

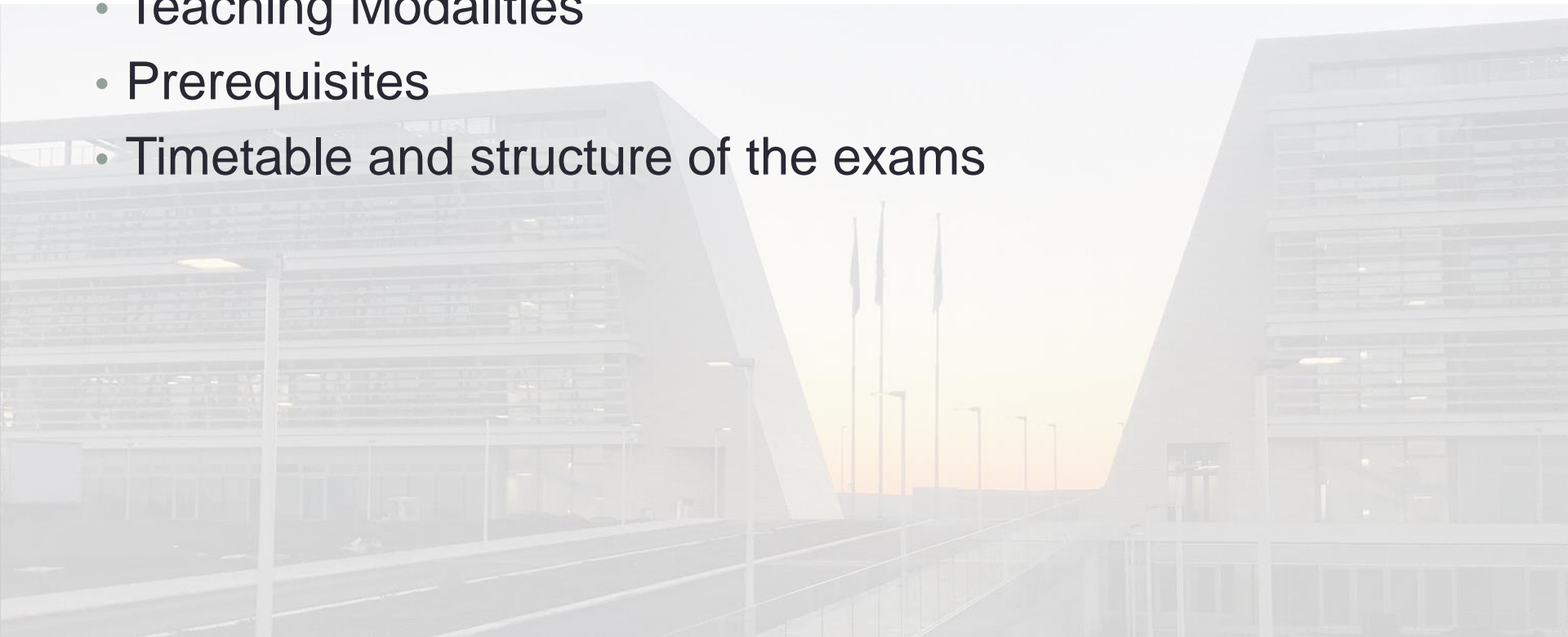
CORSO DI *WEB MINING E RETRIEVAL* *- INTRODUZIONE AL CORSO -*

Corsi di Laurea in Informatica,
Ing. Gestionale, Ing. Informatica, Ing. di Internet
(a.a. 2022-2023)

Roberto Basili

Overview

- WM&R: Objectives and Perspective of the course
- Teaching Modalities
- Prerequisites
- Timetable and structure of the exams



Objectives

- Methods for the accurate access and processing of the information distributed in Web sources
- Foundations:
 - Advanced models of *Machine Learning* and *Deep Learning*
 - *Statistical Learning Theory*
 - *Kernel Machines*
 - *Artificial Neural Networks e Deep Learning*
 - Text Search and Classification
 - Advanced Natural Language Processing
- Applications:
 - Multimedia Data Mining: Machine Vision & Data Analytics
 - *Opinion Mining* & Sentiment Analysis
 - *Textual Inference*, *Semantic Search*, Question Answering
 - Fake News Detection, Data Journalism

Organization

Section I: Advanced **Machine Learning & Deep Learning**

Intro ML. Supervised & Unsupervised Learning

Teoria Statistica dell'Apprendimento e Kernel-based learning.

