Description of the course

Lecturer: Georgios P. Karagiannis — georgios.karagiannis@durham.ac.uk

The following details concern the module Machine Learning and Neural Networks III (MATH3431) in Epiphany term. The description below is informal and aims at helping students organize their study. The official description of the course can be found in

https://www.dur.ac.uk/faculty.handbook/module_description/?year=2022&module_code=MATH3431.

1 Description of the course

Aim

This course provides an introduction to neural network, kernel, and latent models, as well as stochastic learning algorithms in the machine learning framework. We will gain the necessary background for the design, theory, and practical implementation of the aforesaid concepts.

Intended learning outcomes

The students will be able to:

[ILO1] Explain the foundations and theoretical basis of statistical machine learning concepts, as well as possible challenges.

[ILO2] Setup, apply, compare, and extend appropriate statistical model for machine learning in both the classical and Bayesian framework.

[ILO3] Explain, apply, and extend appropriate training/computational approaches.

[ILO4] Explain, extend, and apply theoretical aspects in machine learning

[ILO5] Use appropriate software to facilitate machine learning procedures

Requirements / preparation

A well prepared student aiming to attend this course is expected to have a good understanding of calculus, linear algebra, probability, and statistical inference (classical and Bayesian), as well as in R programming with fluency in seeking information about packages/routines from help or CRAN online resources.

Teaching and learning activities

[TLA1] Lectures

Students will be introduced to the theory, and be exposed to a small number of examples.

- Major focus [ILOs 1-4]

[TLA3] Computer practicals

Students will learn how to implement the introduced methods in a programming language, use existing routines in R related to the introduced concepts.

- Major focus [ILOs 2, 3, & 4]

[TLA4] Office hours

Students will ask further questions. When coming to the office, students are requested to have their questions written down in a piece of paper.

- Major focus [ILO 1-5]

Assessment activities

Formative assessment

[FA1] Four homework assignments will be assigned regularly. The homework sheet will contain a number of problems which have to be assessed and returned. Homework problems and solutions will be available from Blackboard Ultra. The submission of the solution will be done Gradescope. Feedback will be given via Gradescope and emails. Major focus [ILOs 1-5].

Summative assessment

[SA1] ILOs 1-5 will be assessed in written and computer based examinations.

2 Syllabus

Convex/non-convex learning problems [ILO2]:

- Convexity, Lipschitz, loss function, risk functions.

Stochastic learning [ILO1-5]:

- Gradient descent, Stochastic gradient descent, and methodological variations; Stochastic gradient Langevin dynamic algorithms

Support vector machine [ILO1-5]:

- Hard SVM, Soft SVM, duality, Relevant vector machine

Kernel methods & projections in feature space [ILO1-5]:

- The kernel trick; extensions to support vector machine and relevant vector machine

Neural networks [ILO1-5]:

- Feed-forward neural network (Single/multi-layer perceptron, radial basis functions, error functions, relation to other models); Bayesian/classical framework; back-propagation; vanishing gradient problem

Gaussian process regression [ILO1-5]:

- the model; relations to other models

If time allows, we can introduce [ILO1-5]:

- Latent models: Mixture models; mixture of experts; Expectation Maximization algorithm
- Variational inference ; Expectation Propagation
- Learning theory ; concentration inequalities

3 Reading list

Following is a comprehensive list of references that cover all the concepts discussed in the course. However it is possible for some details introduced in lectures to be available in articles available from the library. In such cased, references will be given in the corresponding handouts.

Main texts:

- Bishop, C. M. (2006). Pattern recognition and machine learning. New York: Springer.
  - It is a classical textbook in machine learning (ML) methods. It discusses all the concepts introduced in the course (not necessarily in the same depth). It is one of the main textbooks in the module. The level on difficulty is easy.
  - Students who wish to have a textbook covering traditional concepts in machine learning are suggested to get a copy of this textbook. It is available online from the Microsoft's website `https://www.microsoft.com/en-us/research/publication/pattern-recognition-machine-learning/`
- Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.
  - It has several elements of theory about machine learning algorithms. It is one of the main textbooks in the module. The level on difficulty is advanced as it requires moderate knowledge of maths.
- Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford university press.
  - It is a classical textbook about 'traditional' artificial neural networks (ANN). It is very comprehensive (compared to others) and it goes deep enough for the module although it may be a bit outdated. It is one of the main textbooks in the module for ANN. The level on difficulty is moderate.

Supplementary textbooks:

- Ripley, B. D. (2007). Pattern recognition and neural networks. Cambridge university press.
  - A classical textbook in artificial neural networks (ANN) that also covers other machine learning concepts. It contains interesting theory about ANN.
  - It is suggested to be used as a supplementary reading for neural networks as it contains a few interesting theoretical results. The level on difficulty is moderate.
- Williams, C. K., & Rasmussen, C. E. (2006). Gaussian processes for machine learning (Vol. 2, No. 3, p. 4). Cambridge, MA: MIT press.
  - A classic book in Gaussian process regression (GPR) that covers the material we will discuss in the course about GPR. It can be used as a companion textbook with that of (Bishop, C. M., 2006). The level on difficulty is easy.
- Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.

  - A popular textbook in machine learning methods. It discusses all the concepts introduced in the module. It focuses more on the probabilistic/Bayesian framework but not with great detail. It can be used as a comparison textbook for brief reading about ML methods just to see another perspective than that in (Bishop, C. M., 2006). The level on difficulty is easy.

- Murphy, K. P. (2022). Probabilistic machine learning: an introduction. MIT press.
  - A textbook in machine learning methods. It covers a smaller number of ML concepts than (Murphy, K. P., 2012) but it contains more fancy/popular topics such as deep learning ideas. It is suggested to be used in the same manner as (Murphy, K. P., 2012). The level on difficulty is easy.

- Barber, D. (2012). Bayesian reasoning and machine learning. Cambridge University Press.
  - A textbook in machine learning methods from a Bayesian point of view. It discusses all the concepts introduced apart from ANN and stochastic gradient algorithms. It aims to be more 'statistical' than those of Murphy and Bishop. The level on difficulty is easy.

- Vapnik, V. (1999). The nature of statistical learning theory. Springer science & business media.
  - Important textbook in the statistical machine learning theory. To have have an in deep understanding about statistical learning, one has to read it together with the textbook of Shalev-Shwartz, S., & Ben-David, S. (2014)

- Devroye, L., Györfi, L., & Lugosi, G. (2013). A probabilistic theory of pattern recognition (Vol. 31). Springer Science & Business Media.
  - Theoretical aspects about machine learning algorithms. The level on difficulty is advanced as it requires moderate knowledge of probability.