GP-OML course 'Machine Learning'

Lecturers: Dr. Rui Jorge Almeida (Maastricht University) and Prof. dr. Inneke Van Nieuwenhuyse (Hasselt University)

Dear students,

In this document you can find a number of materials, selected by the instructors of the GP-OML course Machine Learning, which allow you to learn the basics of a number of methods at your own pace. They are intended to offer a "first step" to learn about the methods, without necessarily having to follow the course, for which waiting lists can be long.

In our view, there is a distinct difference between the learning goals of the materials we selected here, and the learning goals of the course. The materials posted on this page are aimed to help you find out "what a method is about", and what are possible advantages/disadvantage in view of our own research. The course allows you to get deeper insights on/hands-on experience with the methods, and uses a flipped classroom format for that purpose (requiring students to prepare exercises/read materials as a preparation for the class discussions).

We encourage all students having an interest in Machine Learning to take a look at these materials, and to register for the course in case you want to take the next step!

Module 1: Gaussian Process Regression

The material consists of 2 weblectures (+/- 1 hr each) and 1 MATLAB package for Gaussian Process Regression. We also suggest 1 book (for advanced study).

Weblecture 1: Gaussian Processes (Part 1): Prof. J. Cunningham, Washington University: https://www.youtube.com/watch?v=BS4Wd5rwNwE

This lecture teaches you the basics of GPs, in view of GP Regression.

- Very useful intro up to min 20: link between GP and MVN distribution
- 20-24: what if data have homogenous noise?
- 24-..: conditioning, making predictions using data already observed

Weblecture 2: Gaussian Processes (Part 2): Prof. J. Cunningham, Washington University: https://www.youtube.com/watch?v=KcB8c3a4LYU

This weblecture covers some already quite advanced topics, which will likely be useful only for those students needing these extensions for their own research.

- Nonstationary kernels
- Building your own kernel
- Vanaf 18: GP classification, comparison with SVM
- Vanaf 28: shortcomings of GPs

Matlab package: DACE package: free download at http://www.omicron.dk/dace.html

This package offers you ready-made codes for GP regression. The nice thing is that it contains a documentation pdf, which allows you to work through an example exercise by yourself. The downside:

the codes *assume that the data have no noise*. For data with noise, the codes are thus ill-suited: see also the discussion in the course, where we go deeper into handling data with heterogenous noise.

Book: Advanced study: http://www.gaussianprocess.org/gpml/

Module 2: Introduction to Machine Learning and Statistical Learning

There are many great books on the subject of machine/statistical learning. A few selected ones are (* indicates freely available):

- Mitchell, T. M. (1997). Machine Learning. New York: McGraw-Hill. ISBN: 978-0-07-042807-2.
- (*) Hastie, T., Tibshirani, R., Friedman, J. (2001). The Elements of Statistical Learning II. New York, NY, USA: Springer New York Inc.. ISBN 978-0-387-84858-7. (ESLII)
- (*) Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. (2013). An introduction to statistical learning : with applications in R. New York: Springer. ISBN: 978-1461471370. (ISLR)
- (*) Goodfellow, I., Bengio, Y. Courville, A. (2016). Deep Learning. MIT Press. ISBN: 978-0-262-035613. www.deeplearningbook.org .

The ISLR book also has labs in R with clear instructions on how to apply the different methods, and these labs are available in Python, for example at https://github.com/JWarmenhoven/ISLR-python/ or https://github.com/mscaudill/IntroStatLearn .

Introductory material to machine learning and statistical learning is:

- Ch. 1, Machine Learning (T. Mitchell) and/or Ch. 5.1 Deep learning (Goodfellow).
- Ch. 1, 2 Introduction to Statistical Learning and/or Ch. 1, 2 Elements of Statistical Learning II.

Participants are encouraged to read the introduction of these different books, since they contain different but overlapping views on the problem of learning by machines using data. The Mitchell and Goodfellow books introduce the problem from a perspective of learning by computers, while ISLR and ESLII take a statistical perspective.