Support Vector Machines.

Deep Learning with Neural Networks.

Approcci probabilistici: HMMs

Section II: Neural Language Processing

Neural Model for Language Processing. Large Scale Language Models.

Textual Inference. Question Answering.

• Section III: Web Mining and Deep Learning Applications

- Visual Object Recognition and Automatic Image Captioning

- Information Extraction. Fake News Detection.

- Sentiment Analysis, Brand Reputation Analysis and Marketing.

Lessons

- Foundational aspects
 - Paradigms and Methods for Machine Learning
 - Optimization: *Feature Selection, Dimensionality Reduction, Statistical Learning Theory*
 - Complex Models for Web Data Management
- Laboratory Lessons
 - Introduction to technologies, systems and tools
 - ML and DL software, NLP software, ML toolkits
 - Final Project
 - Advanced Analytics applications
 - ML for IR and knowledge discovery
 - Natural Language Processing applications (e.g. Visual QA)

Target

- Laurea Magistrale in Informatica (o Ing. Inf., Aut., Int., Gest.)
- Prerequisites:
 - Elements of Calculus (Analisi Matematica) and Geometry
 - Knowledge of Logic and Knowledge Representation
 - Probability Theory and Statistics
 - Data and Knowledge-based Systems
- Some Topics are *shared* with other courses:
 - *Machine Learning* (G. Gambosi)
 - *Information Retrieval* (D. Croce)
 - *Natural Language Processing* (F.M. Zanzotto)

Timetable

- **Monday 14:00 - 16:00 AULA 13**
Macroarea di Scienze
- **Wednesday 9:30 - 11:30 AULA B14**
Macroarea di Ingegneria
- **Thursday 11:30 - 13:30 AULA 13**
Macroarea di Scienze

Meeting with Students:

every Thursday after the lesson or,
on individual demands, *on-line* (MS Teams)

Web page

- Didattica Web (Corso di Laurea in Informatica)
- URL:
- http://sag.art.uniroma2.it/didattica/basili/WmlR_22_23/



Course Official Web Page

Web Mining e Retrieval - Deep Learning (a.a. 2022/23) Secondo Semestre

[Esci dai Frame](#)



Elenco dei File nel deposito



Sommario Contenuti

1. [Novita'](#)
2. [Programma del Corso](#)
3. [Testi di Riferimento](#)
4. [Link Utili](#)
5. [Diapositive delle lezioni](#) 🍷
6. [Progetti ed Esercizi Proposti](#) 🍷

Textbooks

- Deep Learning, Ian Goodfellow, Yoshua Bengio, Aaron Courville, MIT Press, 2016.
- IR Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze, Cambridge University Press. 2008. <http://nlp.stanford.edu/IR-book/>
- Complementary Materials:
 - Pattern Recognition and Data Mining, C. Bishop, 2006.
 - Articles and tutorial:
 - Scientific Papers
 - Lesson slides: http://sag.uniroma2.it/basili/didattica/WmlR_22_23/

