



ADVANCED MACHINE LEARNING AND KNOWLEDGE DISCOVERY

12 CFU - 1° e 2° semestre

Docenti titolari dell'insegnamento

VINCENZA CARCHIOLO - Modulo ADVANCED MACHINE LEARNING - INF/01 - 6 CFU

Email: vincenza.carchiolo@unict.it

Edificio / Indirizzo: Edificio 3 - piano V - stanza 11

Telefono: 095 7382359

Orario ricevimento: martedì e venerdì dalle ore 11 alle ore 12

CONCETTO SPAMPINATO - Modulo KNOWLEDGE DISCOVERY - ING-INF/05 - 6 CFU

Email: cspampin@dieei.unict.it

Edificio / Indirizzo: DIEEI, Plesso 13, stanza 8, Cittadella Universitaria

Telefono: 095/7382057

Orario ricevimento: Su prenotazione via email

OBIETTIVI FORMATIVI

▪ ADVANCED MACHINE LEARNING

The module will focus on the implementations of various machine learning techniques and their applications in various domains. The primary tools used in the class are the Python programming language and several associated libraries.

▪ KNOWLEDGE DISCOVERY

This module covers the fundamental concepts of deep learning methods and how to use them for extracting, modelling and visualizing the learned knowledge.

Topics include: neural networks with backpropagation, convolutional neural networks, recurrent neural networks, methods for representation learning, and how to use them under different learning regimes (supervised, unsupervised and reinforcement learning) and in variety of real-world applications ranging from computer vision, machine translation and medical image analysis.

The learning objectives are:

- to understand and use the main methodologies and techniques for learning from data
- to understand the main methodologies to design and implement neural networks for real-world applications
- to understand how to extract and learn knowledge in scenarios when supervision cannot be provided
- to understand and foresee the reliability of machine learning methods in operational

scenarios.

Knowledge and understanding

- To understand the main concepts of learning from data
- To understand concepts and tools for building intelligent systems using supervision and no supervision
- To understand the most important machine learning and artificial intelligence methodologies and techniques used by industries to make sense of data in order to support the decision process
- To understand what are the most appropriate techniques to be used in different real-world applications

Applying knowledge and understanding

- To be able to effectively understand and use the main tools for creating, loading and manipulating datasets.
- To design and implement from scratch a machine learning system following application-derived constraints in terms of modelling and data
- To understand proper benchmarks and baselines and analysing achieved results and their generalization in real-world applications
- To be able to apply methodologies and techniques to analyse data.

MODALITÀ DI SVOLGIMENTO DELL'INSEGNAMENTO

▪ **ADVANCED MACHINE LEARNING**

Lectures, hands-on exercises, paper reading, student presentations and seminars

Should teaching be carried out in mixed mode or remotely, it may be necessary to introduce changes with respect to previous statements, in line with the programme planned and outlined in the syllabus.

▪ **KNOWLEDGE DISCOVERY**

The main teaching methods are as follows:

- Lectures, to provide theoretical and methodological knowledge of the subject;
- Hands-on exercises, to provide “problem solving” skills and to apply design methodology;
- Laboratories, to learn and test the usage of related tools.
- Paper reading and presentations to enhance understanding of the core concepts
- Seminars by renowned experts (from both universities and industries) in the field to understand the current state of the art.

Should teaching be carried out in mixed mode or remotely, it may be necessary to introduce changes with respect to previous statements, in line with the programme planned and outlined in the syllabus.

PREREQUISITI RICHIESTI

▪ **ADVANCED MACHINE LEARNING**

Python programming language, Linear Algebra

▪ **KNOWLEDGE DISCOVERY**

Python programming language, statistical learning basic concepts

FREQUENZA LEZIONI

▪ **ADVANCED MACHINE LEARNING**

Strongly recommended. Attending and actively participating in the classroom activities will contribute positively towards the overall assessment of the oral exam.

▪ **KNOWLEDGE DISCOVERY**

Strongly recommended. Attending and actively participating in the classroom activities will contribute positively towards the overall assessment of the final exam.

CONTENUTI DEL CORSO

▪ **ADVANCED MACHINE LEARNING**

Introduction to the Course

- Introduction to Machine Learning
- Review of data Characteristics of Data and Preparation and Preprocessing

Supervised Learning

- Classification and Prediction using K-Nearest-Neighbor
- Classifying with Probability Theory; Naïve Bayes
- Building Decision Trees
- Regression models
- Evaluating predictive models
- Ensemble Models: Bagging and Boosting

Unsupervised Learning

- Clustering using K-Means
- Hierarchical Clustering
- Association Rule discovery
- Principal Component Analysis and Dimensionality Reduction

- Singular Value Decomposition

Brief note on Advance Topics

- Matrix Factorization
- Support Vector Machines
- Search and Optimization Techniques
- Markov models; time series analysis, sequential pattern mining

Real application domains

- Text Mining and document analysis/filtering
 - Content analysis, TFxIDF transformation, text categorization, document clustering
- Recommender systems
 - Neighborhood methods (user- and item-based)
 - Matrix factorization
 - Marketing and finance data analysis
- Laboraroty activity
 - Brief Python review
 - The package Numpy, Pandas , matplotlib and seaborn
 - Scikit-learn: a machine learning library for Python
 - Classification, Regression, Clustering, Dimensionality Reduction, Model Selection, Preprocessing

▪ **KNOWLEDGE DISCOVERY**

The KD module consists of two parts: the first one will be addressing the general and modern techniques based on deep learning paradigm to create KD systems from data, while the second one on how to extract, represent and visualize knowledge from data and trained models.