Module 3: Tree based methods

For an overview of tree based methods, we suggest the weblecture on

- **Decision Trees and Ensemble Methods** part of Stanford CS229: Machine Learning (Autumn 2018) by Raphael Townshend - https://www.youtube.com/watch?v=wr9gUr-eWdA

Another great resource are the series of weblectures by Prof. Kilian Q. Weinberger from Cornell University for the course CS4780/CS5780: Machine Learning for Intelligent Systems

- Decision Trees - https://www.youtube.com/watch?v=E1_WCdUAtyE
- Decision Trees - https://www.youtube.com/watch?v=a3ioGSwfVpE
- Bagging - https://www.youtube.com/watch?v=0LB1cy2sCXc
- Random Forests - https://www.youtube.com/watch?v=4EOCQJgqAOY
- Boosting - https://www.youtube.com/watch?v=dosOtgSdbnY
- Boosting - https://www.youtube.com/watch?v=Vd6hzcwEa2k
- Boosting - https://www.youtube.com/watch?v=toOAToTaGV4

Suggested reading material are:

- Ch. 9.2, 10, 15, 16, Elements of Statistical Learning II
- Ch. 8, Introduction to Statistical Learning
- Ch. 3, Machine Learning (T.Mitchell)

Module 4: Neural Networks

For an overview on Neural Networks and Deep Learning, we suggest **one of** the following series of weblectures:

- Neural Networks / Deep Learning - Cornell CS4780 SP17 by Prof. Kilian Weinberger - https://www.youtube.com/watch?v=kPXxbmBsFxs
- Neural Networks / Deep Learning Continued - Cornell CS4780 SP17 by Prof. Kilian Weinberger - https://www.youtube.com/watch?v=zmu9wR2c7Z4
- Neural Networks / Deep Learning - Cornell CS4780 SP17 by Prof. Kilian Weinberger - https://www.youtube.com/watch?v=wqpIohLjsAY

or

- Introduction to Neural Networks | Stanford CS229: Machine Learning (Autumn 2018) by Kian Katanforoosh - https://www.youtube.com/watch?v=MfIjxPh6Pys
- Backprop & Improving Neural Networks | Stanford CS229: Machine Learning (Autumn 2018) by Kian Katanforoosh - https://www.youtube.com/watch?v=zUazLXZZA2U

or

- Backpropagation, Neural Networks 1, Stanford CS231n Winter 2016 by Andrej Karpathy - https://www.youtube.com/watch?v=i94OvYb6noo
- Neural Networks Part 2, Stanford CS231n Winter 2016 by Andrej Karpathy - https://www.youtube.com/watch?v=gYpoJMlgyXA
- Neural Networks Part 3 / Intro to ConvNets, Stanford CS231n Winter 2016 by Andrej Karpathy - https://www.youtube.com/watch?v=hd_KFJ5ktUc

Suggested reading material are:

- Ch. 11, Elements of Statistical Learning II
- Ch. 4, Machine Learning (T. Mitchell)
- Ch. 5, Pattern Recognition and Machine Learning (Bishop) ISBN: 978-0-387-31073-2
- Ch. 6, 7, Deep Learning (Goodfellow)

Other reading material for programming (Deep) Neural Networks in Python or R are:

- Chollet, F. and Allaire, J.J., 2018. Deep Learning with R. ISBN 9781617295546
- Francois, C., 2017. Deep learning with Python. ISBN 9781617294433

Module 4: Unsupervised methods

A few topics on different unsupervised methods can be found in the following weblectures:

- Clustering part of MIT OpenCourseWare by Prof. John Guttag https://www.youtube.com/watch?v=esmzYhuFnds
- Generalization, K-means | Stanford CS221: AI (Autumn 2019) by Prof. Percy Liang - https://www.youtube.com/watch?v=O9vCb30sHBA
- K-means, GMM, and EM Stanford CS229 | Summer 2019 by Prof. Anand Avati https://www.youtube.com/watch?v=LmpkKwsyQj4
- Factor Analysis & ELBO Stanford CS229 | Summer 2019 by Prof. Anand Avati https://www.youtube.com/watch?v=pA-bo8_HNy4
- Principal & Independent CA Stanford CS229 | Summer 2019 by Prof. Anand Avati https://www.youtube.com/watch?v=7pJ6XNvpO8M

Suggested reading material are:

- Clustering:
  - Ch. 13.1-13.2, 14.3, Elements of Statistical Learning II
  - Ch. 10.1, 10.3, Introduction to Statistical Learning

  - Ch. 7, Data Mining: Concepts and Techniques (Jiawei Han) ISBN: 978-1-55860-901-3
- Association rules
  - Ch. 14.2, Elements of Statistical Learning II
  - Ch. 5, Data Mining: Concepts and Techniques (Jiawei Han) ISBN: 978-1-55860-901-3
- Principal Components, Curves and Surfaces
  - Ch. 14.2, Elements of Statistical Learning II
  - Ch. 10.2, Introduction to Statistical Learning

Module 5: What is next?

We encourage all students having an interest in Machine Learning to register for the course in case you want to take the next step!

If you prefer to continue to explore by yourself, all of the courses linked (e.g. cs229, cs230) are invaluable resources. Some of these courses have complete syllabus that you can follow. Combining this with one (or several) of the books suggested in this document, and there is quite a lot of material to explore and have fun with.