

MASTERS SYLLABUS

2487/1920 – Machine Learning, 7 ECTS

Semester 2 (S2)

2019/20

« Qiwei Han »

« qiwei.han@novasbe.pt »

TBD

COURSE UNIT AIMS. (Purpose of the course using broad, general terms)

This course will be an introduction to machine learning techniques and how to use them to help solve business problems. This course is designed for management, economics and finance students who are interested in learning modern, scalable, computational data analysis methods also known as machine learning, and apply them to business and social problems.

This is a hands-on course where students will be expected to use Python to implement solutions to various business problems. Prior experience with Python is required, and students should be familiar with basics Python data science libraries such as numpy, pandas, matplotlib, seaborn. Other programming knowledge, such as R, Matlab, Java, etc would be highly preferred.

COURSE UNIT CONTENT. (Main topics covered in the course)

- Overview of the Machine Learning Process
 1. Understand Problem
 2. Map to Machine Learning formulation
 3. Machine Learning Concepts
 4. End-to-end Machine Learning Pipeline Development
 1. Setup the problem
 2. Feature Development
 3. Modeling
 4. Evaluation
 5. Deployment
- Machine Learning Methods
 1. Supervised
 1. Logistic Regression
 2. KNN
 3. Trees
 4. Random Forests
 5. Ensemble Methods
 2. Unsupervised
 1. Clustering

- 2. Dimensionality reduction
- 3. Deep Learning
 - 1. Neural networks
 - 2. Convolutional neural networks
 - 3. Recurrent neural networks
 - 4. Other advances in the field
- 4. Business applications
 - 1. Recommender systems
 - 2. Uplift modeling
- Feature creation, feature engineering and feature selection
- Interpretability and transparency
- Ethics, Fairness and Bias

LEARNING OBJECTIVES. Upon completion of this course, students should be able to:

A. Knowledge and Understanding

This course will teach students:

- **What role Machine Learning can play in designing, implementing, evaluating, and improving business operations.**
- **How Machine Learning methods work, how to use them, and how to building machine learning pipelines/systems.**
- **How to tackle business problems using Machine Learning methods and tools**
- **How to address key concerns when designing machine learning systems for business applications**

B. Subject-Specific Skills

Students will have hands-on experience using popular programming languages such as Python to perform machine learning tasks for real business problems.

C. General Skills

Student will be able to describe and apply machine learning concepts, steps, tools and technologies that are used in today's data science projects. They will also identify critical issues associated with the storage, extraction and manipulation of data.

DEMONSTRATION OF THE COHERENCE OF THE SYLLABUS WITH COURSE UNIT AIMS

Students will learn essential machine learning concepts and apply practical machine learning techniques to manage data effectively for business analytics problems.

TEACHING AND LEARNING METHODS.

Students are required to bring own laptops for in-class exercises and quizzes. This course adopts learning-by-doing culture that allows students to implement machine learning pipeline through programming in Python. Most of class material will be in the Jupyter notebooks to facilitate reproducible practices.

ASSESSMENT

The overall evaluation of performance consists of 4 parts

- Class participation through 6 quizzes (10%)
- 4 tri-weekly assignment (20%)
- Course project (30%)
- Final exam (40%)

Students need to participate in class quizzes for at least 5 times. If students are present in all quizzes, 5 out of 6 quizzes with highest points will be counted.

Assignments are issued every three weeks. Students need to submit the assignment by the due date and will lose 20% of points for each late day. For example, if the assignment is late for 3 days, student may get 40% of points at maximum. If the assignment is late for 6 days, assignment will be returned without evaluation.

Assignment questions should be sent at least 24 hours before the deadline to assure a timely response. Please CC the TA so that we can all stay coordinated, and include the course code 2489 in the subject line of your emails. Again, don't forget to CC the TA. They are the homework graders and I will usually defer to their judgement on matters of scoring.

Students need to propose a course project using public dataset. Students are responsible to structure the analysis and decide what machine learning techniques to be used. The course project will be discussed with and evaluated by the instructor in the mid-term and final week.

You are encouraged to discuss general approaches and clarification questions with your fellow students. However, you should do your homework yourself.

- Do not copy another student's homework.
- Do not copy from another student's homework.

If you receive any help from another student or outside the class (such as stackoverflow or other forums or websites), you must give credit where credit is due, and clearly identify where you received help. The expectation is that your grade must reflect the work that you alone did.

BIBLIOGRAPHY.

This course does not require any textbook, because data science is a rapidly changing field and no textbook may cover all materials we will teach in the course. However, the following book is recommended for your reference:

Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

Machine Learning: The Art and Science of Algorithms that Make Sense of Data

Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython 2nd Edition

Python Data Science Handbook Essential Tools for Working with Data

Hands-on Machine Learning with scikit-learn, Keras, and Tensorflow

RESOURCES.

The online resources that provides additional information of this course are quite sufficient. Below students may find the following resources that are useful for self-study and exercises:

Online Data Science Encyclopedia

<https://github.com/bulutyzilim/awesome-datascience>

Online Python tutorial:

<https://pythonprogramming.net/python-fundamental-tutorials/>

<http://pandas.pydata.org/pandas-docs/stable/tutorials.html>

Online machine learning library documentation

<https://scikit-learn.org/stable/index.html>

<https://www.tensorflow.org/tutorials>

<https://keras.io/>

<https://pytorch.org/tutorials/>