

Machine Learning

Benno Stein

Theo Lettmann

Michael Völske

Martin Potthast

Contents

I. Introduction

II. Machine Learning Basics

III. Linear Models

IV. Neural Networks

V. Support Vector Machines

VI. Decision Trees

VII. Bayesian Learning

VIII. Learning Theory

[IX. Deep Learning]

X. Ensemble Methods and Meta Learning

XI. Reinforcement Learning

Literature

Machine Learning:

- ❑ T. Mitchell.
Machine Learning
1st edition, McGraw-Hill, 1997.
www.cs.cmu.edu/~tom/mlbook.html
- ❑ C.M. Bishop. [[Interview 2018 @ MS Research](#)]
Pattern Recognition and Machine Learning
2nd edition, Springer 2006.
www.microsoft.com/en-us/research/people/cmbishop/prml-book/
- ❑ T. Hastie, R. Tibshirani, J. Friedman.
The Elements of Statistical Learning
2nd edition, Springer, 2009.
hastie.su.domains/ElemStatLearn/ (2017)
- ❑ I. Goodfellow, Y. Bengio, A. Courville.
Deep Learning
MIT Press, 2016.
deeplearningbook.org

Software

Programming:

- ❑ The Jupyter Project.
JupyterHub
jupyter.org
- ❑ Microsoft Corporation.
Visual Studio Code
code.visualstudio.com
- ❑ JetBrains, Inc.
PyCharm IDE
www.jetbrains.com/pycharm

Machine Learning:

- ❑ *NumPy*
numpy.org
- ❑ *scikit-learn: Machine Learning in Python*
scikit-learn.org

Digital Assistance

Learning:

- ❑ Webis.
Infobot
infobot.webis.de

Curation: ML4NLP

Part	Unit	Section
Bayesian Learning	Probability Basics	 Approaches to probability 4, 14–18, 30–32  Conditional probability 36–41
Introduction	Examples of Learning Tasks	 Learning tasks 13–17
	Specification of Learning Tasks	 Learning paradigms 19–22  Real world $\rightarrow$ model 28–41  LMS algorithm 42–45
	Elements of Machine Learning	 ML stack: LMS 48–53  Discriminative vs. generative 58–61
Machine Learning Basics	Rule-Based Learning	 Learning hypothesis 22–23
	From Regression to Classification	 One-dimensional feature space 69–72, 77  Regression for classification 84–99
	Evaluating Effectiveness	 Misclass. rate 117–126  Label noise, Bayes error 144–147, 151–155  Error bounds 172–175
		 Holdout error 179–183
Linear Models	Logistic Regression	 Logistic regression 4–20, 26–33  BGD$_{\sigma}$ algorithm 34–35  ML stack: BGD$_{\sigma}$ 38–42
		 Loss computation 49–60
	Overfitting and Regularization	 Overfitting 66–85  Regularization 89–98
	Gradient Descent in Detail	 Gradient descent 123–134

Curation: ML4NLP (continued)

Part	Unit	Section
Neural Networks	Perceptron Learning	 Biological inspiration 1–9  PT algorithm 12–13, 18–29
	Multilayer Perceptron	 Linear separability 55–63  Unrestricted classification 75–76
	Advanced MLPs	 $H(P, Q)$ Cross entropy 126–131  Softmax 133–137
Deep Learning Sequence	Recurrent Neural Networks	 RNN sequence encoding 1–30  RNN sequence decoding 1–40
	RNNs for Machine Translation	 Machine translation 1–17, 19–40
	Transformer Architecture	 Transformer architecture 1–36

The road to Large Language Models (LLMs):

1. Logistic regression
2. Perceptron with sigmoid activation
3. Multilayer perceptron (MLP)
4. Recurrent neural networks (RNN)
5. Statistical machine translation (SMT)
6. SMT with RNNs
7. Transformer (attention, position encoding)