web clicks of 100K users in Indiana Univ.
- CAIDA Internet Datasets
- ClueWeb09 1B web pages
- ClueWeb12 733M web pages
- CommonCrawl Web Data over 7 years
- CRAWDAD Wireless datasets from Dartmouth Univ.
- Criteo click-through data
- OONI: Open Observatory of Network Interference Internet censorship data
- Open Mobile Data by MobiPerf
- Rapid7 Sonar Internet Scans
- UCSD Network Telescope, IPv4 /8 net

Data Challenges

- Bruteforce Database
- Challenges in Machine Learning
- CrowdANALYTIX dataX
- <u>D4D Challenge of Orange</u>
- DrivenData Competitions for Social Good
- ICWSM Data Challenge (since 2009)
- Kaggle Competition Data
- KDD Cup by Tencent 2012
- Localytics Data Visualization Challenge
- Netflix Prize
- Space Apps Challenge
- Telecom Italia Big Data Challenge
- TravisTorrent Dataset MSR'2017 Mining Challenge
- Yelp Dataset Challenge

Earth Science

- AQUASTAT Global water resources and uses
- BODC marine data of ~22K vars
- Earth Models
- EOSDIS NASA's earth observing system data
- Integrated Marine Observing System (IMOS) roughly 30TB of ocean measurements or on S3

- Marinexplore Open Oceanographic Data
- Smithsonian Institution Global Volcano and Eruption Database
- USGS Earthquake Archives

Economics

- American Economic Association (AEA)
- EconData from UMD
- Economic Freedom of the World Data
- Historical MacroEconomc Statistics
- International Economics Database and various data tools
- International Trade Statistics
- Internet Product Code Database
- Joint External Debt Data Hub
- Jon Haveman International Trade Data Links
- OpenCorporates Database of Companies in the World
- Our World in Data
- SciencesPo World Trade Gravity Datasets
- The Atlas of Economic Complexity
- The Center for International Data
- The Observatory of Economic Complexity
- UN Commodity Trade Statistics
- UN Human Development Reports

Education

- College Scorecard Data
- Student Data from Free Code Camp

Energy

- AMPds
- BLUEd
- COMBED
- Dataport
- DRED
- ECO
- EIA
- HES Household Electricity Study, UK
- HFED

- iAWE
- PLAID the Plug Load Appliance Identification Dataset
- REDD
- Tracebase
- <u>UK-DALE</u> UK Domestic Appliance-Level Electricity
- WHITED

Finance

- CBOE Futures Exchange
- Google Finance
- Google Trends
- NASDAQ
- NYSE Market Data (see FTP link on <u>RAW</u>)
- OANDA
- OSU Financial data
- Quandl
- St Louis Federal
- Yahoo Finance

GIS

- ArcGIS Open Data portal
- Cambridge, MA, US, GIS data on GitHub
- Factual Global Location Data
- Geo Spatial Data from ASU
- Geo Wiki Project Citizen-driven Environmental Monitoring
- GeoFabrik OSM data extracted to a variety of formats and areas
- GeoNames Worldwide
- Global Administrative Areas Database (GADM)
- Homeland Infrastructure Foundation-Level Data
- Landsat 8 on AWS
- <u>List of all countries in all languages</u>
- National Weather Service GIS Data Portal
- Natural Earth vectors and rasters of the world
- OpenAddresses
- OpenStreetMap (OSM)
- Pleiades Gazetteer and graph of ancient places
- Reverse Geocoder using OSM data & additional high-resolution data files
- TIGER/Line U.S. boundaries and roads

- TwoFishes Foursquare's coarse geocoder
- TZ Timezones shapfiles
- UN Environmental Data
- World boundaries from the U.S. Department of State
- World countries in multiple formats

Government

- A list of cities and countries contributed by community
- Open Data for Africa
- OpenDataSoft's list of 1,600 open data

Healthcare

- EHDP Large Health Data Sets
- Gapminder World demographic databases
- GDC supports several cancer genome programs for CCG, TCGA, TARGET etc.
- PhysioBank Databases a large and growing archive of physiological data
- Medicare Coverage Database (MCD), U.S.
- Medicare Data Engine of medicare.gov Data
- Medicare Data File
- MeSH, the vocabulary thesaurus used for indexing articles for PubMed
- Number of Ebola Cases and Deaths in Affected Countries (2014)
- Open-ODS (structure of the UK NHS)
- OpenPaymentsData, Healthcare financial relationship data
- The Cancer Genome Atlas project (TCGA) (refer to GDC and BigQuery table)
- World Health Organization Global Health Observatory

Image Processing

- 10k US Adult Faces Database
- 2GB of Photos of Cats or Archive version
- Adience Unfiltered faces for gender and age classification
- Affective Image Classification
- Animals with attributes
- Caltech Pedestrian Detection Benchmark
- <u>Chars74K dataset, Character Recognition in Natural Images (both English and Kannada are available)</u>
- Face Recognition Benchmark
- Flickr: 32 Class Brand Logos

- GDXray: X-ray images for X-ray testing and Computer Vision
- ImageNet (in WordNet hierarchy)
- Indoor Scene Recognition
- International Affective Picture System, UFL
- Massive Visual Memory Stimuli, MIT
- MNIST database of handwritten digits, near 1 million examples
- Several Shape-from-Silhouette Datasets
- Stanford Dogs Dataset
- SUN database, MIT
- The Action Similarity Labeling (ASLAN) Challenge
- The Oxford-IIIT Pet Dataset
- Violent-Flows Crowd Violence Non-violence Database and benchmark
- Visual genome
- YouTube Faces Database

Machine Learning

- Context-aware data sets from five domains
- Delve Datasets for classification and regression (Univ. of Toronto)
- Discogs Monthly Data
- eBay Online Auctions (2012)
- IMDb Database
- Keel Repository for classification, regression and time series
- Labeled Faces in the Wild (LFW)
- Lending Club Loan Data
- Machine Learning Data Set Repository
- Free Music Archive
- Million Song Dataset
- More Song Datasets
- MovieLens Data Sets
- New Yorker caption contest ratings
- RDataMining "R and Data Mining" ebook data
- Registered Meteorites on Earth
- Restaurants Health Score Data in San Francisco
- UCI Machine Learning Repository
- Yahoo! Ratings and Classification Data
- Youtube 8m

Museums

- Canada Science and Technology Museums Corporation's Open Data
- Cooper-Hewitt's Collection Database
- Minneapolis Institute of Arts metadata
- Natural History Museum (London) Data Portal
- Rijksmuseum Historical Art Collection
- Tate Collection metadata
- The Getty vocabularies

Natural Language

- Automatic Keyphrase Extraction
- Blogger Corpus
- <u>CLiPS Stylometry Investigation Corpus</u>
- ClueWeb09 FACC
- ClueWeb12 FACC
- DBpedia 4.58M things with 583M facts
- Flickr Personal Taxonomies
- Freebase.com of people, places, and things
- Google Books Ngrams (2.2TB)
- Google MC-AFP, generated based on the public available Gigaword dataset using Paragraph Vectors
- Google Web 5gram (1TB, 2006)
- Gutenberg eBooks List
- Hansards text chunks of Canadian Parliament
- Machine Comprehension Test (MCTest) of text from Microsoft Research
- Machine Translation of European languages
- Microsoft MAchine Reading COmprehension Dataset (or MS MARCO)
- Multi-Domain Sentiment Dataset (version 2.0)
- Open Multilingual Wordnet
- Personae Corpus
- SaudiNewsNet Collection of Saudi Newspaper Articles (Arabic, 30K articles)
- SMS Spam Collection in English
- <u>Universal Dependencies</u>
- USENET postings corpus of 2005~2011
- Webhose News/Blogs in multiple languages
- Wikidata Wikipedia databases
- Wikipedia Links data 40 Million Entities in Context
- WordNet databases and tools

Neuroscience

- Allen Institute Datasets
- Brain Catalogue
- Brainomics
- CodeNeuro Datasets
- Collaborative Research in Computational Neuroscience (CRCNS)
- FCP-INDI
- Human Connectome Project
- NDAR
- NeuroData
- Neuroelectro
- NIMH Data Archive
- OASIS
- OpenfMRI
- Study Forrest

Physics

- CERN Open Data Portal
- Crystallography Open Database
- NASA Exoplanet Archive
- NSSDC (NASA) data of 550 space spacecraft
- Sloan Digital Sky Survey (SDSS) Mapping the Universe

Psychology/Cognition

OSU Cognitive Modeling Repository Datasets

Public Domains

- Amazon
- Archive-it from Internet Archive
- Archive.org Datasets
- CMU JASA data archive
- CMU StatLab collections
- Data.World
- Data360
- Google

- Infochimps
- KDNuggets Data Collections
- Microsoft Azure Data Market Free DataSets
- Microsoft Data Science for Research
- <u>Numbray</u>
- Open Library Data Dumps
- Reddit Datasets
- RevolutionAnalytics Collection
- Sample R data sets
- Stats4Stem R data sets
- StatSci.org
- The Washington Post List
- UCLA SOCR data collection
- **UFO Reports**
- Wikileaks 911 pager intercepts
- Yahoo Webscope

Search Engines

- Academic Torrents of data sharing from UMB
- Datahub.io
- DataMarket (Qlik)
- Harvard Dataverse Network of scientific data
- ICPSR (UMICH)
- Institute of Education Sciences
- National Technical Reports Library
- Open Data Certificates (beta)
- OpenDataNetwork A search engine of all Socrata powered data portals
- Statista.com statistics and Studies
- Zenodo An open dependable home for the long-tail of science

Social Networks

- 72 hours #gamergate Twitter Scrape
- Ancestry.com Forum Dataset over 10 years
- Cheng-Caverlee-Lee September 2009 January 2010 Twitter Scrape
- CMU Enron Email of 150 users
- EDRM Enron EMail of 151 users, hosted on S3
- Facebook Data Scrape (2005)
- Facebook Social Networks from LAW (since 2007)

- Foursquare from UMN/Sarwat (2013)
- GitHub Collaboration Archive
- Google Scholar citation relations
- High-Resolution Contact Networks from Wearable Sensors
- Indie Map: social graph and crawl of top IndieWeb sites
- Mobile Social Networks from UMASS
- Network Twitter Data
- Reddit Comments
- Skytrax' Air Travel Reviews Dataset
- Social Twitter Data
- SourceForge.net Research Data
- Twitter Data for Online Reputation Management
- Twitter Data for Sentiment Analysis
- Twitter Graph of entire Twitter site
- Twitter Scrape Calufa May 2011
- UNIMI/LAW Social Network Datasets
- Yahoo! Graph and Social Data
- Youtube Video Social Graph in 2007,2008

Social Sciences

- ACLED (Armed Conflict Location & Event Data Project)
- Canadian Legal Information Institute
- Center for Systemic Peace Datasets Conflict Trends, Polities, State Fragility, etc.
- Correlates of War Project
- Cryptome Conspiracy Theory Items
- Datacards
- European Social Survey
- FBI Hate Crime 2013 aggregated data
- Fragile States Index
- GDELT Global Events Database
- General Social Survey (GSS) since 1972
- German Social Survey
- Global Religious Futures Project
- Humanitarian Data Exchange
- INFORM Index for Risk Management
- Institute for Demographic Studies
- International Networks Archive
- International Social Survey Program ISSP
- International Studies Compendium Project

- James McGuire Cross National Data
- MacroData Guide by Norsk samfunnsvitenskapelig datatjeneste
- Minnesota Population Center
- MIT Reality Mining Dataset
- Notre Dame Global Adaptation Index (NG-DAIN)
- Open Crime and Policing Data in England, Wales and Northern Ireland
- Paul Hensel General International Data Page
- PewResearch Internet Survey Project
- PewResearch Society Data Collection
- Political Polarity Data
- StackExchange Data Explorer
- Terrorism Research and Analysis Consortium
- Texas Inmates Executed Since 1984
- Titanic Survival Data Set or on Kaggle
- UCB's Archive of Social Science Data (D-Lab)
- UCLA Social Sciences Data Archive
- UN Civil Society Database
- Universities Worldwide
- <u>UPJOHN for Labor Employment Research</u>
- <u>Uppsala Conflict Data Program</u>
- World Bank Open Data
- WorldPop project Worldwide human population distributions

Software

FLOSSmole data about free, libre, and open source software development

Sports

- Basketball (NBA/NCAA/Euro) Player Database and Statistics
- Betfair Historical Exchange Data
- Cricsheet Matches (cricket)
- Ergast Formula 1, from 1950 up to date (API)
- Football/Soccer resources (data and APIs)
- Lahman's Baseball Database
- Pinhooker: Thoroughbred Bloodstock Sale Data
- Retrosheet Baseball Statistics
- <u>Tennis database of rankings, results, and stats for ATP, WTA, Grand Slams</u> and <u>Match Charting Project</u>

Time Series

- Databanks International Cross National Time Series Data Archive
- Hard Drive Failure Rates
- Heart Rate Time Series from MIT
- Time Series Data Library (TSDL) from MU
- UC Riverside Time Series Dataset

Transportation

- Airlines OD Data 1987-2008
- Bay Area Bike Share Data
- Bike Share Systems (BSS) collection
- GeoLife GPS Trajectory from Microsoft Research
- German train system by Deutsche Bahn
- Hubway Million Rides in MA
- Marine Traffic ship tracks, port calls and more
- Montreal BIXI Bike Share
- NYC Taxi Trip Data 2009-
- NYC Taxi Trip Data 2013 (FOIA/FOILed)
- NYC Uber trip data April 2014 to September 2014
- Open Traffic collection
- OpenFlights airport, airline and route data
- Philadelphia Bike Share Stations (JSON)
- Plane Crash Database, since 1920
- RITA Airline On-Time Performance data
- RITA/BTS transport data collection (TranStat)
- Toronto Bike Share Stations (XML file)
- Transport for London (TFL)
- Travel Tracker Survey (TTS) for Chicago
- U.S. Bureau of Transportation Statistics (BTS)
- U.S. Domestic Flights 1990 to 2009
- U.S. Freight Analysis Framework since 2007

Complementary Collections

- Data Packaged Core Datasets
- Database of Scientific Code Contributions
- A growing collection of public datasets: <u>CoolDatasets.</u>

- DataWrangling: <u>Some Datasets Available on the Web</u>
- Inside-r: Finding Data on the Internet
- OpenDataMonitor: An overview of available open data resources in Europe
- Quora: Where can I find large datasets open to the public?
- RS.io: <u>100+ Interesting Data Sets for Statistics</u>
- StaTrek: <u>Leveraging open data to understand urban lives</u>

Awesome TensorFlow

https://github.com/jtoy/awesome-tensorflow

A curated list of awesome TensorFlow experiments, libraries, and projects. Inspired by awesome-machine-learning.

What is TensorFlow?

TensorFlow is an open source software library for numerical computation using data flow graphs. In other words, the best way to build deep learning models.

More info here.

Table of Contents

- Tutorials
- Models/Projects
- Powered by TensorFlow
- Libraries
- <u>Videos</u>

- <u>Papers</u>
- Blog posts
- Community
- Books

Tutorials

- <u>TensorFlow Tutorial 1</u> From the basics to slightly more interesting applications of TensorFlow
- <u>TensorFlow Tutorial 2</u> Introduction to deep learning based on Google's TensorFlow framework. These tutorials are direct ports of Newmu's Theano
- TensorFlow Examples TensorFlow tutorials and code examples for beginners
- <u>Sungioon's TensorFlow-101</u> TensorFlow tutorials written in Python with Jupyter Notebook
- <u>Terry Um's TensorFlow Exercises</u> Re-create the codes from other TensorFlow examples
- <u>Installing TensorFlow on Raspberry Pi 3</u> TensorFlow compiled and running properly on the Raspberry Pi
- <u>Classification on time series</u> Recurrent Neural Network classification in TensorFlow with LSTM on cellphone sensor data
- Getting Started with TensorFlow on Android Build your first TensorFlow Android app
- <u>Predict time series</u> Learn to use a seq2seq model on simple datasets as an introduction to the vast array of possibilities that this architecture offers

- <u>Single Image Random Dot Stereograms</u> SIRDS is a means to present 3D data in a 2D image. It allows for scientific data display of a waterfall type plot with no hidden lines due to perspective.
- <u>CS20 SI: TensorFlow for DeepLearning Research</u> Stanford Course about Tensorflow from 2017 <u>Syllabus</u> <u>Unofficial Videos</u>
- <u>TensorFlow World</u> Concise and ready-to-use TensorFlow tutorials with detailed documentation are provided.
- <u>Effective Tensorflow</u> Tensorflow howtos and best practices. Covers the basics as well as advanced topics.