WEB MINING & RETRIEVAL, a.a. 2022-23		
Sections	Main Topics	Lessons
Introduzione agli elementi di base del ML	Supervised and Unsupervised methods. Generative vs. Discriminative Methods.	Introduction to WM&R.
		Introduction to ML. Supervised vs. Unsupervised Methods. Probabilistic and Generative Methods
		Discriminative Methods: the role of metric spaces. Metrics and semantic similarity.
Clustering and Probabilistic methods	Introduction to clustering algorithms. Generative Models. HMMs	Unsupervised Machine Learning. Clustering. Agglomerative methods: K-means
From PAC learnability to SVM	PAC learnability. VC-dimension. SVMs. Kernel methods	PAC Learnability. Perceptron
		SVM. Hard Margin.
		Soft margin SVM. La Nozione di Kernel.
		Kernel polinomiali e RBF. Sequence & Tree Kernels.
Neural Networks and Deep Learning	Semi-supervised learning: ensemble methods, active learning, EM. On-line learning: Passive-Aggressive. Deep Neural network architectures.	From neural networks to deep learning: perceptrons and MLP
		Deep Learning over MLPs.
		Convolutional Neural Networks: Adopting Convolutional Neural Networks on images,
		Recurrent Neural Networks
		Large Scale Neural Language Models: Probabilistic Language Modeling, HMM.
		Large Scale Neural Language Models: Bengio et al model
		Attention Mechanisms and Transformers
Neural Learning and NLP	Dimensionality Reduction for IR/ML/NLP. Distributional Semantic models for IR. Word Embeddings with NNs.	Richiami ai metodi di Elaborazione del Linguaggio Naturale: il TAL
		Vector Spaces & NLP: from bow to embeddings
		Latent Semantic Kernels & Semantic Kernels.
		Wordspaces. Word Embeddings through Neural Networks.
		Machine Learning for: Natural Language Inference
Social Media Analytics (*)	IR in Social Media. Community detection. User profiling and Recommending. Sentiment and Emotion Analysis.	Opinion Mining e Sentiment Analysis: the task
		Introduzione all'OM & SA: Twitter as a case study
		Fake News Detection: FEVER and other stories
		Social Media Analysis: Recommending
		Social Media Analysis: Forecasting & Brand Reputation

Organisations: Exams

- *Modality*
 - *Full (9 ECTS)*
 - *Informatica, Laurea Tecnologie di Internet*
 - *Light (6 ECTS)*
 - *Laurea Ing. Informatica, Automatica e Gestionale*

Esami Light (6 ECTS)

- 2 MidTerm tests (intorno al 20 April, 10 June)
- Final Test (around 10 June, repeated in July)
 - Sections I, II and III
- Project, not mandatory (auxiliary credits?)
 - Software Project (e.g., Web or mobile app)
 - Deep Learning for Image Processing
 - Deep Learning for Fake News Detection
 - Machine Learning for Voice mediated Robotic Interfacesor
 - Bibliographic survey

Full Exam (9 ECTS)

