Machine Learning

Notes of Lect. 1: Introduction

Alessio Micheli

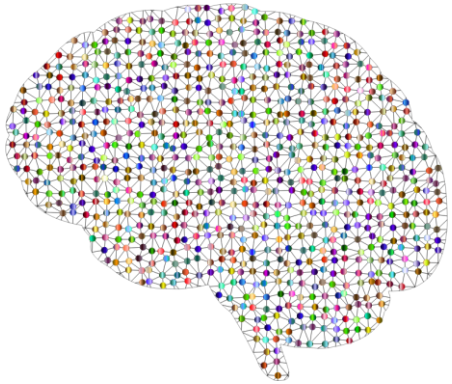
micheli@di.unipi.it

Dipartimento di Informatica
Università di Pisa - Italy

Computational Intelligence & Machine Learning Group

Sept, 2025

Welcome



About ML

Dip. Informatica
University of Pisa

- **Machine Learning (ML)**
  - *Master Programme in Computer Science*
  - *Master Programme in Data Science and Business Informatics*
  - *Master Programme in Digital Humanities*
- **Code: 654AA Credits (ECTS): 9 Semester: 1**
- **Lecturer: Alessio Micheli:** micheli@di.unipi.it
- **Material: Moodle https://elearning.di.unipi.it/**
  - Self-enrolling mechanism
- **Recordings of lectures autumn 25: Teams platform**
  - Connect ONLY through the App "Didactic Agenda" ("Agenda Didattica") for 654AA 25/26
  - See also Moodle: «FAQ and general information»
  - Where are the recordings? ML Teams: File (tab) → Recordings

Practical information

Dip. Informatica
University of Pisa

In class:

- Please, max **silence** during the lecture (to avoid noise) *recording in progress!* (of course, you can make questions)

Official Decrees: Only in-presence class (no streaming):

- <All the lessons of the semester will be delivered "**in person**">, see the general site https://didattica.di.unipi.it/en/master-programme-in-computer-science/timetable-master-computer-science/
- Enrolling students' mechanism for attendance "in presence" **and to connect to Teams:** through the App "Didactic Agenda" ("**agenda didattica**") https://agendadidattica.unipi.it/
  - Informally: You can follow the recording or also the streaming on-line via Teams (but without guaranteed interactions in this case)
  - Note that sometime recording is *not* possible due to tech. issues

Myself & Aims

Dip. Informatica
University of Pisa

- Prof. Alessio Micheli, Dept. of Computer Science, University of Pisa
  - PhD, Full professor of CS, AI, ML
  - Head of CIML@UniPi (Computational Intelligence & Machine Learning Group): More than 25 years of research in ML/AI
  - *Elected President* of the European Neural Network Society, annual ICANN
  - Past co-chair and co-founder of the IEEE Task Force on Reservoir Computing
  - National coordinator of the ML&DM group (including more than 40 research groups) within the AI*IA, annual workshop since 2012
  - Director of the Pisa unit for the CINI-AIIS National Lab (Artificial Intelligence and Intelligent Systems)
  - Associate Editor of the two main journals in ML/Neural Networks (IEEE TNNLS & Neural Networks - Elsevier)
- **My aim for this course**: introduce you, in 3 months, to the current ML, also retracing the basics and history

This Lecture

A. **What is ML, and why to study ML**: a motivational introduction
B. **Interlude**: ML in the curricula and other master degrees (not shown)
C. **ML Course info**: prerequisites, course structure, Q&A, bibliography, exam, math background

What is ML

A first look

Prologue: Learning

Dip. Informatica
University of Pisa

Learning: universal principles for living beings, society, and ….machines

*The problem of **learning** is arguably at the very core of the problem of **intelligence**, both biological and artificial*