Models/Projects

- <u>Domain Transfer Network</u> Implementation of Unsupervised Cross-Domain Image Generation
- Show, Attend and Tell Attention Based Image Caption Generator
- Neural Style Implementation of Neural Style
- Pretty Tensor Pretty Tensor provides a high level builder API
- Neural Style An implementation of neural style
- <u>AlexNet3D</u> An implementations of AlexNet3D. Simple AlexNet model but with 3D convolutional layers (conv3d).
- <u>TensorFlow White Paper Notes</u> Annotated notes and summaries of the TensorFlow white paper, along with SVG figures and links to documentation
- NeuralArt Implementation of A Neural Algorithm of Artistic Style
- Deep-Q learning Pong with TensorFlow and PyGame
- Generative Handwriting Demo using TensorFlow An attempt to implement the random handwriting generation portion of Alex Graves' paper
- <u>Neural Turing Machine in TensorFlow</u> implementation of Neural Turing Machine
- <u>GoogleNet Convolutional Neural Network Groups Movie Scenes By Setting</u> Search, filter, and describe videos based on objects, places, and other things that appear in them
- Neural machine translation between the writings of Shakespeare and modern English using TensorFlow - This performs a monolingual translation, going from modern English to Shakespeare and vice-versa.

- Chatbot Implementation of "A neural conversational model"
- <u>Colornet Neural Network to colorize grayscale images</u> Neural Network to colorize grayscale images
- Neural Caption Generator Implementation of "Show and Tell"
- <u>Neural Caption Generator with Attention</u> Implementation of <u>"Show, Attend</u> and Tell"
- <u>Weakly detector</u> Implementation of <u>"Learning Deep Features for</u> Discriminative Localization"
- <u>Dynamic Capacity Networks</u> Implementation of "<u>Dynamic Capacity Networks</u>"
- <u>HMM in TensorFlow</u> Implementation of viterbi and forward/backward algorithms for HMM
- <u>DeepOSM</u> Train TensorFlow neural nets with OpenStreetMap features and satellite imagery.
- <u>DQN-tensorflow</u> TensorFlow implementation of DeepMind's 'Human-Level Control through Deep Reinforcement Learning' with OpenAl Gym by Devsisters.com
- <u>Highway Network</u> TensorFlow implementation of <u>"Training Very Deep Networks"</u> with a <u>blog post</u>
- <u>Sentence Classification with CNN</u> TensorFlow implementation of <u>"Convolutional Neural Networks for Sentence Classification"</u> with a <u>blog post</u>
- <u>End-To-End Memory Networks</u> Implementation of <u>End-To-End Memory Networks</u>
- <u>Character-Aware Neural Language Models</u> TensorFlow implementation of <u>Character-Aware Neural Language Models</u>
- <u>YOLO TensorFlow ++</u> TensorFlow implementation of 'YOLO: Real-Time Object Detection', with training and an actual support for real-time running on mobile devices.
- <u>Wavenet</u> This is a TensorFlow implementation of the <u>WaveNet generative</u> neural network architecture for audio generation.
- <u>Mnemonic Descent Method</u> Tensorflow implementation of <u>"Mnemonic Descent Method: A recurrent process applied for end-to-end face alignment"
 </u>
- <u>CNN visualization using Tensorflow</u> Tensorflow implementation of <u>"Visualizing and Understanding Convolutional Networks"</u>

- <u>VGAN Tensorflow</u> Tensorflow implementation for MIT <u>"Generating Videos</u> with Scene Dynamics" by Vondrick et al.
- <u>3D Convolutional Neural Networks in TensorFlow</u> Implementation of <u>"3D Convolutional Neural Networks for Speaker Verification application"</u> in TensorFlow by Torfi et al.
- <u>Lip Reading Cross Audio-Visual Recognition using 3D Architectures in TensorFlow</u> TensorFlow Implementation of <u>"Cross Audio-Visual Recognition in the Wild Using Deep Learning"</u> by Torfi et al.
- Attentive Object Tracking Implementation of <u>"Hierarchical Attentive Recurrent Tracking"</u>
- Holographic Embeddings for Graph Completion and Link Prediction -Implementation of Holographic Embeddings of Knowledge Graphs

Powered by TensorFlow

- YOLO TensorFlow Implementation of 'YOLO : Real-Time Object Detection'
- <u>android-yolo</u> Real-time object detection on Android using the YOLO network, powered by TensorFlow.
- <u>Magenta</u> Research project to advance the state of the art in machine intelligence for music and art generation

Libraries

- <u>tf.contrib.learn</u> Simplified interface for Deep/Machine Learning (now part of TensorFlow)
- <u>tensorflow.rb</u> TensorFlow native interface for ruby using SWIG
- <u>tflearn</u> Deep learning library featuring a higher-level API
- <u>TensorFlow-Slim</u> High-level library for defining models
- <u>TensorFrames</u> TensorFlow binding for Apache Spark
- <u>TensorFlowOnSpark</u> initiative from Yahoo! to enable distributed TensorFlow with Apache Spark.
- caffe-tensorflow Convert Caffe models to TensorFlow format
- keras Minimal, modular deep learning library for TensorFlow and Theano
- <u>SyntaxNet: Neural Models of Syntax</u> A TensorFlow implementation of the models described in <u>Globally Normalized Transition-Based Neural Networks</u>, <u>Andor et al.</u> (2016)
- keras-js Run Keras models (tensorflow backend) in the browser, with GPU support
- <u>NNFlow</u> Simple framework allowing to read-in ROOT NTuples by converting them to a Numpy array and then use them in Google Tensorflow.

- <u>Sonnet</u> Sonnet is DeepMind's library built on top of TensorFlow for building complex neural networks.
- <u>tensorpack</u> Neural Network Toolbox on TensorFlow focusing on training speed and on large datasets.

Videos

- <u>TensorFlow Guide 1</u> A guide to installation and use
- TensorFlow Guide 2 Continuation of first video
- <u>TensorFlow Basic Usage</u> A guide going over basic usage
- <u>TensorFlow Deep MNIST for Experts</u> Goes over Deep MNIST
- <u>TensorFlow Udacity Deep Learning</u> Basic steps to install TensorFlow for free on the Cloud 9 online service with 1Gb of data
- Why Google wants everyone to have access to TensorFlow
- Videos from TensorFlow Silicon Valley Meet Up 1/19/2016
- Videos from TensorFlow Silicon Valley Meet Up 1/21/2016
- Stanford CS224d Lecture 7 Introduction to TensorFlow, 19th Apr 2016 CS224d
 Deep Learning for Natural Language Processing by Richard Socher
- <u>Diving into Machine Learning through TensorFlow</u> Pycon 2016 Portland Oregon, <u>Slide</u> & <u>Code</u> by Julia Ferraioli, Amy Unruh, Eli Bixby
- <u>Large Scale Deep Learning with TensorFlow</u> Spark Summit 2016 Keynote by Jeff Dean
- <u>Tensorflow and deep learning without at PhD</u> by Martin Görner
- <u>Tensorflow and deep learning without at PhD, Part 2 (Google Cloud Next '17)</u> by Martin Görner

Papers

- <u>TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems</u> This paper describes the TensorFlow interface and an implementation of that interface
 that we have built at Google
- <u>TF.Learn: TensorFlow's High-level Module for Distributed Machine Learning</u>
- <u>Comparative Study of Deep Learning Software Frameworks</u> The study is performed on several types of deep learning architectures and we evaluate the performance of the above frameworks when employed on a single machine for both (multi-threaded) CPU and GPU (Nvidia Titan X) settings
- <u>Distributed TensorFlow with MPI</u> In this paper, we extend recently proposed Google TensorFlow for execution on large scale clusters using Message Passing Interface (MPI)
- Globally Normalized Transition-Based Neural Networks This paper describes the models behind SyntaxNet.
- <u>TensorFlow: A system for large-scale machine learning</u> This paper describes the TensorFlow dataflow model in contrast to existing systems and demonstrate the compelling performance

Official announcements

- <u>TensorFlow: smarter machine learning, for everyone</u> An introduction to TensorFlow
- <u>Announcing SyntaxNet: The World's Most Accurate Parser Goes Open Source</u> Release of SyntaxNet, "an open-source neural network framework implemented in TensorFlow that provides a foundation for Natural Language Understanding systems.

Blog posts

- Why TensorFlow will change the Game for Al
- <u>TensorFlow for Poets</u> Goes over the implementation of TensorFlow
- <u>Introduction to Scikit Flow Simplified Interface to TensorFlow</u> Key Features Illustrated
- <u>Building Machine Learning Estimator in TensorFlow</u> Understanding the Internals of TensorFlow Learn Estimators
- <u>TensorFlow Not Just For Deep Learning</u>
- The indico Machine Learning Team's take on TensorFlow
- The Good, Bad, & Ugly of TensorFlow A survey of six months rapid evolution (+ tips/hacks and code to fix the ugly stuff), Dan Kuster at Indico, May 9, 2016
- <u>Fizz Buzz in TensorFlow</u> A joke by Joel Grus
- RNNs In TensorFlow, A Practical Guide And Undocumented Features Step-by-step guide with full code examples on GitHub.
- <u>Using TensorBoard to Visualize Image Classification Retraining in TensorFlow</u>
- <u>TFRecords Guide</u> semantic segmentation and handling the TFRecord file format.
- <u>TensorFlow Android Guide</u> Android TensorFlow Machine Learning Example.
- <u>TensorFlow Optimizations on Modern Intel® Architecture</u> Introduces TensorFlow optimizations on Intel® Xeon® and Intel® Xeon Phi™ processor-based platforms based on an Intel/Google collaboration.

Books

- <u>Machine Learning with TensorFlow</u> by Nishant Shukla, computer vision researcher at UCLA and author of Haskell Data Analysis Cookbook. This book makes the math-heavy topic of ML approachable and practicle to a newcomer.
- <u>First Contact with TensorFlow</u> by Jordi Torres, professor at UPC Barcelona Tech and a research manager and senior advisor at Barcelona Supercomputing Center
- <u>Deep Learning with Python</u> Develop Deep Learning Models on Theano and TensorFlow Using Keras by Jason Brownlee
- <u>TensorFlow for Machine Intelligence</u> Complete guide to use TensorFlow from the basics of graph computing, to deep learning models to using it in production environments - Bleeding Edge Press

- <u>Getting Started with TensorFlow</u> Get up and running with the latest numerical computing library by Google and dive deeper into your data, by Giancarlo Zaccone
- Hands-On Machine Learning with Scikit-Learn and TensorFlow by Aurélien Geron, former lead of the YouTube video classification team. Covers ML fundamentals, training and deploying deep nets across multiple servers and GPUs using TensorFlow, the latest CNN, RNN and Autoencoder architectures, and Reinforcement Learning (Deep Q).
- <u>Building Machine Learning Projects with Tensorflow</u> by Rodolfo Bonnin. This book covers various projects in TensorFlow that expose what can be done with TensorFlow in different scenarios. The book provides projects on training models, machine learning, deep learning, and working with various neural networks. Each project is an engaging and insightful exercise that will teach you how to use TensorFlow and show you how layers of data can be explored by working with Tensors.

Awesome Mac

https://github.com/jaywcjlove/awesome-mac

This repo is a collection of **AWESOME** Mac applications and tools for developers and designers. Feel free to **Star** and **Fork**. Any comments, suggestions? <u>Let us know</u>. we love PRs :), please follow the <u>awesome</u> format.

Explanation

English | 中文

means open source, click to enter open source repo;

means **free** to use, or **free** personal license;

means hot app;

means recommended app;

means must have app;

means App store hyperlink;

means highly recommended, must-have app. The number of stars represents how strongly I recommend it;

You may come across some non-English characters. Those apps don't have an English version yet, so you might want steer clear of them.

Contents

- Editors and IDE
- Development Tools
- Test Tools
- Command Line Tools
- Version Control Systems
- Databases
- Design and Product
- Virtual Machines
- Communication
- Data Recovery Tools
- Audio and Video Tools
- Reading and Writing Tools

- Ebooks
- FTP Clients
- Frameworks For Hybrid Applications
- Download Management Tools
- Online Storage
- Input Methods
- Web Browsers
- Translation Tools
- Proxy and VPN Tools
- <u>Utilities</u>
- Remote Login Software
- QuickLook Plugins

Editors and IDE

applications to edit text, I suggest the open-source editors

Text Editors

- Atom A hackable text editor for the 21st century made by GitHub. Atom Plugins.
- <u>Sublime Text</u> A sophisticated text editor for code, markup and prose. You'll love the slick user interface, extraordinary features and amazing performance, <u>Sublime Text</u> <u>Plugins</u>.
- <u>Brackets</u> A modern, open-source text editor that understands web design by Adobe.
- <u>Espresso</u> The web editor for Mac is back. For people who make delightful, innovative and fast websites.
- <u>Visual Studio Code</u> Microsoft's free & open-source editor, TypeScript friendly, <u>VSCode Plugins</u>
- <u>Emacs</u> A popular text editor used mainly on Unix-based systems by programmers, scientists, engineers, students, and system administrators
- <u>LightTable</u> The next generation code editor.
- <u>TextMate</u> An editor that brings Apple's approach to operating systems into the world of text editors.
- <u>BBEdit</u> The leading professional HTML and text editor for Macintosh.
- Vim A highly configurable text editor built to make creating and changing any kind of text very efficient, Vim Plugins.
- Vimr Refined Vim Experience for OS X.
- ONI An IDE powered by Neovim.
- micro a modern and intuitive terminal-based text editor.
 [(https://github.com/ory/editor)]
- <u>HBuilder</u> An IDE for web development (Support HTML5), built by DCloud. (Missing English UI and Docs)
- Tincta A text editor for Mac OS X.
- <u>CotEditor</u> Lightweight plain-text editor for macOS.
- Chocolat Native text editor.

IDEs

- Xcode The essential IDE for iOS/macOS development.
- IntelliJ IDEA A powerful Java IDE. (Free for Students)
- <u>Coda2</u> A fast, clean and powerful text editor.

- <u>Eclipse</u> Popular open-source IDE, mainly for Java but with plugin support for a wide array of languages and platforms.
- WebStorm The smartest JavaScript IDE by JetBrains. FREE for Students,
 check here for more info.
 - NodeJS Node.js integration. You definitely need this, quite a few features require it.
 - o <u>EditorConfig</u> A JetBrains IDE plugin supporting the EditorConfig standard.
 - o Material Theme UI Provides 3 modes, nice and clean.
- <u>Deco IDE</u> The best IDE for building React Native apps.
- Xamarin Studio Free cross platform C# IDE. Xamarin Studio supports iOS, Android and .Net development
- <u>NetBeans IDE</u> A free and open-source IDE, mainly used for Java development, but supports many other languages and frameworks.
- Android Studio The official IDE for Android, based on Intellij IDEA.

Development Tools

- <u>WeFlow</u> Development tools for front-end workflows based on <u>tmt-workflow</u>.
 - (Missing English UI and Docs)
- <u>Koala</u> A GUI application for Less, Sass, Compass and CoffeeScript compilation.
- <u>CodeKit</u> Web development tool which can automatically compile Less, Sass, CoffeeScript, TypeScript, Jade and JavaScript, auto-refresh browsers and much more.
- <u>PaintCode</u> PaintCode is a unique vector drawing app that generates Objective-C or Swift code in real time, acting as a bridge between developers and graphic designers.
- <u>Hosts.prefpane</u> A system preference pane to manage your hosts file.
- iHosts The only /etc/hosts editor on Mac App Store.
- SwitchHosts A free and open-source app for hosts management & switching.
- Gas Mask A simple hosts file manager for Mac OS X.
- <u>DiffMerge</u> An application to visually compare and merge files.
- <u>Gemini</u> An intelligent duplicate file finder.
- <u>BetterRename</u> The most powerful and complete Mac file renaming application on the market.
- PPRows An application which can calculate how many lines of code you write.

- Beyond Compare Compare files and folders using simple, powerful commands that focus on the differences you are interested in and ignore those you are not.
- <u>Kaleidoscope</u> A powerful compare tool for text, images and folders. It works perfectly with git, svn or other version control tools.
- <u>Fanvas</u> Convert swf to html canvas animation. (**Missing English Docs**)
- EnvPane An OS X preference pane for environment variables.
- <u>Dash</u> Awesome API documentation browser and code snippet manager.
- <u>DLite</u> The simplest way to use Docker on OS X.
- <u>SnippetsLab</u> Easy-to-use code snippets manager.
- <u>StarUML</u> Powerful UML app.
- <u>Vagrant Manager</u> Manage your vagrant machines in one place.
- <u>zeplin</u> A collaboration tool for work between designers and developers.
- Go2Shell Open terminal from Finder.
- SecureCRT A terminal emulation which supports SSH, Telnet or other protocols.
- <u>Finicky</u> An application that allows you to set up rules that decide which browser is opened for every link that would open the default browser.
- MJML A framework that utilizes a semantic syntax and a rich standard components library, which allow users to easily create responsive emails.
- <u>TeXstudio</u> An integrated writing environment for creating LaTeX documents.
- <u>Vagrant</u> A tool for building and distributing development environments.
- <u>FinderGo</u> M Open terminal quickly from Finder
- Localname Provide access to your local development server
- <u>LaunchRocket</u> Brew services management in Mac System Preferences.

Regular Expression Editors

- <u>Patterns</u> A regular expression editor.
- Reginald A regular expression test application. RegexKitLite.
- Regex A regular expression testing tool with an emphasis on simplicity.

- Reggy An open-source regular expression editor.
- RegExRX A development tool for regular expressions.

