Contents

Introduction

Number of topics Number of resources [Reddit](#) [Subscribe](#)

Machine Learning My Way (MLMW) is a list of topics and selected links for you to plan your studies or identify and fill up your knowledge gaps.

Topics covered include probability, statistics, econometrics, machine learning, data quality, model productization as well as society and business outcomes of ML adoption.

Get the MLMW guide:

- [view online](#)
- download [recent PDF](#)
- subscribe to the [mailing list](#) for the newest PDF version.

You may also want to follow MLMW updates on [Reddit](#) or [Telegram](#).

image

Beginner track

Where do I start?

My suggestion for beginners is getting two free textbooks for classic machine learning (ML) and enrolling into a video course for deep learning (DL):

- Mathematics for Machine Learning (MML) by Marc Peter Deisenroth, A. Aldo Faisal and Cheng Soon Ong;
- [Introduction to Statistical Learning with Python (ISLP)](#) by Gareth James, Daniela Witten, Trevor Hastie and Rob Tibshirani;
- Deep Learning Specialization by Andrew Ng.

Links to other textbooks and supplementary materials are provided below.

Roadmap

```
Math   ML      DL                Subfields
----   -----   --------------    ----------------------------

MML -> ISLP -> deeplearning.ai -> NLP and Transformers
         |                     -> CV
         |                     -> Agents, robotics and RL
         |                     -> Tabular data and time series
       Practical manuals:
       - scipy lectures (free)
       - Muller, Geron or Burkov (paid)
```

Prerequisites

You will need a working knowledge of Python and ability to operate with mathematical concepts from linear algebra and calculus.

Core path
1. Check you math knowledge with Mathematics for Machine Learning (MML) Part 1.
2. Read chapter 8 "When Models Meet Data" in MML for introduction to machine learning.
3. Proceed to [Introduction to Statistical Learning with Python (ISLP)](#) textbook.
4. Read from `scikit-learn` documentation about neural network models.
5. Start Andrew Ng Deep Learning Specialization.

Reference texts

There are more dense textbooks than ISLP or Andrew Ng course that you can use as references.

For classic machine learning they are [Bishop (2006)](#) and [Murphy (2022)](#).

For deep learning there are several open textbooks:

- [Ian Goodfellow, Yoshua Bengio and Aaron Courville (2016). Deep Learning Book (DLB)](#)
- [Aston Zhang, Zack Lipton, Mu Li and Alex Smola (2021). Dive into Deep Learning (d2l).](#)
- [Simon Prince (2023). Understanding Deep Learning (UDL).](#)

DLB (2016) is a reference text that enjoys [a continious stream of citations](#), while d2l and UDL are newer and keep updating code and content.

[^1]: This open-source book represents our attempt to make deep learning approachable, teaching readers the concepts, the context, and the code. The entire book is drafted in Jupyter notebooks, seamlessly integrating exposition figures, math, and interactive examples with self-contained code. Our goal is to offer a resource that could (i) be freely available for everyone; (ii) offer sufficient technical depth to provide a starting point on the path to actually becoming an applied machine learning scientist; (iii) include runnable code, showing readers how to solve problems in practice; (iv) allow for rapid updates, both by us and also by the community at large; (v) be complemented by a forum for interactive discussion of technical details and to answer questions. [[arxiv abstract](#)]

What else?

You can supplement the core path above with the following:

- introductory courses on probability (like [P4D](#) or [Seeing Theory](#)) and [statistics](#)
- [scipy lectures](#) is an underappreciated resource by the authors on foundational `scikit-learn` package themselves;
- [Python Data Science Handbook](#), including [a chapter on machine learning](#) and [DSML](#) textbook;
- practical books that combine machine learning concepts and programming practice – either of [Müller](#), [Géron](#) or [Burkov](#) (all are paid editions);
- review subfields of machine learning like natural language processing (NLP), computer vision (CV), reinforcement learning (RL) or robotics;
- for popular attention and transformer architectures check out talks by [Andrej Karpathy](#) and his [NanoGPT](#) sample code;
- familiarize with approaches to data modelling in econometrics (eg. chapter 1, 2, 4, 5 from [Econ 1630](#))
- glance at data analysis vocabulary as endorsed by the industry leaders like [Google](#), [Mathworks](#), [H2O](#) and [NVIDIA](#)
- last but not least [StatQuest videos by Josh Starmer](#) and [3Blue1Brown videos by Grant Sanderson](#) make great study companions.

Python packages

Machine learning:

- [scikit-learn](#), see also [Gaël Varoquaux 2023 interview](#).

Deep learning:

- [PyTorch](#) (most popular)
- [TensorFlow](#) and [Keras](#) (easier to learn)

Does reading these materials make you a machine learning engineer?

Not until you make projects for real tasks on real data (where textbook examples do not play out that well).

Not in scope

This page puts no recommendation for various skills that are also important for a quantitative modeller or an engineer:

- Python programming, Linux and cloud computing;
- data processing, pipelines and model productisation;
- experiment design and iterative workflows;
- advanced topics in statistics and machine learning;
- domain knowledge, business sense and outcomes of ML adoption.

Please refer to larger [MLMW guide](#) for coverage of these topics.

Preliminaries

Linear algebra

Calculus

Probability

Beginner
- [Seeing Theory](#) visual textbook.
- [Probability distributions](#).

Courses
- [Statistics 110 (Probability) by Joe Blitzstein](#).

Reference
- [Chapter 6 "Probability and Distributions"](#) in the MML book.
- [P4D](#).

Advanced
- [Aubrey Clayton](#) reading of [The Logic of Science by E.T. Jaynes](#).

Statistics

MLMW Guide

Models and methods

Data types, sources and quality

From research design to model productization

Software tools

Business change and society impact

Appendix