Poggio, Shelton, *AI Magazine 1999*

What is ML? First view

Dip. Informatica
University of Pisa

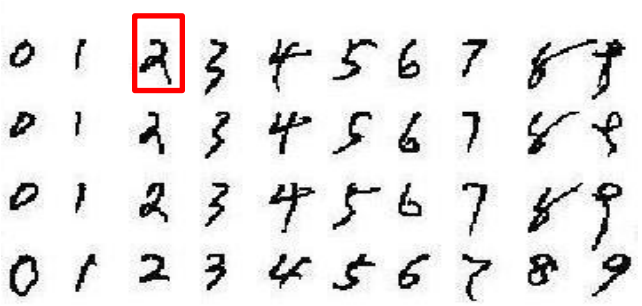
- In Computer Science, theoretical and applicative field called:
  - Apprendimento Automatico *in italiano (it)*
  - Machine Learning (ML) *English and literature* (aka learning systems)
- **Machine Learning** has emerged as an area of research combining the aims of creating *computers that could learn* (AI) and new *powerful adaptive/statistical tools* with rigorous foundation in computational science
- Machines that *learn* by themself. Why? Luxury or necessity?
  - Growing availability and need for analysis of empirical data *Central/methodological role due to changing of paradigm in science:* ***data-driven***
  - Difficult to provide adaptivity/intelligence by programming (A. Turing)
    → *learning* as the only choice to provide *intelligence* into the systems …
    - Pave the way to a new AI era (Growing data+ HPC+ ML flexibility)

What is ML? Simple examples

Dip. Informatica
University of Pisa

- Automatized learning by the system of the experience (set of examples) to address a computational task

Email spam classification



Character/face/speech recognition

… no (or poor) prior knowledge/rules for the solution
but it is (more) easy to have a source of
training experience (data with known results)

Zoom: Handwritten Digits Recognition: Learn a classifier

Dip. Informatica
University of Pisa

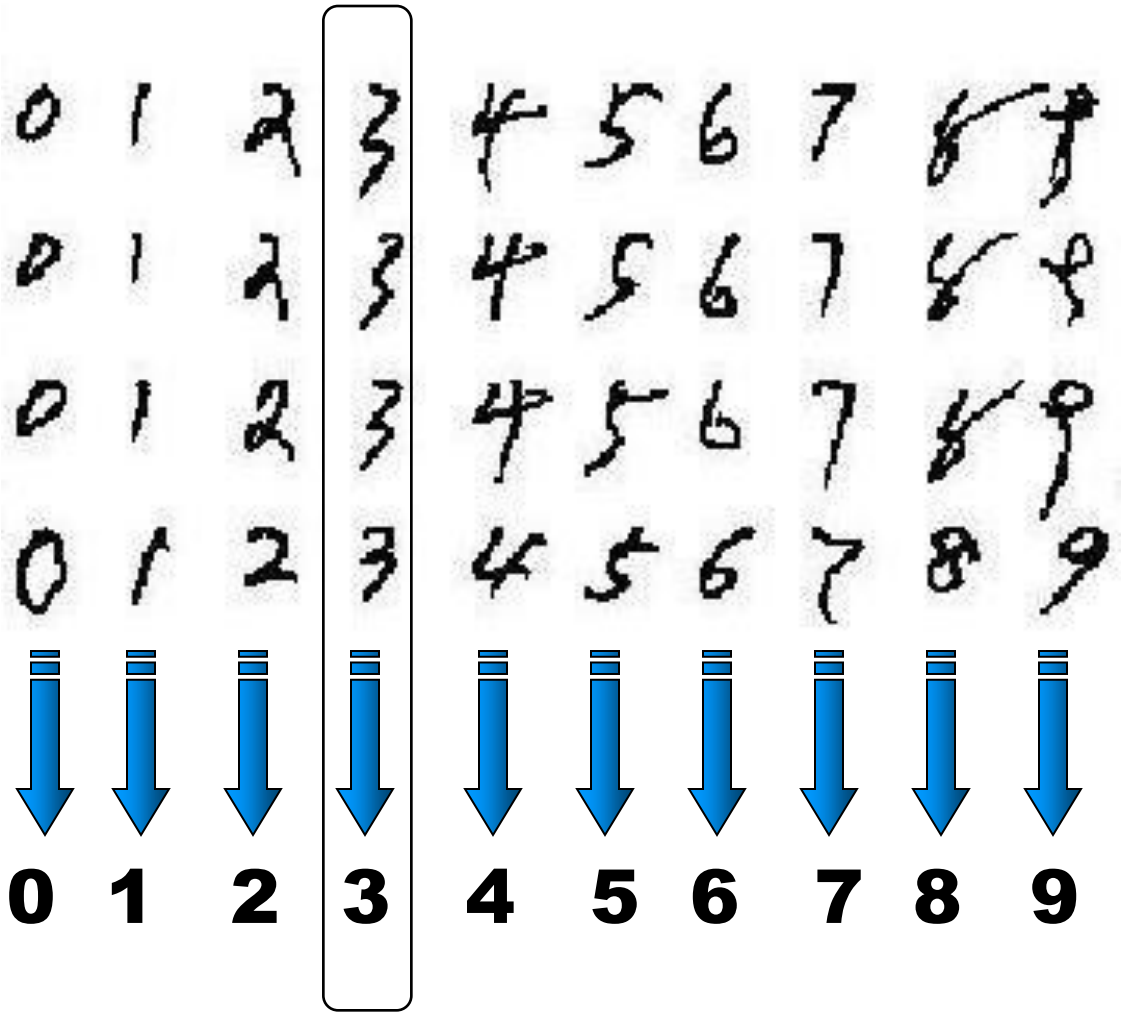
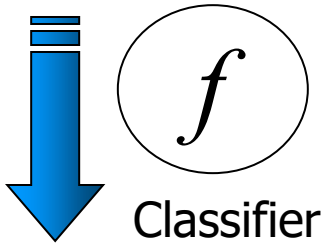