Test Tools

- Wireshark The world's foremost and widely-used network protocol analyzer.
- <u>Charles</u> An HTTP proxy / HTTP monitor / Reverse Proxy that enables a developer to view all of the HTTP and SSL / HTTPS traffic between their machine and the Internet.
- <u>James</u> Open-source proxy tool for checking and mapping requests with http as well
 as https.
- mitmproxy An interactive intercepting HTTP proxy for penetration testers and software developers
- Insomnia 3.0 A beautiful HTTP testing tool.
- <u>Cocoa Rest Client</u> A free open-source, native Apple OS X app for testing HTTP/REST endpoints.
- Paw An advanced HTTP client.
- <u>Cellist</u> HTTP debugging proxy for OS X.
- Integrity A free website link checker for Mac.
- <u>Postman</u> A powerful GUI platform to make your API development faster & easier, from building API requests through testing, documentation and sharing.

Command Line Tools

- Mac OS X Manual Pages Manual pages recommended reference for a number of BSD and POSIX functions and tools.
- <u>iTerm2</u> iTerm2 is an amazing terminal emulator for OS
 X.
- <u>tmux</u> A "terminal multiplexer", it enables a number of terminals (or windows) to be accessed and controlled from a single terminal. tmux is intended to be a simple, modern, BSD-licensed alternative to programs such as GNU screen.
- mas A simple command line interface for the Mac App Store.

- <u>cool-retro-term</u> A good looking terminal emulator which mimics the old cathode display.
- Oh my zsh An open-source, community-driven framework for managing your ZSH configuration. It comes bundled with a ton of helpful functions, helpers, plugins, themes.
- <u>autojump</u> A cd command that learns easily navigate directories from the command line
- <u>Prezto Instantly Awesome Zsh</u>

 The configuration framework for Zsh; it enriches the command line interface environment with sane defaults, aliases, functions, auto completion, and prompt themes.
- Glances Glances is a cross-platform curses-based system monitoring tool.
- <u>Cakebrew</u> A GUI client for <u>Homebrew</u>. Install, check or remove apps, no commandline needed.
- ndm Manage npm straight from the couch.
- <u>Black Screen</u> A terminal emulator for the 21st century.
- Fish Shell A smart and user-friendly terminal, which is similar with zsh
- <u>oh-my-fish</u> Alternative to oh-my-zsh, on fishShell framework
- <u>bash-it</u> Shameless ripoff of oh-my-zsh for bash.
- <u>HyperTerm</u> A terminal built on web technologies.
- <u>itunes-remote</u> Software for controlling iTunes via the terminal.
- mycli A command line client for MySQL that can do auto-completion and syntax highlighting.
- Mac-CLI The ultimate tool to manage your Mac. automatize the usage of your OS X system.
- <u>m-cli</u> Swiss Army Knife for macOS.
- <u>lnav</u> A log file navigator.
- <u>cmus</u> A small, fast and powerful console music player for Unix-like operating systems.
- Serial Full-featured serial terminal for the Mac.

Version Control Systems

- Git A distributed version control tool, the official website provides dozens of GUI clients for Mac.
- SVN An open-source version control system.

Version Control GUIs

- GitFinder A fast and lightweight Git client for Mac with Finder integration.
- Fork A fast and friendly Git client for Mac.
- <u>Gitbar</u> Open-source, display Github contribution statistics on your menu bar.
- GitHub Desktop The GitHub official GUI.
- OhMyStar Beautiful and efficient way to manage, explore and share your Github Stars.
- <u>GitUp</u> A simple and powerful Git client.
- <u>Hub</u> A command-line wrapper for Git that makes you better at GitHub.
- SourceTree A free Git & Mercurial client for Windows or Mac.
- Tower2 The most powerful Git client for Mac and Windows.
- Versions A Mac Subversion Client (SVN).
- <u>Cornerstone</u> Powerful version control with a gorgeous interface.
- SmartGit A Git client with support.
- GitKraken The most popular Git GUI for Windows, Mac AND Linux.

Version Control Managers

- <u>GitLab</u> Open-source Git repo management software.
- <u>Github</u> Code hosting, project management and deployment for software projects.
- <u>Coding.net</u> A one-stop cloud platform for developers, free repository hosting,
 project collaboration, code quality assurance and Page service.
- <u>phabricator</u> Open software engineering platform and fun adventure game.
- Gogs Gogs (Go Git Service) is a painless self-hosted Git service.

- Gerrit Provides web based code review and repository management for the Git version control system.
- <u>Gitblit</u> Pure java Git solution for managing, viewing, and serving Git repositories.

Databases

- Sequel Pro MySQL database management for Mac OS X.
- MySQL Workbench The official MySQL GUI.
- <u>Navicat Data Modeler</u> Is a powerful and cost-effective database design tool which helps you build high-quality conceptual, logical and physical data models.
- <u>SQLPro Studio</u> A simple, powerful database manager for macOS.
- <u>Bdash</u> A modern SQL client application, supports MySQL, PostgreSQL (Redshift) and
 BigQuery.
- <u>Postico</u> A Modern PostgreSQL client for Mac.
- <u>ElectroCRUD</u> A modern MySQL CRUD application.
- <u>Base 2</u> An application for creating, designing, editing and browsing SQLite 3 database files.
- <u>SQLight</u> Is an SQLite database manager tool.
- <u>DB Browser for SQLite</u> The Official home of the DB Browser for SQLite.
- <u>Postgres.app</u> The easiest way to get started with PostgreSQL on the Mac.
- PSequel PostgreSQL GUI tool for Mac OS X.
- Robo 3T Native and cross-platform MongoDB management tool (Admin UI).
- MongoBooster Shell-centric and cross-platform GUI tool for MongoDB. It provides update-in-place, fluent query builder, ES6 syntax support and True IntelliSense experience.
- Mongo Management Studio A simple MongoDB GUI.
- MongoChef The GUI for MongoDB.
- Mongotron A Mongo DB GUI built using Electron, and Angular Js.
- Chrome MySQL Admin A powerful Chrome app to manage your MySQL.

- <u>JackDB</u> A secure, collaborative environment for your queries and data-driven insights.
- Medis GUI Manager for Redis.
- RedisDesktopManager Cross-platform GUI management tool for Redis.
- MDB Explorer MDB tool to open, read, export your MDB files to other formats and databases.
- <u>Datum SQLite</u> A full featured, modern, and fast SQLite manager.
- mongoDB.app The easiest way to get started with mongoDB on the Mac.
- <u>DataGrip</u> Database management software by JetBrains. **FREE** for Students, check here for more info
- Tableau Public Free data-visualization software.
- <u>Core Data Editor</u> Core Data Editor lets you easily view, edit and analyze applications' data.
- <u>Keylord</u> Keylord is a desktop GUI client for Redis, Bolt, LevelDB and Memcached keyvalue databases.
- DBeaver A Universal SQL Client

Design and Product

Design Tools

- Acorn A great Mac OS X picture and photo editor, built for humans.
- Affinity Designer Professional graphic design software for Mac.
- Affinity Photo Professional image editing software for Mac.
- <u>Blender</u> Free and open 3D creation software.
- <u>Pixelmator</u> Full-featured image editor for Mac.
- Sketch Professional digital design for mac.
 - Sketch Toolbox A super simple plugin manager for Sketch.
 - Measure Make it a fun to create spec for developers and teammates.
 - <u>User Flows</u> generating flow diagrams from Artboards.
- Sketch Cache Cleaner -Deletes hidden Sketch history files that can take a lot of space on your hard drive and that you would probably never use.

- <u>Gravit Designer</u> Gravit Designer is a full featured free vector design app right at your fingertips.
- inklet Turn your Mac trackpad into drawing board.
- Alchemy An experimental, open-source drawing application with an emphasis on creating conceptual art.
- MyPaint Is a simple drawing and painting program that works well with Wacom-style graphics tablets.
- GIMP The GNU Image Manipulation Program.
- Inkscape is a professional vector graphics editor.
- Monodraw Powerful ASCII art editor designed for the Mac.
- <u>SketchBook</u> Drawing software for concept design, comic art, and digital sketching.
- <u>Tayasui Sketches</u> Professional drawing software.
- Art Text 3 This is graphic design software specifically tuned for lettering, typography,
 text mockups and various artistic text effects.
- FontForge Free (libre) font editor.
- Nik Collection Google Nik collection.
- Paintbrush A bitmap image editor.
- <u>Krita</u> Open-source digital painting software for concept artists, digital painters, and illustrators.
- <u>Vectr</u> Free graphics editor used to create vector graphics easily and intuitively.
- Principle An application for designing animated and interactive user interfaces.
- <u>MagicaVoxel</u> A free lightweight 8-bit voxel editor and interactive path tracing renderer.
- <u>ScreenToLayers</u> Easily export your screen into a layered PSD file.
- <u>Pixel Perfect</u> A tool to overlay design images over implementation and make it pixel perfect.
- Sparkle Pro visual web design.

Prototyping And Mind-mapping Tools

 Origami Studio - A new tool for designing modern interfaces, built and used by designers at Facebook.

- FLINTO Quickly create interactive prototypes of mobile, desktop, or web apps.
- Kite A powerful animation and prototyping application for Mac & iOS.
- <u>Justinmind</u> Prototyping platform for web and mobile apps.
- MockFlow Online prototyping suite for web-design and usability testing.
- Axure RP 8 Prototypes, specifications and diagrams in one tool.
- <u>pencil</u> A free, open-source tool for making diagrams and GUI prototyping.
- Mockplus Prototype faster, smarter and easier.
- OmniGraffle Diagramming and graphic design for Mac, iPhone, and iPad.
- XMind The most popular mind-mapping tool on the planet.
- <u>Lighten</u> The best way to clarify thinking, boost productivity, brainstorm, and visualize concepts.
- <u>Scapple</u> Practical mind-mapping software with free whiteboard-like layout.
- <u>Framer</u> A tool for interactive prototyping.
- <u>Balsamiq Mockups</u> A wire-framing tool that helps you work faster and smarter.
- Marvel Simple design, prototyping and collaboration.
- <u>MindNode</u> Mind-mapping software with an emphasis on simplicity and ease-of-use.
- WriteMapper Get from idea to final draft in no time.
- <u>SimpleMind</u> Is the world leader in cross platform Mind Mapping tools.
- Adobe XD (Experience Design) The first all-in-one cross-platform tool for designing and prototyping websites and mobile apps.

Other Tools

- <u>TinyPNG4Mac</u> Open-source tool to compress images.
- <u>Image2icon</u> Create and personalize icons from your pictures.
- <u>ImageAlpha</u> Compress images with PNG format and remove transparency.
- Resize Master Batch resize and watermark your images fast and easy.
- JPEGmini Reduce image size by up to 80%, without compromising quality.
- <u>ImageOptim</u> Compress images and remove EXIF information.
- Sip The best way to collect, organize & share your colors.
- Mark Man Measure & Spec Fast.

- Frank DeLoupe A color-picking tool, supports Retina.
- <u>ColorSchemer</u> A professional color matching application for your Mac.
- <u>lconKit</u> An App icon generator.
- Snagit Simple, Powerful Screen Capture Software and Screen Recorder.
- GifCapture GIF capture app for macOS.
- Gifox Gif Recording and Sharing.
- <u>APNGb</u> A png image assembler/disassembler app.
- <u>LICEcap</u> Record your screen and export to GIF. You can change the recording area anytime during recording.
- <u>Kap</u> An open-source screen-recorder built with web technology.
- GIPHY Capture Capture and share your screen as a GIF.
- Skitch A screen capture application with a powerful annotation capabilities.
- Teampaper Snap Let your screenshots speak up.
- Monosnap Make screenshots. Draw on it. Shoot video and share your files. It's fast, easy and free.
- <u>Jietu</u> Screenshot with a powerful annotation capabilities, by Tencent. (**Missing English UI and Docs**)
- <u>ScreenShot PSD</u> Capture the screen as a layered PSD for easy editing.
- Snip An application for sharing captured images on QQ Mail.
- <u>iPic</u> Easily upload images with Markdown supported.
- <u>lconjar</u> lcon management tool to organize or search your icons.
- <u>svgus</u> Organize, clean and transform your SVGs.
- <u>RightFont</u> Preview, sync, install and manage fonts on Mac, Dropbox or Google
 Drive.
- <u>Solarized</u> Clean and beautiful color theme. Works well with iTerm, JetBrains products,
 Vim etc.

Virtual Machines

- Parallels Powerful, easy-to-use VM. No free upgrade for each new Mac OS.
- Virtual Box A powerful x86 and AMD64/Intel64 virtualization product.
- VMWare Fusion A powerful, commercial VM developed by VMware.
- <u>Veertu</u> The lightest VM on Mac. A responsive, sandboxed & native way to run VM on your Mac.

Communication

Team communication and collaboration tools

- <u>Franz</u> An <u>Electron</u> based, multi-protocol wrapper for web-based chat client. One application, 23 messenger services.
- QQ Official QQ app for Mac. (Missing English Docs)
- WeChat Official WeChat app for Mac.
- <u>Electronic WeChat</u> An open-source WeChat client, build with <u>Electron</u>.
- <u>Skype</u> A cross-platform application that provides video chat and voice call services. Users can exchange images, text, video and any other digital documents.
- Maipo A third-party Weibo client for Mac OS. (Missing English UI and Docs)
- 御飯 A third-party FanFou client for Mac OS. (Missing English UI and Docs)
- ChitChat Unofficial WhatsApp.
- Telegram A messaging app with a focus on speed and security.
- Messenger For Mac A third-party Facebook messenger for Mac.
- Adium A free instant messaging application for Mac OS X, connecting to AIM, MSN,
 SMPP, Yahoo and more.
- <u>Textual</u> Textual is the world's most popular application for interacting with Internet Relay Chat (IRC) chatrooms on OS X.
- Gitter An instant messaging and chat room system for developers as well as GitHub users. Developer friendly with Markdown syntax support.
- <u>简聊</u> A communication tool for enterprise use. You can build your own service with its open-source version. (**Missing English Docs**)
- <u>钉钉</u> A free office communication platform for enterprise use. (**Missing English UI** and **Docs**)

- LimeChat An open-source IRC client for Mac OS X.
- Slack Awesome tool for team collaboration and communication.
- 零信 Work anywhere at anytime. Cross platform. (Missing English UI and Docs)
- 今目标 An internet work platform for small-to-medium sized enterprises. (Missing English UI and Docs)
- <u>BearyChat</u> Team collaboration and communication tool for internet-based teams. (**Missing English UI and Docs**)
- <u>Bitpost</u> A private decentralized messaging system. Instead of connecting to centralized servers(like Facebook, Gmail etc.) or federated servers(like email, IRC, Jabber), it uses P2P protocol to send encrypted messages to another person or to many subscribers.
- <u>Teambition</u> A team collaboration tool, including many features like task plan, schedule, file sharing, instant discussion and everything you need when collaborating with other team members.
- WeeChat The extensible command-line chat client.
- Rambox messaging and emailing app that combines common web applications into one.

Email Clients

- Spark A fast email client. For both Mac OS and iOS.
- Airmail A fast email client. For both Mac OS and iOS.
- Nylas Mail An extensible desktop mail app built on the modern web.
- Foxmail A fast email client.
- Mail Master Netease full platform of the mailbox management client.
- MailTags Use tags to organize email and schedule
- N1 An extensible, open-source mail app, free for developers and \$7/month for Pro.
- Postbox A powerful, simple and beautiful email client, need to pay for a license.
- <u>Polymail</u> Simple, beautiful and powerful email client.
- <u>Newton(formerly Cloudmagic)</u> An excellent email client with concise interface, scheduling and read receipts.
- <u>ThunderBird</u> Software made to make email easier.
- Yomail Maybe the best email client for email.

Data Recovery Tools

- <u>DiskWarrior</u> The world's most advanced repair and recovery tool for Mac.
- <u>Data Rescue</u> A comprehensive and professional data recovery tool for most cases.
- <u>Stellar Phoenix Mac Data Recovery</u> A powerful recovery tool for Mac file recovery, Time machine recovery, Encrypted Disk recovery and much more.
- R-Studio for Mac A powerful tool which recovers data on disks, even if their partitions are formatted, damaged or deleted.

Audio and Video Tools

- <u>Popcorn Time</u> Watch torrent movies instantly, This Popcorn Time service will never be taken down. Download and enjoy.
- <u>Kodi</u> An award-winning free and open-source (GPL) software media center for playing videos, music, pictures, games, and more.
- <u>mpv</u> A free, open-source, and cross-platform media player.
- <u>IINA</u> The modern video player for macOS, Based on mpv, the powerful media player project.
- VOX Player #1 high-definition audio player for Mac and iPhone. Music just sounds better!
- Radiant Player Third party Google Play Music client.
- Sonora A minimal, beautifully designed music player.
- <u>Audacity</u> A free, open-source, cross-platform audio software for multi-track recording and editing.
- <u>Natron</u> Open-source compositing software. Node-graph based. Similar in functionalities to Adobe After Effects and Nuke by The Foundry.
- Ardour Cross-platform audio software for multi-track recording and editing.
- <u>Hydrogen</u> Professional yet simple and intuitive pattern-based drum programming for GNU/Linux.
- <u>Audio Hijack</u> Record any application's audio, including VoIP calls from Skype, web streams from Safari, and much more.
- <u>Stringed 2</u> Music practice software designed to help users learn how to play their favorite songs.
- Mixxx The most advanced free DJ software.