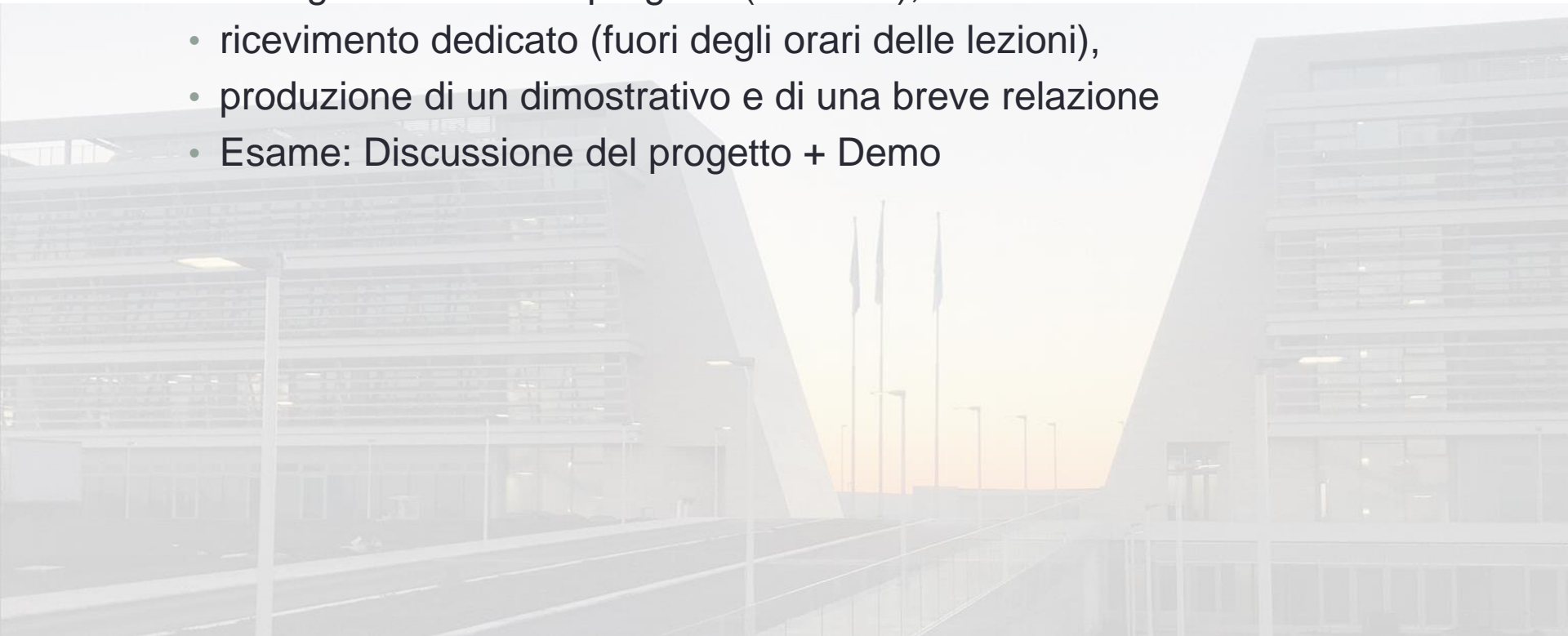
- 2 MidTerm tests (around April 20, June 10)
- 1 Final Test (around June 10, repeated in July)
- Project, mandatory
 - Software Project (e.g., Web or mobile app)
 - Deep Learning for Image Processing
 - Deep Learning for Fake News Detection
 - Machine Learning for Voice mediated Robotic Interfacesor
 - Bibliographic survey

Esami (1)

- **Approfondimento Bibliografico**
 - assegnamento di un tema e della relativa bibliografia,
 - ricevimento dedicato (fuori degli orari delle lezioni),
 - produzione di una relazione finale (stile Tesi),
 - Esame: presentazione finale della tesina (poss. entro la fine del Corso)

Esami (2)

- **Progetto** (max 2/3 persone)
 - Assegnamento di un progetto (stile tesi),
 - ricevimento dedicato (fuori degli orari delle lezioni),
 - produzione di un dimostrativo e di una breve relazione
 - Esame: Discussione del progetto + Demo



Esempi di Progetti

- **Opinion Tracker.** Riconoscimento e tracking di opinioni in social networks sulla base di descrizioni tematiche
- **Multimedia Search Engine.** Enterprise search for picture and video repositories
- **Graph Neural Networks.** Algoritmi graph-based per machine learning e ragionamento automatico.
- **Automatic Metadata creator.** Titolazione automatica di immagini e video
- **Fake News Detection.**
- **Zero-Shot Learning for domain specific classification**

Questions?



Action List

- Please register to the Course on Delphi :
 - URL: <https://delphi.uniroma2.it/totem/jsp/>
- Please use the field Note» to communicate your Laurea Degree and the number of ECTS foreseen by your curriculum
- **Regularly** access the Course Web page for:
 - Slides and teaching materials (*in progress*)
 - Timetables and Scheduling changes
 - Laboratory and Project Topics