Image
8 x 8



Output class

Applicative Areas

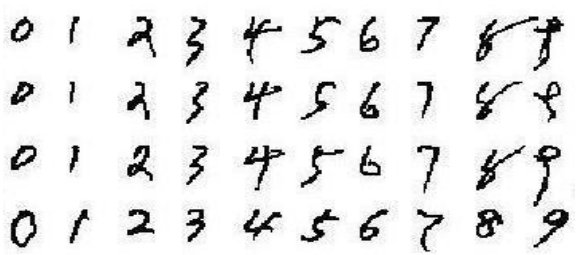
Dip. Informatica
University of Pisa

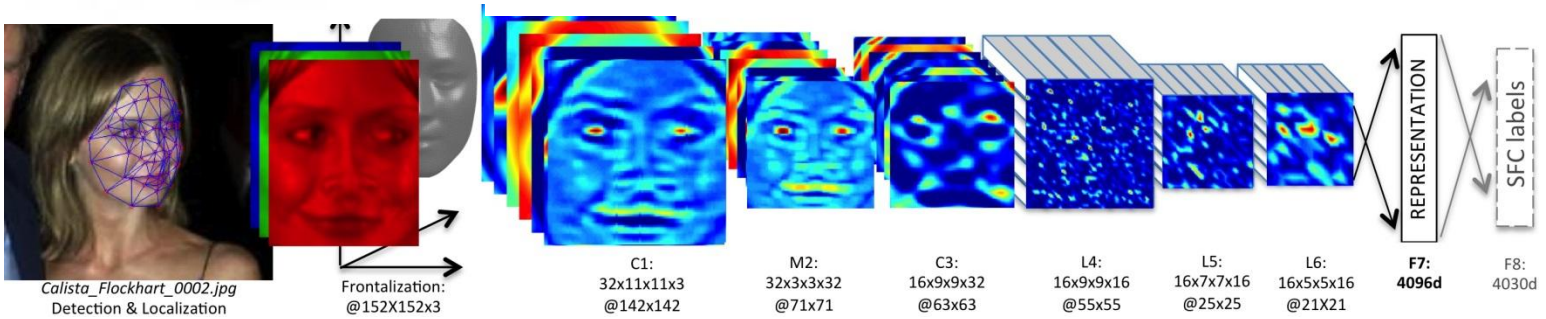
Applicative and related areas:

1. Real-World systems (pervasive, from OCR, to human computer interface, to search engine, …)
2. New interdisciplinary area, encompassing (since the 80s):
   - **Pattern Recognition** (e.g. face and speech recognition), **Computer Vision**, **Natural Language Processing** (from text, e.g. email spam classification, to speech processing, to generative AI),
   - **Robotics**, Adaptive Systems and Filters, **Intelligent Sensor Networks**, Personalized components, …
   - **Knowledge Discovery and Data Mining**, **Information Retrieval**, Analysis of complex data (**Med**, **Bio**, Chem, Web, Marketing), Financial forecasting, …