- <u>MuseScore</u> Free, open-source music notation software.
- Cog A free, open-source audio player.
- <u>Elmedia Player</u> This media player is a super versatile app for any file format you probably may think of: FLV, MP4, AVI, MOV, DAT, MKV, MP3, FLAC, M4V are all supported as well as many others.
- VLC A free, open-source, cross-platform multimedia player as well as framework that plays most multimedia files, DVDs, Audio CDs, VCDs and various streaming protocols.
- XLD A tool to decode, convert and play various 'lossless' audio files.
- <u>HandBrake</u> A tool for converting video from nearly any format to a selection of modern, widely supported codecs.
- MPlayerX A simple, powerful, beautiful media player.
- Playback Experimental video player.
- <u>ScreenFlow</u> Screencasting and video editing software.
- <u>Shotcut</u> Free open-source video editor.
- ArcTime A simple, powerful, efficient subtitle creation software. (Missing English UI and Docs)
- <u>Perian</u> (Retired, no longer supported) <u>Let QuickTime play all the common formats</u>
 of free plug-ins.
- Adapter Free audio, video and image conversion software.
- Synfig Studio Synfig Studio is free, open-source 2D animation software.
- <u>Aegisub</u> Aegisub is a free, cross-platform open source tool for creating and modifying subtitles. Aegisub makes it quick and easy to time subtitles to audio, and features many powerful tools for styling them, including a built-in real-time video preview.
- <u>iFFmpeg</u> A Comprehensive Media Tool for macOS. Making High Quality Video
 Encoding Accessible for Everyone.

Reading and Writing Tools

 OpenOffice - Compatible with other major office suites, Apache OpenOffice is free to download, use, and distribute.

- <u>LibreOffice</u> LibreOffice is free and open-source software. Development is open to new talent and new ideas, and our software is tested and used daily by a large and devoted user community.
- KOffice KOffice contains a word processor (KWord), a spreadsheet (KSpread), a
 presentation program (KPresenter), and a number of other components that varied
 over the course of KOffice's development.
- Spillo Powerful, beautiful and amazingly fast Pinboard client for OS X.
- <u>iChm</u> An ebook reader for CHM (Microsoft Compiled HTML help) files.
- <u>Chmox</u> Read CHM documents on your Mac.
- <u>CHM Reader</u> Read Compiled HTML (.chm) documents on your Mac.
- Skim A PDF reader and note-taker for OS X.
- PDF Expert Read, annotate and edit PDFs, change text and images.
- MarginNote An in-depth PDF and EPUB reading, learning, managing and note taking app.
- <u>Kindle App</u> Amazon official reading app of kindle.
- Klib A new way to manage highlights for Kindle and iBooks.
- texpad A great LaTeX editor for Mac with auto-update PDF and autocomplete LaTeX commands.
- Bear Writer Bear is a beautiful, flexible writing app for crafting notes and prose.
- <u>Boostnote</u> note-taking app made for programmers.
- QOwnNotes Is the open source notepad with markdown support and todo list manager.
- RSS
 - Feeds 2 Keep tabs on your favorite website and RSS feeds from your Mac's menubar.
 - ReadKit A Mac read later client supporting all major providers: Instapaper,
 Pocket and Readability. Even more, ReadKit is a full-featured RSS reader as well.
 - Reeder 3 A news reader for Feedbin, Feedly, Feed Wrangler and so
 on.
 - <u>Leaf</u> An amazing news reader dedicated to help you enjoy your daily news and easily manage your subscriptions.
 - Vienna An RSS/Atom reader for Mac OS X.
- Markdown

- Yu Writer A Markdown text editor that can find writing fun.
- Mou A Markdown editor for developers, on Mac OS X.
- o Marp A Markdown presentation writer with cross-platform support.
- <u>TextNut</u> A rich-format editor featuring Markdown export and Markdown syntax hints.
- MWeb Pro Markdown writing, note taking and static blog generator App.
- <u>Typora</u> A truly minimal Markdown editor featuring seamless live preview.
- o MacDown An open-source Markdown editor for OS X.
- EME Recently launched a Markdown editor, interface is like Chrome browser interface, very simple.
- <u>LightPaper</u> Simple, beautiful, yet powerful text editor for your Mac.
- Marked 2 Markdown previewer, reviewer, and exporter. Great for people that already have a preferred way of editing Markdown, but want to get a live preview of the document they are working on.
- o iA Writer Writing app with an emphasis on simplicity and design.
- <u>Ultimate</u> The Ultimate Writing App for Mac, iPad and iPhone.
- Cmd Markdown Cmd Markdown Open the journey of excellence.

Note-taking

- Quiver The Programmer's Notebook, lets you easily mix text, code, Markdown and LaTeX within one note, edit code with an awesome code editor and live preview Markdown and LaTeX.
- <u>Evernote</u> Infamous note-taking app, available on many platforms.
- OneNote Note-taking app by Microsoft.
- o <u>Inkdrop</u> The notebook app for Markdown lovers built on top of Electron.
- Notes A clean, simple note-taking app.
- Notebook Note-taking app.
- o <u>有道云笔记</u> A note-taking app. It features multi-level notebook structure, Markdown syntax and iWork/Office preview. (**Missing English UI and**

Docs)

- 为知笔记 A note-taking app. It features multi-level notebook structure, multi tags, Markdown syntax and unlimited cloud storage space. (Missing English UI and Docs)
- <u>Leanote</u> A note-taking app. Features Markdown syntax and blog export.
 (Missing English Docs)

Ebooks

- <u>Calibre</u> A free and open-source e-book computer software application suite which runs on multiple platforms, allows users to manage e-book collections as well as create, edit, and read e-books.
- <u>Sigil</u> A multi-platform EPUB ebook Editor.
- <u>Scribus</u> Professional layout and publishing software supporting EPS and SVG import/export, and PDF support.

FTP Clients

- <u>Transmit</u> A highly flexible and intuitive FTP client, supports SFTP, S3 and iDisk/WebDAV.
- Flow An award-winning, beautiful, fast, and reliable FTP + SFTP client.
- Yummy FTP Pro-level, fast, reliable FTP/S + SFTP + WebDAV/S file transfer app.
- <u>Cyberduck</u> A libre FTP, SFTP, WebDAV, S3, Backblaze B2, Azure and OpenStack Swift browser.
- <u>FileZilla</u> A free software, cross-platform FTP application. Supports FTP, SFTP and FTPS
 (FTP over SSL/TLS).

Frameworks For Hybrid Applications

- <u>create-dmg</u> Create a good-looking DMG for your macOS app in seconds.
- nw.js Build desktop application with HTML and JavaScript. It lets you call all Node.js modules directly from DOM and enables a new way of writing applications with all Web technologies.
- <u>Electron</u> Build cross platform desktop application with JavaScript, HTML and
 CSS. _ _ _

- <u>Electrino</u> Desktop runtime for apps built on web technologies, using the system's own web browser engine.
- <u>react-desktop</u> React UI Components for macOS Sierra.
- <u>ReactXP</u> Microsoft official production, support platform Web, iOS, Android and Windows UWP is still an ongoing work.
- <u>React Native macOS</u> Build OS X desktop apps using React Native and Cocoa.
- <u>React Native for Ubuntu</u> Build Ubuntu desktop apps using React Native.
- AppJS A lightweight JavaScript UI library for creating mobile webapps that behave like native apps.
- <u>HEX</u> Build cross-platform desktop application with HTML and JavaScript. Made by YouDao. (**Broken official website**)
- AlloyDesktop Build cross-platform desktop application with HTML and JavaScript.
 Made by Tencent (Not very recommended). (Missing English Docs)
- MacGap Provides a lightweight JavaScript API for OS X integration, such as displaying native notifications or writing data to a file.
- <u>ionic</u> Build amazing native and progressive web apps with Angular and open web technologies. One app running on everything.

Download Management Tools

- <u>Transmission</u> A Fast, Easy, Free BitTorrent Client.
- <u>aria2</u> A lightweight multi-protocol & multi-source command-line download utility.
- <u>JDownloader</u> A free, open-source download management tool with a huge community of developers that makes downloading as easy and fast as it should be.
- You-Get A tiny command-line utility to download media contents (videos, audios, images) from the web.
- <u>Free Download Manager</u> A powerful, modern download accelerator and organizer for Windows and Mac. (FREE)
- <u>FOLX</u> A free download manager for Mac OS X with a true Mac-style interface.

<u>Downie</u> - Video downloader for macOS with support for YouTube and other 1200 sites.

Online Storage

I recommend using online storage with Mac clients

- <u>Dropbox</u> A file hosting service that offers cloud storage and file synchronization with collaborative edit features.
- <u>Baidu cloud</u> Baidu cloud official client. (**Missing English UI and Docs**)
- <u>Tecent cloud</u> Tencent cloud client. (Require Tencent Account, Missing English UI and Docs)
- <u>Jianguo cloud</u> Jianguo cloud client. (Not recommended)
- 115 115 cloud client. (Missing English UI and Docs)
- 360 360 cloud client. (Missing English UI and Docs)
- <u>ownCloud</u> Cloud storage.
- NextCloud An actively maintained fork of ownCloud, faster and completely opensource
- Mega Free cloud service, offers 50GB free storage.
- <u>Seafile</u> Reliable and High Speed File Sync and Share.

Input Methods

- QQ Input for Chinese Produced by Tencent. (Missing English UI and Docs)
- Sogou Input for Chinese Sogou input. (Missing English UI and Docs)
- <u>Baidu Input for Chinese</u> Baidu input. (**Missing English UI and Docs**)
- Qingge Wubi Input Wubi input produced for iOS and Mac. (Missing English UI and Docs)
- WBIM Wubi input method.
- <u>Emoticon Input</u> Emoticons icons or characters input method. (**Missing English UI** and Docs)
- RIME Rime input method. (Missing English UI and Docs)

- <u>hallelujahIM</u> hallelujah English input method.
- Rocket Makes typing emoji faster and easier using Slack-style shortcuts.

Web Browsers

Browsers on Mac OS

- Safari Built-in browser of Mac OS.
- <u>Chrome</u> Chrome, produced by Google
- <u>Chromium</u> Chromium is an open-source web browser project by Google, to provide the source code for Google Chrome.
- <u>Firefox</u> Firefox is a free, open-source web browser developed by the Mozilla
 Foundation.
- Opera Opera.
- QQ Browser QQ browser, produced by Tencent.
- 傲游云浏览器 (Missing English UI and Docs)
- <u>Vivaldi</u> Vivaldi, new browser from Opera's developers.
- <u>Öryöki</u> Öryöki, small web browser with a thin interface. This is an experimental project, currently in development.
- <u>Brave</u> Web browser with an emphasis on privacy and speed.

Translation Tools

(Or you could just use the Mac OS built-in dictionary)

- <u>iTranslate</u> Translate entire website instantly with its built-in browser or with iTranslate
 Safari extension into over 40 languages.
- YouDao Translate (Missing English UI and Docs)
- 海词词典 (Missing English UI and Docs)
- <u>Eudic</u> (Missing English UI and Docs)
- Grammarly Refine your english

Proxy and VPN Tools

- SpechtLite A rule-based proxy app for macOS.
- <u>ShadowsocksX</u> A secure socks5 proxy, designed to protect your internet traffic
- <u>ShadowsocksX-NG</u> Next generation of ShadowsocksX.
- <u>Lantern</u> Lantern is a free application that delivers fast, reliable and secure access to the open internet.
- 鱼摆摆 An application for the access to open internet.
- <u>Tunnelbear</u> Really simple VPN to browse the web privately & securely. Unblock websites around the world with applications for Mac, PC, iOS, Android & Chrome.
- <u>Tunnelblick</u> A free, open-source graphic user interface for OpenVPN on OS X.
- GoAgentX An application for the access to open internet. (Missing English UI and
 Docs)
- Surge Surge is a web developer tool and proxy utility for iOS 9.
- <u>srocket</u> An application for the access to open internet. (Missing English UI and Docs)
- <u>LoCoVPN</u> An application for the access to open internet. (**Missing English UI and Docs**)
- <u>二师兄 VPN</u> An application that connects you to the VPN. (Missing English UI and Docs)
- GTX 加速器 An application that connects you to the VPN. (Missing English UI and Docs)
- GreenVPN A powerful and fun tools that helps you to solve any network issues.
 (Mainly used for the access to open internet)
- WiseVPN Unlimited traffic, free VPN.
- 风驰 VPN An application that connects you to the VPN. (Missing English UI and Docs)
- <u>开眼</u> A Chrome plugin for the access to open internet.
- <u>PlutoX</u> An application that connects you to the VPN.
- tinc Secure mesh VPN software.

- <u>Shimo</u> VPN Client for Mac, provides many features related to VPN usage.
- <u>FreeVPN Plus</u> Never expired Mac free VPN.

Utilities

General Tools

- Alfred An award-winning app for Mac OS X which boosts your efficiency with hotkeys, keywords, text expansion and more. Search your Mac and the web, and be more productive with custom actions to control your Mac.
- <u>BetterZip 3</u> An archive tool supports ZIP, TAR, TGZ, TBZ, TXZ (new), 7-ZIP,
 RAR_o
- <u>Keka</u> A free file archiver for macOS. Compression formats supported:7z, Zip, Tar, Gzip, Bzip2, DMG, ISO. Extraction formats supported: RAR, 7z, Lzma, xz, Zip, Tar, Gzip, Bzip2, ISO, EXE, CAB, PAX.
- Numi A beautiful calculator app for Mac.
- Memo Memo is a simple and elegant app, Unlock memos even more quickly using Touch ID.
- <u>Itsycal</u> A tiny calendar for your Mac's menu bar.
- <u>BitBar</u> An app lets you put the output from any script or program right in your Mac
 OS X menu bar. Has a good community.
- ClipMenu A clipboard manager for Mac OS X.
- Paste Smart clipboard history & snippets manager.
- <u>iPaste</u> Lightweight and efficient clipboard tool.
- <u>ControlPlane</u> Manages configuration profiles for your Mac. Determines where you
 are or what you are doing based on a number of available evidence sources and then
 automatically reconfigures your Mac based on your preferences.
- <u>AirServer</u> The most advanced screen mirroring software receiver for Mac, PC and Xbox One.
- <u>Hazel</u> Automated file organization for your Mac. Responsibly and beautifully designed.
- MacAssistant Google Assistant for macOS
- <u>muCommander</u> A lightweight file manager with a dual-pane interface.

GTD (Getting Things Done)

- <u>Todoist</u> A cross-platform todo list app.
- <u>dida365</u> Lightweight to-do application。
- <u>TaskPaper</u> Plain text to-do lists.
- Fantastical The calendar app you won't be able to live without.
- Day-O 2 Menu bar clock replacement with built-in calendar.
- Things A delightful and easy to use task manager. (Award-winning App)
- OmniFocus A nice GTD app, made by OmniGroups.
- 2Do A nice todo app.
- Wunderlist The easiest way to get stuff done.
- OnyX Multifunction utility to verify disks and files, run cleaning and system maintenance tasks, configure hidden options and more.
- Focus A beautiful pomodoro-based time manager.

Productivity

- <u>xScope</u> A powerful set of tools that are ideal for measuring, inspecting & testing onscreen graphics and layouts.
- <u>Karabiner</u> A powerful and stable keyboard customizer for OS X.
- <u>Ukelele</u> Unicode Keyboard Layout Editor.
- <u>Keytty</u> A app to keep your hands on the keyboard. Move, click, scroll, drag and more with a few strokes.
- Mos A simple tool can offer the smooth scrolling and reverse the mouse scrolling direction on your Mac.
- <u>BetterTouchTool</u> A great, feature-packed app that allows you to configure many gestures for your Magic Mouse, Macbook Trackpad, Magic Trackpad and also Mouse Gestures for normal mice.
- <u>Hammerspoon</u> A tool for powerful automation of OS X, programmable by Lua scripting engine.
- <u>Qbserve</u> Time tracking automation: freelance project tracking, timesheets, invoicing & real-time productivity feedback.
- <u>Timing</u> Automatic time and productivity tracking for Mac. Helps you stay on track with your work and ensures no billable hours get lost if you are billing hourly.

Window Management

- Shiftlt Managing window size and position in OSX.
- Moom Allows you to easily move and zoom windows, or to another display—using either the mouse or the keyboard.
- <u>Divvy</u> Window management at its finest with its amazing Divvy Grid system.
- Slate A window management application similar to Divvy and SizeUp (except better and free!). (Needs config file)
- o <u>SizeUp</u> Powerful, keyboard-centric window management.
- Spectacle Move and resize windows with simple and customizable keyboard shortcuts.
- Amethyst A tiling window manager.
- Magnet A window manager that keeps your workspace organized.

Password Management

- o <u>1Password</u> A cross-platform password management tool.
- <u>LastPass</u> A password management tool for Mac OS and browser.
- KeePassX A light-weight and open-source password management
 app. _
- MacPass An open-source KeePass Mac OS client.
- Keeweb Free cross-platform password manager compatible with KeePass.