Part I: Methods and Architectures

Neural Networks and Backpropagation

- Derivatives and Gradient Descent
- Neural Network Representation, Gradient descent for Neural Networks
- Forward and Back Propagation
- The revolution of depth: deep learning

- Optimization algorithms: Mini-batch gradient descent, Exponentially weighted average, Gradient descent with momentum, RMSprop, Adam optimization algorithm, Learning rate decay
- Training aspects of deep learning: Regularization, Dropout, Normalizing inputs, Vanishing / Exploding gradients, Weight Initialization for Deep Networks

Convolutional Neural Networks

- Foundations: padding, strided convolution, dilation, 2D and 3D convolution, pooling
- State of the art models: AlexNet, ResNets, DenseNets, Inception
- Transfer Learning and Data Augmentation

Recurrent Neural Networks

- LSTM and variants
- Attention mechanisms

Part II: Knowledge Discovery from Data and Models

Unsupervised Learning with Deep Networks

- Representation and Feature Learning
- Autoencoders and Variational Autoencoders
- Generative Adversarial Networks

Reinforcement Learning

- Introduction to Reinforcement Learning
- Policy Gradients
- Actor-Critic Algorithms
- Value Function Methods
- Deep RL with Q-functions

Deep Learning Frameworks:

- Overview of the most used DL frameworks
- PyTorch and Jupyter Notebooks

Applications:

- Computer vision
- Medical Image Analysis
- Machine translation

TESTI DI RIFERIMENTO

▪ ADVANCED MACHINE LEARNING

1. Introduction to Machine Learning, Fourth Edition, By Ethem Alpaydin, MitPress ISBN: 9780262043793. 2020

2. Python Data Science Essentials - Third Edition by Alberto Boschetti, Luca Massaron, Packt Publishing, ISBN: 9781789537864, 2020

3. Teaching materials and reading paper list provided by the instructor

▪ **KNOWLEDGE DISCOVERY**

1. Deep Learning. I. Goodfellow, Y. Bengio and A. Courville, MIT Press, 2016

2. Programming PyTorch for Deep Learning, I. Pointer, O'Reilly Media

3. Teaching materials and reading paper list provided by the instructor

ALTRO MATERIALE DIDATTICO

▪ **ADVANCED MACHINE LEARNING**

All teaching material will be published both on Studium and on Microsoft Team

▪ **KNOWLEDGE DISCOVERY**

All teaching material will be published both on Studium and on www.perceivelab.com/teaching

PROGRAMMAZIONE DEL CORSO

ADVANCED MACHINE LEARNING

	Argomenti	Riferimenti testi
1	Introduction to Machine Learning	1,3
2	Python review	3
3	Pandas, Numpy, Matplotlib	2,3
4	Classification and Prediction: K-Nearest-Neighbor	1
5	Classification and Naive Bayes	1
6	Decision Tree	1,3
7	Regression models	1
8	Evaluating predictive models	1
9	Ensemble Models: Bagging and Boosting	1
10	Clustering using K-Means	1
11	Hierarchical Clustering	1
12	Association Rule discovery	1

13	Dimensional reduction	1
14	Singular Value Decomposition	1
15	Advanced topic	1,3
16	Sckit learn	3
17	ML in NLP	3
18	ML for Reccomender System	3

KNOWLEDGE DISCOVERY

	Argomenti	Riferimenti testi
1	Neural networks: derivatives, gradient descent, back-propagation	3
2	Deep Learning: basic concepts, optimization algorithms, training procedures	1, 3
3	Convolutional Neural Networks	1,3
4	Recurrent Neural Networks	1,3
5	Unsupervised Learning with Deep Networks: Representation and Feature Learning	1, 3
6	Autoencoders and Variational Autoencoders	1, 3
7	Generative Adversarial Networks	1, 3
8	Reinforcement Learning: Deep Q-Networks and Policy Gradient	3
9	Deep Learning Frameworks: PyTorch and Jupyter Notebooks	2, 3

VERIFICA DELL'APPRENDIMENTO

MODALITÀ DI VERIFICA DELL'APPRENDIMENTO

▪ **ADVANCED MACHINE LEARNING**

There will be one assignment and one final exam. The assignments will contain written questions that require some Python programming. The final exam consists a final assignment and an oral discussion concerning all course material.

The final assignment concerns comparative analysis on a given problem that must be presented in a final report and discussed in an oral discussion. The vote on the advanced machine learning module will account for 40% of the total grade for the entire course.

The grading policy for the AML module is:

- 40%: Final assignments
- 20% Intermediate assignments

- 40%: Oral discussion

Learning assessment may also be carried out on line, should the conditions require it.

- **KNOWLEDGE DISCOVERY**

The final exam consists of the development of a project in Pytorch, addressing one of the topics discussed during classes, together with a final report (structured as a scientific paper) discussing motivation, models, datasets and results used in the project.

The exam is evaluated according to the ability to create a deep learning model from scratch for extracting and learning knowledge from data on a given real-world problem, to understand how to properly measure its performance and to motivate the devised solutions.

The vote on the knowledge discovery module will account for 50% of the total grade for the entire course.

The module also foresees intermediate assignments. These assignments (between two to four) include: a) python scripts to solve simple basic learning problems on datasets discussed during with the instructor in order to avoid overlap and b) quizzes to verify the correct understanding of the presented techniques.

The grading policy for the KD module is:

- 50%: Final project
- 35%: Programming assignments
- 15%: Quizzes

Learning assessment may also be carried out on line, should the conditions require it.

ESEMPI DI DOMANDE E/O ESERCIZI FREQUENTI

- **ADVANCED MACHINE LEARNING**

Examples of questions and exercises are available on the Studium platform and on the course website.

- **KNOWLEDGE DISCOVERY**

Examples of questions and exercises are available on the Studium platform and on the course website.