Finder

- o Quicklook-Plugins A list of useful "Quick Look" plugins for developers.
- ForkLift The most advanced dual pane file manager and file transfer client for macOS.
- Path Finder An all-around file management app.
- TotalFinder A Chrome-like app as finder substitute.
- XtraFinder Adds tabs and cut to Mac Finder.

Quality of Life Improvements

- <u>CheatSheet</u> Hold the \(\mathbb{H}\)-Key to get a list of all active shortcuts of the current
 application. It's as simple as that.
- Snap Launch an app in a snap. Ridiculously easy shortcut management.
- <u>KeyCastr</u> An open-source keystroke visualizer.

 <u>f.lux</u> - An application makes the color of your computer's display adapt to the time of day.

System Related Tools

- AppCleaner A small application which allows you to thoroughly uninstall unwanted apps.
- <u>iStats</u> iStats is a command-line tool that allows you to easily grab the CPU temperature, fan speeds and battery information on OS X.
- <u>Juice</u> Make your battery information a bit more interesting by making your own measurement scale instead of a boring battery indicator.
- <u>NitroShare</u> CROSS-PLATFORM NETWORK FILE TRANSFER APPLICATION.
- Monity A system monitoring widget for OS X.
- <u>SSH Tunnel</u> An application to manage your SSH.
- Mounty A tiny tool to re-mount write-protected NTFS volumes under Mac OS X
 10.9+ in read-write mode.
- <u>Paragon NTFS</u> Read/write access to NTFS in macOS Sierra.
- <u>Tuxera NTFS</u> Full read-write compatibility with NTFS-formatted drives on a Mac.
- <u>gfxCardStatus</u> An unobtrusive menu bar app for OS X that allows MacBook Pro users to see which apps are affecting their battery life by using the more power-hungry graphics.
- <u>DaisyDisk</u> Gives you a great overview of disk usage. Can also make more disk-space available by cleaning up your disk.
- OmniDiskSweeper Shows you the files on your drive, ordered by size. It can be used to find and remove unused files.
- <u>iStat Menus</u> An advanced Mac system monitor on the menubar.
- HTML5 Player An HTML 5 video player. Keep your Mac from "burning".
- <u>InsomniaX</u> Disable either lid and/or idle sleep (so you can play music with your lid closed, for example).
- <u>Caffeine</u> An app that doesn't let your Mac to fall asleep.
- <u>KeepingYouAwake</u> Alternative to Caffeine with better support for dark mode in Mac.
- NoSleep NoSleep makes closing of your MacBook lid possible without going into sleep mode.
- <u>Coolant</u> A menubar app that lets you know when an app is consuming 100% CPU or more than a gigabyte of memory (or any arbitrary limits you choose).

HandShaker - Mac on the management of Android mobile phone content.

Gaming Software

 OpenEmu - A great video game console emulator, supports many different emulators in a single application. (e.g. Sony PSP, GameBoy, NDS and so on)

Other

WWDC - The Mac OS unofficial WWDC app.

Remote Login Software

- <u>TeamViewer</u> A proprietary computer software package for remote control, desktop sharing, online meetings, web conferencing and file transfer between computers.
- RealVNC The original and best software for remote access across desktop and mobile.
- AnyDesk An application which provides Remote access across multiple machines.

QuickLook Plugins

Using <u>Homebrew Cask</u> to install. List of useful <u>Quick Look</u> plugins for developers. If you install it manually, you can move the downloaded .qLgenerator file to~ / Library / QuickLook qLmanage -r

- QuicklookStephen Preview plain text files without or with unknown file extension. such as README、INSTALL、Capfile、CHANGELOG...brew cask install qlstephen
- QLColorCode Preview source code files with syntax highlighting. brew cask install qlcolorcode
- QLMarkdown Preview Markdown files. brew cask install qlmarkdown
- QuickLookJSON Preview JSON files. brew cask install quicklook-json
- QLPrettyPatch Preview .patch files. brew cask install glprettypatch
- QuickLookCSV Preview CSV files. brew cask install quicklook-csv
- <u>BetterZipQL</u> Preview the compressed format file for BetterZip software. brew cask install betterzipql
- <u>qllmageSize</u> Display image size and resolution. brew cask install qlimagesize
- WebP Preview WebP images. brew cask install webpquicklook
- <u>Suspicious Package</u> Preview the contents of a standard Apple installer package. brew cask install suspicious-package
- QuickLookASE Preview Adobe ASE Color Swatches generated with Photoshop,
 Illustrator, Prisma among many others. brew cask install quicklookase

- QLVideo Preview most types of video files, as well as their thumbnails, cover art and metadata. brew cask install glvideo
- QuickLookAPK Preview Android APK files. brew cask install quicklookapk
- <u>ProvisionQL</u> Preview iOS / macOS app and provision information. brew cask install provisionql
- <u>quicklook-pat</u> Preview Adobe Photoshop pattern files. brew cask install quicklook-pat

Third Party App Markets

If you come across websites offering pirated software or cracks, please post <u>HERE</u>. We love apps, but only authentic ones. :)

Get Apps

Here are some of the major software download sites, there are a number of OSX Mac software sites

- MacUpdate Desktop Simplifies finding, buying and installing apps for your Mac.
- Homebrew Cask A command line installation manager which extends Homebrew and brings its elegance, simplicity, and speed to Mac OS applications and large binaries alike.
- <u>Homebrew</u> The missing package manager for macOS.
- <u>MacPorts</u> The MacPorts Project is an open-source community initiative to design an
 easy-to-use system for compiling, installing, and upgrading either command-line, X11
 or Aqua based open-source software on the Mac OS X operating system.

Mac App Download Sites

Here are some of the major software download sites, there are a number of OSX Mac software sites

Genuine Sites

- <u>Slant</u> I personally recommend this. This is a platform where you can compare apps side-by-side, you might get an idea by seeing other users recommendations. Please
 - contribute if you find an application from this list!
- <u>alternativeTo</u> Also a very nice community. If you are looking for some alternative apps **FOR** Windows or another platform, check this site.
- Other sites like <u>MacStories</u>, <u>LifeHacker</u>, <u>ProductHunt</u> are great resources.
- Also, Quora, Reddit, you know the drill.

- MacUpdate : https://www.macupdate.com/
- App Shopper : http://appshopper.com/
- <u>HackStore</u> An amazing free alternative to the App Store for side developers and apps for Mac OS which Apple doesn't accept.

• Chinese community

。 少数派:http://sspai.com/tag/Mac

。 Mac 玩儿法:<u>http://www.waerfa.com</u>

Awesome Big Data

https://github.com/onurakpolat/awesome-bigdata

A curated list of awesome big data frameworks, resources and other awesomeness. Inspired by awesome-python, <a href="mailto:awesome-python-python-python-python-python-python-python-python-python-python-python-python-python-python-python-pytho

Your contributions are always welcome!

- Awesome Big Data
 - o RDBMS
 - o Frameworks
 - Distributed Programming
 - Distributed Filesystem
 - o Key-Map Data Model
 - Document Data Model
 - o Key-value Data Model
 - o Graph Data Model
 - NewSQL Databases
 - Columnar Databases
 - o Time-Series Databases
 - o SQL-like processing
 - o Data Ingestion
 - o Service Programming
 - Scheduling
 - Machine Learning
 - Benchmarking

- Security
- o System Deployment
- Applications
- o Search engine and framework
- MySQL forks and evolutions
- PostgreSQL forks and evolutions
- Memcached forks and evolutions
- Embedded Databases
- o <u>Business Intelligence</u>
- Data Visualization
- Internet of things and sensor data
- o <u>Interesting Readings</u>
- Interesting Papers
- o <u>Videos</u>
- o Books
- Other Awesome Lists

RDBMS

- MySQL The world's most popular open source database.
- <u>PostgreSQL</u> The world's most advanced open source database.
- <u>Oracle Database</u> object-relational database management system.

Frameworks

- <u>IBM Streams</u> platform for distributed processing and real-time analytics. Integrates with many of the popular technologies in the Big Data ecosystem (Kafka, HDFS, Spark, etc.)
- <u>Apache Hadoop</u> framework for distributed processing. Integrates MapReduce (parallel processing), YARN (job scheduling) and HDFS (distributed file system).

- <u>Tigon</u> High Throughput Real-time Stream Processing Framework.
- <u>Pachyderm</u> Pachyderm is a data storage platform built on Docker and Kubernetes to provide reproducible data processing and analysis.

Distributed Programming

- AddThis Hydra distributed data processing and storage system originally developed at AddThis.
- AMPLab SIMR run Spark on Hadoop MapReduce v1.
- Apache APEX a unified, enterprise platform for big data stream and batch processing.
- Apache Beam an unified model and set of language-specific SDKs for defining and executing data processing workflows.
- Apache Crunch a simple Java API for tasks like joining and data aggregation that are tedious to implement on plain MapReduce.
- Apache DataFu collection of user-defined functions for Hadoop and Pig developed by LinkedIn.
- Apache Flink high-performance runtime, and automatic program optimization.
- Apache Gearpump real-time big data streaming engine based on Akka.
- <u>Apache Gora</u> framework for in-memory data model and persistence.
- Apache Hama BSP (Bulk Synchronous Parallel) computing framework.
- <u>Apache MapReduce</u> programming model for processing large data sets with a parallel, distributed algorithm on a cluster.
- Apache Pig high level language to express data analysis programs for Hadoop.
- Apache REEF retainable evaluator execution framework to simplify and unify the lower layers of big data systems.
- <u>Apache S4</u> framework for stream processing, implementation of S4.
- <u>Apache Spark</u> framework for in-memory cluster computing.
- Apache Spark Streaming framework for stream processing, part of Spark.
- <u>Apache Storm</u> framework for stream processing by Twitter also on YARN.
- Apache Samza stream processing framework, based on Kafka and YARN.
- <u>Apache Tez</u> application framework for executing a complex DAG (directed acyclic graph) of tasks, built on YARN.
- <u>Apache Twill</u> abstraction over YARN that reduces the complexity of developing distributed applications.
- <u>Cascalog</u> data processing and querying library.
- <u>Cheetah</u> High Performance, Custom Data Warehouse on Top of MapReduce.
- Concurrent Cascading framework for data management/analytics on Hadoop.
- <u>Damballa Parkour</u> MapReduce library for Clojure.
- <u>Datasalt Pangool</u> alternative MapReduce paradigm.

- <u>DataTorrent StrAM</u> real-time engine is designed to enable distributed, asynchronous, real time in-memory big-data computations in as unblocked a way as possible, with minimal overhead and impact on performance.
- Facebook Corona Hadoop enhancement which removes single point of failure.
- Facebook Peregrine Map Reduce framework.
- <u>Facebook Scuba</u> distributed in-memory datastore.
- Google Dataflow create data pipelines to help themæingest, transform and analyze data.
- Google MapReduce map reduce framework.
- <u>Google MillWheel</u> fault tolerant stream processing framework.
- <u>IBM Streams</u> platform for distributed processing and real-time analytics. Provides toolkits for advanced analytics like geospatial, time series, etc. out of the box.
- <u>JAQL</u> declarative programming language for working with structured, semistructured and unstructured data.
- <u>Kite</u> is a set of libraries, tools, examples, and documentation focused on making it easier to build systems on top of the Hadoop ecosystem.
- Metamarkets Druid framework for real-time analysis of large datasets.
- Netflix PigPen map-reduce for Clojure which compiles to Apache Pig.
- Nokia Disco MapReduce framework developed by Nokia.
- Onyx Distributed computation for the cloud.
- <u>Pinterest Pinlater</u> asynchronous job execution system.
- <u>Pydoop</u> Python MapReduce and HDFS API for Hadoop.
- Rackerlabs Blueflood multi-tenant distributed metric processing system
- Skale High performance distributed data processing in NodeJS.
- <u>Stratosphere</u> general purpose cluster computing framework.
- <u>Streamdrill</u> useful for counting activities of event streams over different time windows and finding the most active one.
- <u>streamsx.topology</u> Libraries to enable building IBM Streams application in Java, Python or Scala.
- <u>Tuktu</u> Easy-to-use platform for batch and streaming computation, built using Scala, Akka and Play!
- <u>Twitter Heron</u> Heron is a realtime, distributed, fault-tolerant stream processing engine from Twitter replacing Storm.
- <u>Twitter Scalding</u> Scala library for Map Reduce jobs, built on Cascading.
- <u>Twitter Summingbird</u> Streaming MapReduce with Scalding and Storm, by Twitter.
- <u>Twitter TSAR</u> TimeSeries AggregatoR by Twitter.
- Wallaroo The ultrafast and elastic data processing engine. Big or fast data no fuss, no Java needed.

Distributed Filesystem

- Ambry a distributed object store that supports storage of trillion of small immutable objects as well as billions of large objects.
- Apache HDFS a way to store large files across multiple machines.
- <u>BeeGFS</u> formerly FhGFS, parallel distributed file system.
- <u>Ceph Filesystem</u> software storage platform designed.
- <u>Disco DDFS</u> distributed filesystem.
- <u>Facebook Haystack</u> object storage system.
- Google Colossus distributed filesystem (GFS2).
- Google GFS distributed filesystem.
- Google Megastore scalable, highly available storage.
- GridGain GGFS, Hadoop compliant in-memory file system.
- <u>Lustre file system</u> high-performance distributed filesystem.
- Microsoft Azure Data Lake Store HDFS-compatible storage in Azure cloud
- Quantcast File System QFS open-source distributed file system.
- Red Hat GlusterFS scale-out network-attached storage file system.
- <u>Seaweed-FS</u> simple and highly scalable distributed file system.
- Alluxio reliable file sharing at memory speed across cluster frameworks.
- <u>Tahoe-LAFS</u> decentralized cloud storage system.
- <u>Baidu File System</u> distributed filesystem.

Document Data Model

- Actian Versant commercial object-oriented database management systems .
- <u>Crate Data</u> is an open source massively scalable data store. It requires zero administration.
- Facebook Apollo Facebook's Paxos-like NoSQL database.
- jumboDB document oriented datastore over Hadoop.
- <u>LinkedIn Espresso</u> horizontally scalable document-oriented NoSQL data store.
- MarkLogic Schema-agnostic Enterprise NoSQL database technology.
- <u>Microsoft Azure DocumentDB</u> NoSQL cloud database service with protocol support for MongoDB
- MongoDB Document-oriented database system.
- RavenDB A transactional, open-source Document Database.
- <u>RethinkDB</u> document database that supports queries like table joins and group by.

Key Map Data Model

Note: There is some term confusion in the industry, and two different things are called "Columnar Databases". Some, listed here, are distributed, persistent databases built around the "key-map" data model: all data has a (possibly composite) key, with

which a map of key-value pairs is associated. In some systems, multiple such value maps can be associated with a key, and these maps are referred to as "column families" (with value map keys being referred to as "columns").

Another group of technologies that can also be called "columnar databases" is distinguished by how it stores data, on disk or in memory -- rather than storing data the traditional way, where all column values for a given key are stored next to each other, "row by row", these systems store all *column* values next to each other. So more work is needed to get all columns for a given key, but less work is needed to get all values for a given column.

The former group is referred to as "key map data model" here. The line between these and the <u>Key-value Data Model</u> stores is fairly blurry.

The latter, being more about the storage format than about the data model, is listed under Columnar Databases.

You can read more about this distinction on Prof. Daniel Abadi's blog: <u>Distinguishing</u> two major types of Column Stores.

- Apache Accumulo distributed key/value store, built on Hadoop.
- Apache Cassandra column-oriented distributed datastore, inspired by BigTable.
- Apache HBase column-oriented distributed datastore, inspired by BigTable.
- <u>Baidu Tera</u> an Internet-scale database, inspired by BigTable.
- Facebook HydraBase evolution of HBase made by Facebook.
- Google BigTable column-oriented distributed datastore.
- <u>Google Cloud Datastore</u> is a fully managed, schemaless database for storing non-relational data over BigTable.
- <u>Hypertable</u> column-oriented distributed datastore, inspired by BigTable.
- <u>InfiniDB</u> is accessed through a MySQL interface and use massive parallel processing to parallelize queries.
- <u>Tephra</u> Transactions for HBase.
- <u>Twitter Manhattan</u> real-time, multi-tenant distributed database for Twitter scale.
- <u>ScyllaDB</u> column-oriented distributed datastore written in C++, totally compatible with Apache Cassandra.

Key-value Data Model

- <u>Aerospike</u> NoSQL flash-optimized, in-memory. Open source and "Server code in 'C' (not Java or Erlang) precisely tuned to avoid context switching and memory copies."
- <u>Amazon DynamoDB</u> distributed key/value store, implementation of Dynamo paper.
- <u>Badger</u> a fast, simple, efficient, and persistent key-value store written natively in Go.
- Bolt an embedded key-value database for Go.

- <u>BTDB</u> Key Value Database in .Net with Object DB Layer, RPC, dynamic IL and much more
- <u>BuntDB</u> a fast, embeddable, in-memory key/value database for Go with custom indexing and geospatial support.
- Edis is a protocol-compatible Server replacement for Redis.
- <u>ElephantDB</u> Distributed database specialized in exporting data from Hadoop.
- EventStore distributed time series database.
- GridDB suitable for sensor data stored in a timeseries.
- <u>HyperDex</u> a scalable, next generation key-value and document store with a wide array of features, including consistency, fault tolerance and high performance.
- <u>LinkedIn Krati</u> is a simple persistent data store with very low latency and high throughput.
- <u>Linkedin Voldemort</u> distributed key/value storage system.
- Oracle NoSQL Database distributed key-value database by Oracle Corporation.
- Redis in memory key value datastore.
- Riak a decentralized datastore.
- Storehaus library to work with asynchronous key value stores, by Twitter.
- <u>SummitDB</u> an in-memory, NoSQL key/value database, with disk persistance and using the Raft consensus algorithm.
- <u>Tarantool</u> an efficient NoSQL database and a Lua application server.
- <u>TiKV</u> a distributed key-value database powered by Rust and inspired by Google Spanner and HBase.
- <u>Tile38</u> a geolocation data store, spatial index, and realtime geofence, supporting a variety of object types including latitude/longitude points, bounding boxes, XYZ tiles, Geohashes, and GeoJSON
- <u>TreodeDB</u> key-value store that's replicated and sharded and provides atomic multirow writes

Graph Data Model

- AgensGraph a new generation multi-model graph database for the modern complex data environment.
- Apache Giraph implementation of Pregel, based on Hadoop.
- Apache Spark Bagel implementation of Pregel, part of Spark.
- <u>ArangoDB</u> multi model distributed database.
- <u>DGraph</u> A scalable, distributed, low latency, high throughput graph database aimed at providing Google production level scale and throughput, with low enough latency to be serving real time user queries, over terabytes of structured data.
- <u>EliasDB</u> a lightweight graph based database that does not require any third-party libraries.
- <u>Facebook TAO</u> TAO is the distributed data store that is widely used at facebook to store and serve the social graph.

- GCHQ Gaffer Gaffer by GCHQ is a framework that makes it easy to store large-scale graphs in which the nodes and edges have statistics.
- Google Cayley open-source graph database.
- <u>Google Pregel</u> graph processing framework.
- <u>GraphLab PowerGraph</u> a core C++ GraphLab API and a collection of highperformance machine learning and data mining toolkits built on top of the GraphLab API.
- GraphX resilient Distributed Graph System on Spark.
- Gremlin graph traversal Language.
- Infovore RDF-centric Map/Reduce framework.
- Intel GraphBuilder tools to construct large-scale graphs on top of Hadoop.
- MapGraph Massively Parallel Graph processing on GPUs.
- Neo4j graph database writting entirely in Java.
- OrientDB document and graph database.
- <u>Phoebus</u> framework for large scale graph processing.
- <u>Titan</u> distributed graph database, built over Cassandra.
- Twitter FlockDB distributed graph database.

Columnar Databases

Note please read the note on <u>Key-Map Data Model</u> section.

- <u>Columnar Storage</u> an explanation of what columnar storage is and when you might want it.
- Actian Vector column-oriented analytic database.
- C-Store column oriented DBMS.
- <u>ClickHouse</u> an open-source column-oriented database management system that allows generating analytical data reports in real time.
- <u>EventQL</u> a distributed, column-oriented database built for large-scale event collection and analytics.
- MonetDB column store database.
- <u>Parquet</u> columnar storage format for Hadoop.
- <u>Pivotal Greenplum</u> purpose-built, dedicated analytic data warehouse that offers a columnar engine as well as a traditional row-based one.
- <u>Vertica</u> is designed to manage large, fast-growing volumes of data and provide very fast query performance when used for data warehouses.
- <u>SQream DB</u> A GPU powered big data database, designed for analytics and data warehousing, with ANSI-92 compliant SQL, suitable for data sets from 10TB to 1PB.
- Google BigQuery Google's cloud offering backed by their pioneering work on Dremel.
- <u>Amazon Redshift</u> Amazon's cloud offering, also based on a columnar datastore backend.

 IndexR an open-source columnar storage format for fast & realtime analytic with big data.

NewSQL Databases

- <u>Actian Ingres</u> commercially supported, open-source SQL relational database management system.
- Amazon RedShift data warehouse service, based on PostgreSQL.
- BayesDB statistic oriented SQL database.
- <u>Bedrock</u> a simple, modular, networked and distributed transaction layer built atop SQLite.
- <u>CitusDB</u> scales out PostgreSQL through sharding and replication.
- Cockroach Scalable, Geo-Replicated, Transactional Datastore.
- <u>Comdb2</u> a clustered RDBMS built on optimistic concurrency control techniques.
- <u>Datomic</u> distributed database designed to enable scalable, flexible and intelligent applications.
- <u>FoundationDB</u> distributed database, inspired by F1.
- Google F1 distributed SQL database built on Spanner.
- Google Spanner globally distributed semi-relational database.
- <u>H-Store</u> is an experimental main-memory, parallel database management system that is optimized for on-line transaction processing (OLTP) applications.
- <u>Haeinsa</u> linearly scalable multi-row, multi-table transaction library for HBase based on Percolator.
- HandlerSocket NoSQL plugin for MySQL/MariaDB.
- <u>InfiniSQL</u> infinity scalable RDBMS.
- MemSQL in memory SQL database witho optimized columnar storage on flash.
- <u>NuoDB</u> SQL/ACID compliant distributed database.
- <u>Oracle TimesTen in-Memory Database</u> in-memory, relational database management system with persistence and recoverability.
- <u>Pivotal GemFire XD</u> Low-latency, in-memory, distributed SQL data store. Provides SQL interface to in-memory table data, persistable in HDFS.
- <u>SAP HANA</u> is an in-memory, column-oriented, relational database management system.
- <u>SenseiDB</u> distributed, realtime, semi-structured database.
- <u>Sky</u> database used for flexible, high performance analysis of behavioral data.
- <u>SymmetricDS</u> open source software for both file and database synchronization.
- Map-D GPU in-memory database, big data analysis and visualization platform
- <u>TiDB</u> TiDB is a distributed SQL database. Inspired by the design of Google F1.
- VoltDB claims to be fastest in-memory database

Time-Series Databases

- <u>Axibase Time Series Database</u> distributed time series database on top of HBase. Includes built-in Rule Engine, data forecasting and visualization.
- <u>Chronix</u> a time series storage built to store time series highly compressed and for fast access times.
- <u>Cube</u> uses MongoDB to store time series data.
- Heroic is a scalable time series database based on Cassandra and Elasticsearch.
- InfluxDB distributed time series database.
- <u>Kairosdb</u> similar to OpenTSDB but allows for Cassandra.
- Newts a time series database based on Apache Cassandra.
- OpenTSDB distributed time series database on top of HBase.
- <u>Prometheus</u> a time series database and service monitoring system.
- <u>Beringei</u> Facebook's in-memory time-series database.
- <u>TrailDB</u> an efficient tool for storing and querying series of events.
- <u>Druid</u> Column oriented distributed data store ideal for powering interactive applications
- <u>Riak-TS</u> Riak TS is the only enterprise-grade NoSQL time series database optimized specifically for IoT and Time Series data.
- <u>Akumuli</u> Akumuli is a numeric time-series database. It can be used to capture, store and process time-series data in real-time. The word "akumuli" can be translated from esperanto as "accumulate".
- Rhombus A time-series object store for Cassandra that handles all the complexity of building wide row indexes.
- Dalmatiner DB Fast distributed metrics database
- <u>Blueflood</u> A distributed system designed to ingest and process time series data
- <u>Timely</u> Timely is a time series database application that provides secure access to time series data based on Accumulo and Grafana.

SQL-like processing

- Actian SQL for Hadoop high performance interactive SQL access to all Hadoop data.
- <u>Apache Drill</u> framework for interactive analysis, inspired by Dremel.
- Apache HCatalog table and storage management layer for Hadoop.
- Apache Hive SQL-like data warehouse system for Hadoop.
- <u>Apache Calcite</u> framework that allows efficient translation of queries involving heterogeneous and federated data.
- Apache Phoenix SQL skin over HBase.
- <u>Cloudera Impala</u> framework for interactive analysis, Inspired by Dremel.
- <u>Concurrent Lingual</u> SQL-like query language for Cascading.
- <u>Datasalt Splout SQL</u> full SQL query engine for big datasets.
- <u>Facebook PrestoDB</u> distributed SQL query engine.
- Google BigQuery framework for interactive analysis, implementation of Dremel.

- <u>PipelineDB</u> an open-source relational database that runs SQL queries continuously on streams, incrementally storing results in tables.
- <u>Pivotal HDB</u> SQL-like data warehouse system for Hadoop.
- <u>RainstorDB</u> database for storing petabyte-scale volumes of structured and semistructured data.
- Spark Catalyst is a Query Optimization Framework for Spark and Shark.
- SparkSQL Manipulating Structured Data Using Spark.
- Splice Machine a full-featured SQL-on-Hadoop RDBMS with ACID transactions.
- Stinger interactive query for Hive.
- <u>Tajo</u> distributed data warehouse system on Hadoop.
- <u>Trafodion</u> enterprise-class SQL-on-HBase solution targeting big data transactional or operational workloads.

Data Ingestion

- Amazon Kinesis real-time processing of streaming data at massive scale.
- Apache Chukwa data collection system.
- Apache Flume service to manage large amount of log data.
- Apache Kafka distributed publish-subscribe messaging system.
- Apache Sqoop tool to transfer data between Hadoop and a structured datastore.
- Cloudera Morphlines framework that help ETL to Solr, HBase and HDFS.
- <u>Embulk</u> open-source bulk data loader that helps data transfer between various databases, storages, file formats, and cloud services.
- Facebook Scribe streamed log data aggregator.
- Fluentd tool to collect events and logs.
- <u>Google Photon</u> geographically distributed system for joining multiple continuously flowing streams of data in real-time with high scalability and low latency.
- <u>Heka</u> open source stream processing software system.
- <u>HIHO</u> framework for connecting disparate data sources with Hadoop.
- <u>Kestrel</u> distributed message queue system.
- <u>LinkedIn Databus</u> stream of change capture events for a database.
- LinkedIn Kamikaze utility package for compressing sorted integer arrays.
- <u>LinkedIn White Elephant</u> log aggregator and dashboard.
- <u>Logstash</u> a tool for managing events and logs.
- Netflix Suro log agregattor like Storm and Samza based on Chukwa.
- <u>Pinterest Secor</u> is a service implementing Kafka log persistance.
- <u>Linkedin Gobblin</u> linkedin's universal data ingestion framework.
- <u>Skizze</u> sketch data store to deal with all problems around counting and sketching using probabilistic data-structures.
- <u>StreamSets Data Collector</u> continuous big data ingest infrastructure with a simple to use IDE.

- Yahoo Pulsar a distributed pub-sub messaging platform with a very flexible messaging model and an intuitive client API.
- <u>Alooma</u> data pipeline as a service enabling moving data sources such as MySQL into data warehouses.

Service Programming

- Akka Toolkit runtime for distributed, and fault tolerant event-driven applications on the JVM.
- Apache Avro data serialization system.
- Apache Curator Java libaries for Apache ZooKeeper.
- Apache Karaf OSGi runtime that runs on top of any OSGi framework.
- Apache Thrift framework to build binary protocols.
- Apache Zookeeper centralized service for process management.
- Google Chubby a lock service for loosely-coupled distributed systems.
- <u>Hydrosphere Mist</u> a service for exposing Apache Spark analytics jobs and machine learning models as realtime, batch or reactive web services.
- <u>Linkedin Norbert</u> cluster manager.
- OpenMPI message passing framework.
- Serf decentralized solution for service discovery and orchestration.
- Spotify Luigi a Python package for building complex pipelines of batch jobs. It handles dependency resolution, workflow management, visualization, handling failures, command line integration, and much more.
- <u>Spring XD</u> distributed and extensible system for data ingestion, real time analytics, batch processing, and data export.
- <u>Twitter Elephant Bird</u> libraries for working with LZOP-compressed data.
- <u>Twitter Finagle</u> asynchronous network stack for the JVM.

Scheduling

- Apache Airflow a platform to programmatically author, schedule and monitor workflows.
- <u>Apache Aurora</u> is a service scheduler that runs on top of Apache Mesos.
- Apache Falcon data management framework.
- Apache Oozie workflow job scheduler.
- Azure Data Factory cloud-based pipeline orchestration for on-prem, cloud and HDInsight
- Chronos distributed and fault-tolerant scheduler.
- <u>Linkedin Azkaban</u> batch workflow job scheduler.
- <u>Schedoscope</u> Scala DSL for agile scheduling of Hadoop jobs.
- <u>Sparrow</u> scheduling platform.

Machine Learning

- Azure ML Studio Cloud-based Azure ML, R, Python Machine Learning platform
- <u>brain</u> Neural networks in JavaScript.
- <u>Cloudera Oryx</u> real-time large-scale machine learning.
- <u>Concurrent Pattern</u> machine learning library for Cascading.
- <u>convnetjs</u> Deep Learning in Javascript. Train Convolutional Neural Networks (or ordinary ones) in your browser.
- <u>DataVec</u> A vectorization and data preprocessing library for deep learning in Java and Scala. Part of the Deeplearning4j ecosystem.
- <u>Deeplearning4j</u> Fast, open deep learning for the JVM (Java, Scala, Clojure). A neural network configuration layer powered by a C++ library. Uses Spark and Hadoop to train nets on multiple GPUs and CPUs.
- <u>Decider</u> Flexible and Extensible Machine Learning in Ruby.
- <u>ENCOG</u> machine learning framework that supports a variety of advanced algorithms, as well as support classes to normalize and process data.
- etcML text classification with machine learning.
- <u>Etsy Conjecture</u> scalable Machine Learning in Scalding.
- <u>GraphLab Create</u> A machine learning platform in Python with a broad collection of ML toolkits, data engineering, and deployment tools.
- <u>H2O</u> statistical, machine learning and math runtime with Hadoop. R and Python.
- <u>Keras</u> An intuitive neural net API inspired by Torch that runs atop Theano and Tensorflow.
- <u>Mahout</u> An Apache-backed machine learning library for Hadoop.
- <u>MLbase</u> distributed machine learning libraries for the BDAS stack.
- MLPNeuralNet Fast multilayer perceptron neural network library for iOS and Mac OS X.
- MOA MOA performs big data stream mining in real time, and large scale machine learning.
- MonkeyLearn Text mining made easy. Extract and classify data from text.
- ND4J A matrix library for the JVM. Numpy for Java.
- <u>nupic</u> Numenta Platform for Intelligent Computing: a brain-inspired machine intelligence platform, and biologically accurate neural network based on cortical learning algorithms.
- <u>PredictionIO</u> machine learning server buit on Hadoop, Mahout and Cascading.
- <u>RL4J</u> Reinforcement learning for Java and Scala. Includes Deep-Q learning and A3C algorithms, and integrates with Open Al's Gym. Runs in the Deeplearning4j ecosystem.
- <u>SAMOA</u> distributed streaming machine learning framework.
- <u>scikit-learn</u> scikit-learn: machine learning in Python.
- Spark MLlib a Spark implementation of some common machine learning (ML) functionality.
- Sibyl System for Large Scale Machine Learning at Google.

- <u>TensorFlow</u> Library from Google for machine learning using data flow graphs.
- <u>Theano</u> A Python-focused machine learning library supported by the University of Montreal.
- Torch A deep learning library with a Lua API, supported by NYU and Facebook.
- <u>Velox</u> System for serving machine learning predictions.
- Vowpal Wabbit learning system sponsored by Microsoft and Yahoo!.
- WEKA suite of machine learning software.
- <u>BidMach</u> CPU and GPU-accelerated Machine Learning Library.

Benchmarking

- <u>Apache Hadoop Benchmarking</u> micro-benchmarks for testing Hadoop performances.
- <u>Berkeley SWIM Benchmark</u> real-world big data workload benchmark.
- Intel HiBench a Hadoop benchmark suite.
- <u>PUMA Benchmarking</u> benchmark suite for MapReduce applications.
- Yahoo Gridmix3 Hadoop cluster benchmarking from Yahoo engineer team.
- <u>Deeplearning4j Benchmarks</u>

Security

- Apache Ranger Central security admin & fine-grained authorization for Hadoop
- Apache Eagle real time monitoring solution
- Apache Knox Gateway single point of secure access for Hadoop clusters.
- Apache Sentry security module for data stored in Hadoop.
- <u>BDA</u> The vulnerability detector for Hadoop and Spark

System Deployment

- Apache Ambari operational framework for Hadoop mangement.
- <u>Apache Bigtop</u> system deployment framework for the Hadoop ecosystem.
- Apache Helix cluster management framework.
- Apache Mesos cluster manager.
- <u>Apache Slider</u> is a YARN application to deploy existing distributed applications on YARN.
- Apache Whirr set of libraries for running cloud services.
- <u>Apache YARN</u> Cluster manager.
- <u>Brooklyn</u> library that simplifies application deployment and management.
- <u>Buildoop</u> Similar to Apache BigTop based on Groovy language.
- <u>Cloudera HUE</u> web application for interacting with Hadoop.
- <u>Facebook Prism</u> multi datacenters replication system.

- Google Borg job scheduling and monitoring system.
- Google Omega job scheduling and monitoring system.
- Hortonworks HOYA application that can deploy HBase cluster on YARN.
- Marathon Mesos framework for long-running services.

Applications

- 411 an web application for alert management resulting from scheduled searches into Elasticsearch.
- Adobe spindle Next-generation web analytics processing with Scala, Spark, and Parquet.
- Apache Kiji framework to collect and analyze data in real-time, based on HBase.
- Apache Metron a platform that integrates a variety of open source big data technologies in order to offer a centralized tool for security monitoring and analysis.
- Apache Nutch open source web crawler.
- Apache OODT capturing, processing and sharing of data for NASA's scientific archives.
- Apache Tika content analysis toolkit.
- Argus Time series monitoring and alerting platform.
- Atlas a backend for managing dimensional time series data.
- <u>Countly</u> open source mobile and web analytics platform, based on Node.js & MongoDB.
- <u>Domino</u> Run, scale, share, and deploy models without any infrastructure.
- Eclipse BIRT Eclipse-based reporting system.
- <u>ElastAert</u> ElastAlert is a simple framework for alerting on anomalies, spikes, or other patterns of interest from data in ElasticSearch.
- <u>Eventhub</u> open source event analytics platform.
- <u>Hermes</u> asynchronous message broker built on top of Kafka.
- <u>HIPI Library</u> API for performing image processing tasks on Hadoop's MapReduce.
- Hunk Splunk analytics for Hadoop.
- Imhotep Large scale analytics platform by indeed.
- MADlib data-processing library of an RDBMS to analyze data.
- <u>Kapacitor</u> an open source framework for processing, monitoring, and alerting on time series data.
- <u>Kylin</u> open source Distributed Analytics Engine from eBay.
- <u>PivotalR</u> R on Pivotal HD / HAWQ and PostgreSQL.
- <u>Rakam</u> open-source real-time custom analytics platform powered by Postgresql, Kinesis and PrestoDB.
- Qubole auto-scaling Hadoop cluster, built-in data connectors.
- <u>Sense</u> Cloud Platform for Data Science and Big Data Analytics.

- <u>SnappyData</u> a distributed in-memory data store for real-time operational analytics, delivering stream analytics, OLTP (online transaction processing) and OLAP (online analytical processing) built on Spark in a single integrated cluster.
- <u>Snowplow</u> enterprise-strength web and event analytics, powered by Hadoop, Kinesis, Redshift and Postgres.
- SparkR R frontend for Spark.
- Splunk analyzer for machine-generated data.
- Sumo Logic cloud based analyzer for machine-generated data.
- <u>Talend</u> unified open source environment for YARN, Hadoop, HBASE, Hive, HCatalog & Pig.
- Warp query by example tool for big data (OS X app)

Search engine and framework

- Apache Lucene Search engine library.
- <u>Apache Solr</u> Search platform for Apache Lucene.
- <u>Elassandra</u> is a fork of Elasticsearch modified to run on top of Apache Cassandra in a scalable and resilient peer-to-peer architecture.
- <u>ElasticSearch</u> Search and analytics engine based on Apache Lucene.
- <u>Enigma.io</u> Freemium robust web application for exploring, filtering, analyzing, searching and exporting massive datasets scraped from across the Web.
- <u>Facebook Unicorn</u> social graph search platform.
- Google Caffeine continuous indexing system.
- <u>Google Percolator</u> continuous indexing system.
- <u>TeraGoogle</u> large search index.
- <u>HBase Coprocessor</u> implementation of Percolator, part of HBase.
- Lily HBase Indexer quickly and easily search for any content stored in HBase.
- <u>LinkedIn Bobo</u> is a Faceted Search implementation written purely in Java, an extension to Apache Lucene.
- <u>LinkedIn Cleo</u> is a flexible software library for enabling rapid development of partial, out-of-order and real-time typeahead search.
- LinkedIn Galene search architecture at LinkedIn.
- <u>LinkedIn Zoie</u> is a realtime search/indexing system written in Java.
- MG4J MG4J (Managing Gigabytes for Java) is a full-text search engine for large document collections written in Java. It is highly customisable, high-performance and provides state-of-the-art features and new research algorithms.
- Sphinx Search Server fulltext search engine.

MySQL forks and evolutions

Amazon RDS - MySQL databases in Amazon's cloud.

- <u>Drizzle</u> evolution of MySQL 6.0.
- Google Cloud SQL MySQL databases in Google's cloud.
- MariaDB enhanced, drop-in replacement for MySQL.
- MySQL Cluster MySQL implementation using NDB Cluster storage engine.
- <u>Percona Server</u> enhanced, drop-in replacement for MySQL.
- <u>ProxySQL</u> High Performance Proxy for MySQL.
- <u>TokuDB</u> TokuDB is a storage engine for MySQL and MariaDB.
- WebScaleSQL is a collaboration among engineers from several companies that face similar challenges in running MySQL at scale.

PostgreSQL forks and evolutions

- HadoopDB hybrid of MapReduce and DBMS.
- IBM Netezza high-performance data warehouse appliances.
- <u>Postgres-XL</u> Scalable Open Source PostgreSQL-based Database Cluster.
- RecDB Open Source Recommendation Engine Built Entirely Inside PostgreSQL.
- <u>Stado</u> open source MPP database system solely targeted at data warehousing and data mart applications.
- Yahoo Everest multi-peta-byte database / MPP derived by PostgreSQL.
- <u>TimescaleDB</u> An open-source time-series database optimized for fast ingest and complex queries
- <u>PipelineDB</u> The Streaming SQL Database. An open-source relational database that runs SQL queries continuously on streams, incrementally storing results in tables

Memcached forks and evolutions

- <u>Facebook McDipper</u> key/value cache for flash storage.
- Facebook Memcached fork of Memcache.
- <u>Twemproxy</u> A fast, light-weight proxy for memcached and redis.
- <u>Twitter Fatcache</u> key/value cache for flash storage.
- <u>Twitter Twemcache</u> fork of Memcache.

Embedded Databases

- <u>Actian PSQL</u> ACID-compliant DBMS developed by Pervasive Software, optimized for embedding in applications.
- <u>BerkeleyDB</u> a software library that provides a high-performance embedded database for key/value data.
- <u>HanoiDB</u> Erlang LSM BTree Storage.
- <u>LevelDB</u> a fast key-value storage library written at Google that provides an ordered mapping from string keys to string values.

- <u>LMDB</u> ultra-fast, ultra-compact key-value embedded data store developed by Symas.
- RocksDB embeddable persistent key-value store for fast storage based on LevelDB.

Business Intelligence

- <u>BIME Analytics</u> business intelligence platform in the cloud.
- Chartio lean business intelligence platform to visualize and explore your data.
- <u>datapine</u> self-service business intelligence tool in the cloud.
- GoodData platform for data products and embedded analytics.
- <u>Jaspersoft</u> powerful business intelligence suite.
- <u>Jedox Palo</u> customisable Business Intelligence platform.
- <u>Jethrodata</u> Interactive Big Data Analytics.
- Microsoft business intelligence software and platform.
- <u>Microstrategy</u> software platforms for business intelligence, mobile intelligence, and network applications.
- Pentaho business intelligence platform.
- Qlik business intelligence and analytics platform.
- <u>Redash</u> Open source business intelligence platform, supporting multiple data sources and planned queries.
- <u>Saiku</u> open source analytics platform.
- <u>SpagoBl</u> open source business intelligence platform.
- <u>Tableau</u> business intelligence platform.
- Zoomdata Big Data Analytics.
- Metabase The simplest, fastest way to get business intelligence and analytics to everyone in your company

Data Visualization

- Airpal Web UI for PrestoDB.
- AnyChart fast, simple and flexible JavaScript (HTML5) charting library featuring pure JS API.
- <u>Arbor</u> graph visualization library using web workers and jQuery.
- <u>Banana</u> visualize logs and time-stamped data stored in Solr. Port of Kibana.
- <u>Bloomery</u> Web UI for Impala.
- <u>Bokeh</u> A powerful Python interactive visualization library that targets modern web browsers for presentation, with the goal of providing elegant, concise construction of novel graphics in the style of D3.js, but also delivering this capability with highperformance interactivity over very large or streaming datasets.
- C3 D3-based reusable chart library
- <u>CartoDB</u> open-source or freemium hosting for geospatial databases with powerful front-end editing capabilities and a robust API.

- <u>chartd</u> responsive, retina-compatible charts with just an img tag.
- Chart.js open source HTML5 Charts visualizations.
- Chartist.js another open source HTML5 Charts visualization.
- <u>Crossfilter</u> JavaScript library for exploring large multivariate datasets in the browser. Works well with dc.js and d3.js.
- <u>Cubism</u> JavaScript library for time series visualization.
- Cytoscape JavaScript library for visualizing complex networks.
- <u>DC.js</u> Dimensional charting built to work natively with crossfilter rendered using d3.js. Excellent for connecting charts/additional metadata to hover events in D3.
- <u>D3</u> javaScript library for manipulating documents.
- <u>D3.compose</u> Compose complex, data-driven visualizations from reusable charts and components.
- <u>D3Plus</u> A fairly robust set of reusable charts and styles for d3.js.
- Echarts Baidus enterprise charts.
- Envisionis dynamic HTML5 visualization.
- FnordMetric write SQL queries that return SVG charts rather than tables
- Freeboard pen source real-time dashboard builder for IOT and other web mashups.
- <u>Gephi</u> An award-winning open-source platform for visualizing and manipulating large graphs and network connections. It's like Photoshop, but for graphs. Available for Windows and Mac OS X.
- Google Charts simple charting API.
- Grafana graphite dashboard frontend, editor and graph composer.
- Graphite scalable Realtime Graphing.
- <u>Highcharts</u> simple and flexible charting API.
- <u>IPython</u> provides a rich architecture for interactive computing.
- <u>Kibana</u> visualize logs and time-stamped data
- Lumify open source big data analysis and visualization platform
- Matplotlib plotting with Python.
- Metricsgraphic.js a library built on top of D3 that is optimized for time-series data
- NVD3 chart components for d3.js.
- <u>Peity</u> Progressive SVG bar, line and pie charts.
- <u>Plot.ly</u> Easy-to-use web service that allows for rapid creation of complex charts, from heatmaps to histograms. Upload data to create and style charts with Plotly's online spreadsheet. Fork others' plots.
- <u>Plotly.js</u> The open source javascript graphing library that powers plotly.
- Recline simple but powerful library for building data applications in pure Javascript and HTML.
- Redash open-source platform to query and visualize data.
- ReCharts A composable charting library built on React components
- Shiny a web application framework for R.
- <u>Sigma.js</u> JavaScript library dedicated to graph drawing.

- <u>Superset</u> a data exploration platform designed to be visual, intuitive and interactive, making it easy to slice, dice and visualize data and perform analytics at the speed of thought.
- <u>Vega</u> a visualization grammar.
- Zeppelin a notebook-style collaborative data analysis.
- Zing Charts JavaScript charting library for big data.

Internet of things and sensor data

- <u>Apache Edgent (Incubating)</u> a programming model and micro-kernel style runtime that can be embedded in gateways and small footprint edge devices enabling local, real-time, analytics on the edge devices.
- Azure IoT Hub Cloud-based bi-directional monitoring and messaging hub
- <u>TempolQ</u> Cloud-based sensor analytics.
- <u>2lemetry</u> Platform for Internet of things.
- Pubnub Data stream network
- ThingWorx Rapid development and connection of intelligent systems
- IFTTT If this then that
- <u>Evrything</u>- Making products smart

Interesting Readings

- Big Data Benchmark Benchmark of Redshift, Hive, Shark, Impala and Stiger/Tez.
- <u>NoSQL Comparison</u> Cassandra vs MongoDB vs CouchDB vs Redis vs Riak vs HBase vs Couchbase vs Neo4j vs Hypertable vs ElasticSearch vs Accumulo vs VoltDB vs Scalaris comparison.
- <u>Monitoring Kafka performance</u> Guide to monitoring Apache Kafka, including native methods for metrics collection.
- Monitoring Hadoop performance Guide to monitoring Hadoop, with an overview of Hadoop architecture, and native methods for metrics collection.

Interesting Papers

2015 - 2016

• 2015 - Facebook - One Trillion Edges: Graph Processing at Facebook-Scale.

2013 - 2014

• <u>2014</u> - **Stanford** - Mining of Massive Datasets.

- 2013 AMPLab Presto: Distributed Machine Learning and Graph Processing with Sparse Matrices.
- 2013 AMPLab MLbase: A Distributed Machine-learning System.
- 2013 AMPLab Shark: SQL and Rich Analytics at Scale.
- <u>2013</u> **AMPLab** GraphX: A Resilient Distributed Graph System on Spark.
- <u>2013</u> **Google** HyperLogLog in Practice: Algorithmic Engineering of a State of The Art Cardinality Estimation Algorithm.
- <u>2013</u> **Microsoft** Scalable Progressive Analytics on Big Data in the Cloud.
- 2013 Metamarkets Druid: A Real-time Analytical Data Store.
- 2013 Google Online, Asynchronous Schema Change in F1.
- 2013 Google F1: A Distributed SQL Database That Scales.
- 2013 Google MillWheel: Fault-Tolerant Stream Processing at Internet Scale.
- <u>2013</u> Facebook Scuba: Diving into Data at Facebook.
- 2013 Facebook Unicorn: A System for Searching the Social Graph.
- 2013 Facebook Scaling Memcache at Facebook.

2011 - 2012

- 2012 **Twitter** The Unified Logging Infrastructure for Data Analytics at Twitter.
- <u>2012</u> **AMPLab** Blink and It's Done: Interactive Queries on Very Large Data.
- <u>2012</u> **AMPLab** Fast and Interactive Analytics over Hadoop Data with Spark.
- 2012 AMPLab Shark: Fast Data Analysis Using Coarse-grained Distributed Memory.
- 2012 Microsoft Paxos Replicated State Machines as the Basis of a High-Performance Data Store.
- 2012 Microsoft Paxos Made Parallel.
- 2012 AMPLab BlinkDB: Queries with Bounded Errors and Bounded Response Times on Very Large Data.
- 2012 Google Processing a trillion cells per mouse click.
- <u>2012</u> **Google** Spanner: Google's Globally-Distributed Database.
- 2011 AMPLab Scarlett: Coping with Skewed Popularity Content in MapReduce Clusters.
- <u>2011</u> **AMPLab** Mesos: A Platform for Fine-Grained Resource Sharing in the Data Center.
- <u>2011</u> **Google** Megastore: Providing Scalable, Highly Available Storage for Interactive Services.

2001 - 2010

- 2010 Facebook Finding a needle in Haystack: Facebook's photo storage.
- 2010 AMPLab Spark: Cluster Computing with Working Sets.
- <u>2010</u> **Google** Pregel: A System for Large-Scale Graph Processing.

- <u>2010</u> **Google** Large-scale Incremental Processing Using Distributed Transactions and Notifications base of Percolator and Caffeine.
- 2010 Google Dremel: Interactive Analysis of Web-Scale Datasets.
- 2010 Yahoo S4: Distributed Stream Computing Platform.
- 2009 HadoopDB: An Architectural Hybrid of MapReduce and DBMS Technologies for Analytical Workloads.
- <u>2008</u> **AMPLab** Chukwa: A large-scale monitoring system.
- 2007 Amazon Dynamo: Amazon's Highly Available Key-value Store.
- 2006 Google The Chubby lock service for loosely-coupled distributed systems.
- 2006 Google Bigtable: A Distributed Storage System for Structured Data.
- 2004 Google MapReduce: Simplied Data Processing on Large Clusters.
- 2003 **Google** The Google File System.

Videos

Books

Streaming

- <u>Streaming Data</u> Streaming Data introduces the concepts and requirements of streaming and real-time data systems.
- <u>Storm Applied</u> Storm Applied is a practical guide to using Apache Storm for the realworld tasks associated with processing and analyzing real-time data streams.
- <u>Fundamentals of Stream Processing: Application Design, Systems, and Analytics</u> This comprehensive, hands-on guide combining the fundamental building blocks and emerging research in stream processing is ideal for application designers, system builders, analytic developers, as well as students and researchers in the field.
- <u>Stream Data Processing: A Quality of Service Perspective</u> Presents a new paradigm suitable for stream and complex event processing.
- <u>Unified Log Processing</u> Unified Log Processing is a practical guide to implementing a unified log of event streams (Kafka or Kinesis) in your business

Distributed systems

 <u>Distributed Systems for fun and profit</u> – Theory of distributed systems. Include parts about time and ordering, replication and impossibility results.

Data Visualization

- The beauty of data visualization
- Designing Data Visualizations with Noah Iliinsky
- Hans Rosling's 200 Countries, 200 Years, 4 Minutes

• <u>Ice Bucket Challenge Data Visualization</u>

Other Awesome Lists

- Other awesome lists <u>awesome-awesomeness</u>.
- Even more lists <u>awesome</u>.
- Another list? <u>list</u>.
- WTF! <u>awesome-awesome-awesome</u>.
- Analytics <u>awesome-analytics</u>.

Top-down learning path: Machine Learning for Software Engineers

https://github.com/ZuzooVn/machine-learning-for-software-engineers

What is it?

This is my multi-month study plan for going from mobile developer (self-taught, no CS degree) to machine learning engineer.

My main goal was to find an approach to studying Machine Learning that is mainly hands-on and abstracts most of the Math for the beginner. This approach is unconventional because it's the top-down and results-first approach designed for software engineers.

Please, feel free to make any contributions you feel will make it better.

Table of Contents

- What is it?
- Why use it?
- How to use it
- Follow me
- Don't feel you aren't smart enough
- About Video Resources
- Prerequisite Knowledge
- The Daily Plan
- Motivation
- Machine learning overview
- Machine learning mastery
- Machine learning is fun
- Inky Machine Learning
- Machine Learning: An In-Depth Guide
- Stories and experiences

- Machine Learning Algorithms
- Beginner Books
- Practical Books
- Kaggle knowledge competitions
- Video Series
- MOOC
- Resources
- Becoming an Open Source Contributor
- Games
- Podcasts
- Communities
- Conferences
- Interview Questions
- My admired companies

Why use it?

I'm following this plan to prepare for my near-future job: Machine learning engineer. I've been building native mobile applications (Android/iOS/Blackberry) since 2011. I have a Software Engineering degree, not a Computer Science degree. I have an itty-bitty amount of basic knowledge about: Calculus, Linear Algebra, Discrete Mathematics, Probability & Statistics from university. Think about my interest in machine learning:

- Can I learn and get a job in Machine Learning without studying CS Master and PhD?
 - o "You can, but it is far more difficult than when I got into the field." Drac Smith
- How do I get a job in Machine Learning as a software programmer who self-studies Machine Learning, but never has a chance to use it at work?
 - o "I'm hiring machine learning experts for my team and your MOOC will not get you the job (there is better news below). In fact, many people with a master's in machine learning will not get the job because they (and most who have taken MOOCs) do not have a deep understanding that will help me solve my problems." Ross C. Taylor
- What skills are needed for machine learning jobs?
 - "First, you need to have a decent CS/Math background. ML is an advanced topic so most textbooks assume that you have that background. Second, machine learning is a very general topic with many sub-specialties requiring unique skills. You may want to browse the curriculum of an MS program in Machine Learning to see the course, curriculum and textbook." Utility
 - o "Probability, distributed computing, and Statistics." <u>Hydrangea</u>

I find myself in times of trouble.

AFAIK, There are two sides to machine learning:

- Practical Machine Learning: This is about querying databases, cleaning data, writing scripts to transform data and gluing algorithm and libraries together and writing custom code to squeeze reliable answers from data to satisfy difficult and ill-defined questions. It's the mess of reality.
- Theoretical Machine Learning: This is about math and abstraction and idealized scenarios and limits and beauty and informing what is possible. It is a whole lot neater and cleaner and removed from the mess of reality.

I think the best way for practice-focused methodology is something like 'practice — learning — practice', that means where students first come with some existing projects with problems and solutions (practice) to get familiar with traditional methods in the area and perhaps also with their methodology. After practicing with some elementary experiences, they can go into the books and study the underlying theory, which serves to guide their future advanced practice and will enhance their toolbox of solving practical problems. Studying theory also further improves their understanding on the elementary experiences, and will help them acquire advanced experiences more quickly.

It's a long plan. It's going to take me years. If you are familiar with a lot of this already it will take you a lot less time.

How to use it

Everything below is an outline, and you should tackle the items in order from top to bottom.

I'm using Github's special markdown flavor, including tasks lists to check progress.

• Create a new branch so you can check items like this, just put an x in the brackets: [x]

More about Github-flavored markdown

Follow me

I'm a Vietnamese Software Engineer who is really passionate and wants to work in the USA.

How much did I work during this plan? Roughly 4 hours/night after a long, hard day at work.

I'm on the journey.

Twitter: <u>@Nam Vu</u>

USA as heck

Don't feel you aren't smart enough

I get discouraged from books and courses that tell me as soon as I open them that multivariate calculus, inferential statistics and linear algebra are prerequisites. I still don't know how to get started...

- What if I'm Not Good at Mathematics
- <u>5 Techniques To Understand Machine Learning Algorithms Without the Background in Mathematics</u>
- How do I learn machine learning?

About Video Resources

Some videos are available only by enrolling in a Coursera or EdX class. It is free to do so, but sometimes the classes are no longer in session so you have to wait a couple of months, so you have no access. I'm going to be adding more videos from public sources and replacing the online course videos over time. I like using university lectures.

Prerequisite Knowledge

This short section were prerequisites/interesting info I wanted to learn before getting started on the daily plan.

- What is the difference between Data Analytics, Data Analysis, Data Mining, Data Science, Machine Learning, and Big Data?
- Learning How to Learn
- Don't Break The Chain
- How to learn on your own

The Daily Plan

Each subject does not require a whole day to be able to understand it fully, and you can do multiple of these in a day.

Each day I take one subject from the list below, read it cover to cover, take notes, do the exercises and write an implementation in Python or R.

Motivation

• Dream

Machine learning overview

- A Visual Introduction to Machine Learning
- A Gentle Guide to Machine Learning
- Introduction to Machine Learning for Developers
- Machine Learning basics for a newbie
- How do you explain Machine Learning and Data Mining to non Computer Science people?
- Machine Learning: Under the hood. Blog post explains the principles of machine learning in layman terms. Simple and clear
- What is machine learning, and how does it work?
- Deep Learning A Non-Technical Introduction

wiachine learning mastery	
• □	The Machine Learning Mastery Method Machine Learning for Programmers Applied Machine Learning with Machine Learning Mastery Python Machine Learning Mini-Course Machine Learning Algorithms Mini-Course ine learning is fun
• □	Machine Learning is Fun! Part 2: Using Machine Learning to generate Super Mario Maker levels Part 3: Deep Learning and Convolutional Neural Networks Part 4: Modern Face Recognition with Deep Learning Part 5: Language Translation with Deep Learning and the Magic of Sequences Part 6: How to do Speech Recognition with Deep Learning Part 7: Abusing Generative Adversarial Networks to Make 8-bit Pixel Art Machine Learning
. =	Part 1: What is Machine Learning? Part 2: Supervised Learning and Unsupervised Learning ine Learning: An In-Depth Guide
	Overview, goals, learning types, and algorithms Data selection, preparation, and modeling Model evaluation, validation, complexity, and improvement Model performance and error analysis Unsupervised learning, related fields, and machine learning in practice es and experiences
	Machine Learning in a Week Machine Learning in a Year How I wrote my first Machine Learning program in 3 days Learning Path: Your mentor to become a machine learning expert You Too Can Become a Machine Learning Rock Star! No PhD How to become a Data Scientist in 6 months: A hacker's approach to career nning

Clida
o Slide
5 Skills You Need to Become a Machine Learning Engineer
 Are you a self-taught machine learning engineer? If yes, how did you do it & how
long did it take you?
How can one become a good machine learning engineer?
A Learning Sabbatical focused on Machine Learning
Machine Learning Algorithms
• 10 Machine Learning Algorithms Explained to an 'Army Soldier'
• Top 10 data mining algorithms in plain English
• 10 Machine Learning Terms Explained in Simple English
A Tour of Machine Learning Algorithms
The 10 Algorithms Machine Learning Engineers Need to Know
Comparing supervised learning algorithms
Machine Learning Algorithms: A collection of minimal and clean implementations
of machine learning algorithms
Beginner Books
• Data Smart: Using Data Science to Transform Information into Insight 1st Edition
Data Science for Business: What you need to know about data mining and data
analytic-thinking
analytic trilliang
Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die
Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books Machine Learning for Hackers
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books Machine Learning for Hackers GitHub repository(R)
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books Machine Learning for Hackers GitHub repository(R) GitHub repository(Python)
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books Machine Learning for Hackers GitHub repository(R) GitHub repository(Python) Python Machine Learning
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books Machine Learning for Hackers GitHub repository(R) GitHub repository(Python) Python Machine Learning GitHub repository
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books Machine Learning for Hackers GitHub repository(R) GitHub repository(Python) Python Machine Learning GitHub repository Programming Collective Intelligence: Building Smart Web 2.0 Applications
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books Machine Learning for Hackers GitHub repository(R) GitHub repository(Python) Python Machine Learning GitHub repository Programming Collective Intelligence: Building Smart Web 2.0 Applications
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books Machine Learning for Hackers GitHub repository(R) GitHub repository(Python) Python Machine Learning GitHub repository Programming Collective Intelligence: Building Smart Web 2.0 Applications Machine Learning: An Algorithmic Perspective, Second Edition
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books Machine Learning for Hackers GitHub repository(R) GitHub repository(Python) Python Machine Learning GitHub repository Programming Collective Intelligence: Building Smart Web 2.0 Applications Machine Learning: An Algorithmic Perspective, Second Edition GitHub repository Resource repository
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books Machine Learning for Hackers GitHub repository(R) GitHub repository(Python) Python Machine Learning GitHub repository Programming Collective Intelligence: Building Smart Web 2.0 Applications Machine Learning: An Algorithmic Perspective, Second Edition GitHub repository
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books Machine Learning for Hackers GitHub repository(R) GitHub repository(Python) Python Machine Learning GitHub repository Programming Collective Intelligence: Building Smart Web 2.0 Applications Machine Learning: An Algorithmic Perspective, Second Edition GitHub repository Resource repository Introduction to Machine Learning with Python: A Guide for Data Scientists
 Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die Practical Books Machine Learning for Hackers GitHub repository(R) GitHub repository(Python) Python Machine Learning GitHub repository Programming Collective Intelligence: Building Smart Web 2.0 Applications Machine Learning: An Algorithmic Perspective, Second Edition GitHub repository Resource repository Introduction to Machine Learning with Python: A Guide for Data Scientists GitHub repository

		 Slides for Chapters 6-8 (zip)
•		Machine Learning in Action
		o <u>GitHub repository</u>
•		Reactive Machine Learning Systems(MEAP)
		o <u>GitHub repository</u>
•		An Introduction to Statistical Learning
		o <u>GitHub repository(R)</u>
		o <u>GitHub repository(Python)</u>
		o <u>Videos</u>
•		Building Machine Learning Systems with Python
		o <u>GitHub repository</u>
•		<u>Learning scikit-learn: Machine Learning in Python</u>
	_	o <u>GitHub repository</u>
•		Probabilistic Programming & Bayesian Methods for Hackers
•		Probabilistic Graphical Models: Principles and Techniques
•		Machine Learning: Hands-On for Developers and Technical Professionals
		 Machine Learning Hands-On for Developers and Technical Professionals review
	_	o <u>GitHub repository</u>
•		<u>Learning from Data</u>
		o Online tutorials
•		Reinforcement Learning: An Introduction (2nd Edition)
		o <u>GitHub repository</u>
•		Machine Learning with TensorFlow(MEAP) o GitHub repository
		o <u>GitHub repository</u>
(ag	g	e knowledge competitions
•		Kaggle Competitions: How and where to begin?
•		How a Beginner Used Small Projects To Get Started in Machine Learning and
	<u>Co</u>	empete on Kaggle
•		Master Kaggle By Competing Consistently
/id	eo	Series
•		Machine Learning for Hackers
•		Fresh Machine Learning
•		Machine Learning Recipes with Josh Gordon
•		Everything You Need to know about Machine Learning in 30 Minutes or Less
•		A Friendly Introduction to Machine Learning
•		Nuts and Bolts of Applying Deep Learning - Andrew Ng
•		BigML Webinar
		o Video

		o <u>Resources</u>
•		mathematicalmonk's Machine Learning tutorials
•		Machine learning in Python with scikit-learn
		o <u>GitHub repository</u>
		o <u>Blog</u>
•		My playlist – Top YouTube Videos on Machine Learning, Neural Network & Deep
	Lea	<u>arning</u>
•		16 New Must Watch Tutorials, Courses on Machine Learning
•		<u>DeepLearning.TV</u>
•		Learning To See
•		Neural networks class - Université de Sherbrooke
•		21 Deep Learning Videos, Tutorials & Courses on Youtube from 2016
•		30 Top Videos, Tutorials & Courses on Machine Learning & Artificial Intelligence
	fro	<u>om 2016</u>
•		Practical Deep Learning For Coders
	00	C
•		edX's Introduction to Artificial Intelligence (AI)
•		Udacity's Intro to Machine Learning
	_	 <u>Udacity Intro to Machine Learning Review</u>
•		<u>Udacity's Supervised, Unsupervised & Reinforcement</u>
•		Machine Learning Foundations: A Case Study Approach
•		Coursera's Machine Learning
		o <u>Video only</u>
		o Coursera Machine Learning review
	_	o <u>Coursera: Machine Learning Roadmap</u>
•		Machine Learning Distilled
•		BigML training
•		Coursera's Neural Networks for Machine Learning
		o Taught by Geoffrey Hinton, a pioneer in the field of neural networks
•		Machine Learning - CS - Oxford University
•		Creative Applications of Deep Learning with TensorFlow
•		Intro to Descriptive Statistics
•		Intro to Inferential Statistics
•		6.S094: Deep Learning for Self-Driving Cars
•		6.S191: Introduction to Deep Learning
•		Coursera's Deep Learning
_		

Resources

• Learn Machine Learning in a Single Month

•	The Non-Technical Guide to Machine Learning & Artificial Intelligence
•	Best practices rule book for Machine Learning engineering from Google
•	Machine Learning for Software Engineers on Hacker News
•	Machine Learning for Developers
•	Machine Learning for Humans 😈 📀
•	Machine Learning Advice for Developers
•	Machine Learning For Complete Beginners
•	Getting Started with Machine Learning: For absolute beginners and fifth graders
•	How to Learn Machine Learning: The Self-Starter Way
•	Machine Learning Self-study Resources
•	Level-Up Your Machine Learning
•	An Honest Guide to Machine Learning
•	Enough Machine Learning to Make Hacker News Readable Again
	o <u>Video</u>
	o <u>Slide</u>
•	Dive into Machine Learning
•	{Machine, Deep} Learning for software engineers
•	Deep Learning For Beginners
•	Foundations for deep learning
•	Machine Learning Mindmap / Cheatsheet
•	Machine Learning courses in Universities
	o L Stanford
	 Machine Learning Summer Schools Oxford
	Flipboard Topics
•	o Machine learning
	o <u>Deep learning</u>
	o Artificial Intelligence
•	Medium Topics
	Machine learning
	o Deep learning
	 Artificial Intelligence
•	Monthly top 10 articles
	Machine Learning
	• <u>July 2016</u>
	• August 2016
	• <u>September 2016</u>
	 October 2016
	November 2016
	<u>Year 2016</u>
	o Algorithms

- September 2016
- October-November 2016
- Comprehensive list of data science resources
- <u>DigitalMind's Artificial Intelligence resources</u>
- Awesome Machine Learning
- CreativeAi's Machine Learning

Games

- Halite: A.I. Coding Game
- Vindinium: A.I. Programming Challenge
- General Video Game Al Competition
- Angry Birds Al Competition
- The Al Games
- Fighting Game AI Competition
- CodeCup
- Student StarCraft Al Tournament
- AIIDE StarCraft AI Competition
- CIG StarCraft AI Competition
- CodinGame Al Bot Games

Becoming an Open Source Contributor

•	tensorflow/magenta: Magenta: Music and Art Generation with Machine
	<u>Intelligence</u>
•	tensorflow/tensorflow: Computation using data flow graphs for scalable machine
	<u>learning</u>
•	cmusatyalab/openface: Face recognition with deep neural networks.
•	tensorflow/models/syntaxnet: Neural Models of Syntax.

Podcasts

Podcasts for Beginners:

- Talking Machines
- o Linear Digressions
- o Data Skeptic
- This Week in Machine Learning & Al

"More" advanced podcasts

- Partially Derivative
- o O'Reilly Data Show
- Not So Standard Deviation

Podcasts to think outside the box:

Data Stories

Communities

- Quora
 - o Machine Learning
 - Statistics
 - o Data Mining
- Reddit
 - o Machine Learning
 - Computer Vision
 - o Natural Language
 - o Data Science
 - o Big Data
 - Statistics
- Data Tau
- Deep Learning News
- KDnuggets

Conferences

- Neural Information Processing Systems (NIPS)
- International Conference on Learning Representations (ICLR)
- Association for the Advancement of Artificial Intelligence (AAAI)
- IEEE Conference on Computational Intelligence and Games (CIG)
- IEEE International Conference on Machine Learning and Applications (ICMLA)
- International Conference on Machine Learning (ICML)

Interview Questions

How To Prepare For A Machine Learning Interview

•		40 Interview Questions asked at Startups in Machine Learning / Data Science
•		21 Must-Know Data Science Interview Questions and Answers
•		Top 50 Machine learning Interview questions & Answers
•		Machine Learning Engineer interview questions
•		Popular Machine Learning Interview Questions
•		What are some common Machine Learning interview questions?
•		What are the best interview questions to evaluate a machine learning researcher?
•		Collection of Machine Learning Interview Questions
•		121 Essential Machine Learning Questions & Answers
y	ac	dmired companies

M

• ELSA - Your virtual pronunciation coach