- The impact has been revolutionary for some fields, e.g. for NLP, Pattern Recognition and Computer Vision, … **A "new AI"**

A look by pictures to some current successful applications (SoA)





Face recognition (Facebook)

Go winner (DeepMind - Big G)

march 2016

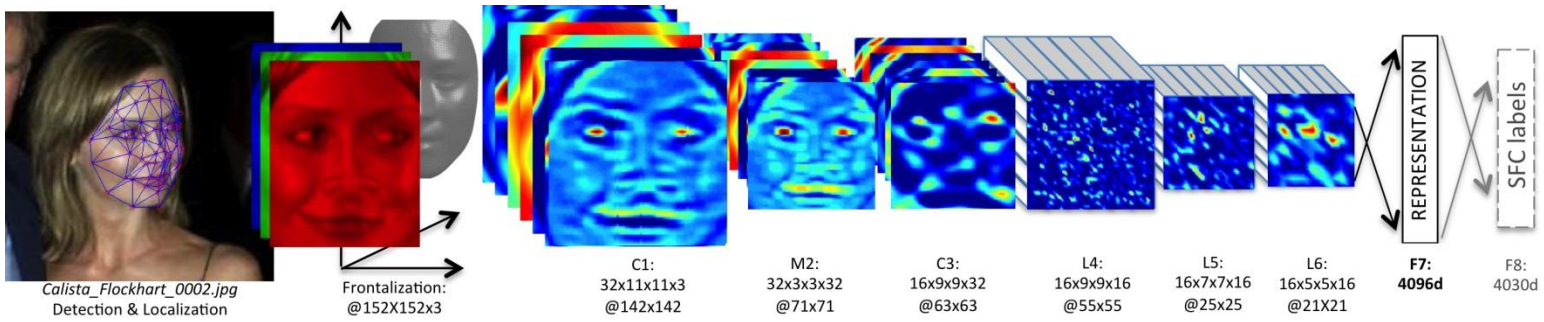
Self-driving cars

Sensor data
Smart *
IoT

An instance on a (not) recent result (CVPR, 2014)

Dip. Informatica
University of Pisa

- **Face recognition** combining (deep) **Neural Networks** and other ML approaches
- Starting from four million facial images belonging to more than 4,000 identities



- Asked whether two photos show the same person, DeepFace answers correctly 97.25% of the time … just a shade behind humans (97.53%).
- From a news on newspaper LaRepubblica April14 ;-)

Health (Examples)

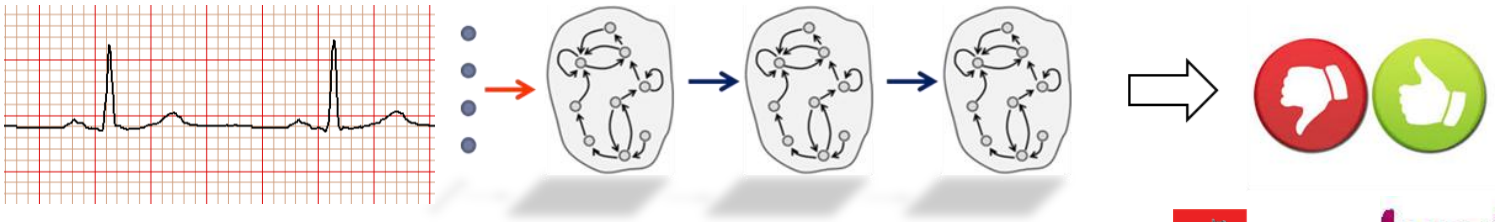
Dip. Informatica
University of Pisa

Diagnosis, therapies, personalized medicine, health monitoring… drug design

Skin Cancer Classification with Deep Learning

- The system (a deep neural network) can learn from 130,000 cases, far more than a doctor can "in many lifetimes"
- It achieves the accuracy of certified dermatologists (Nature 2017)
- It can be implemented with **app**

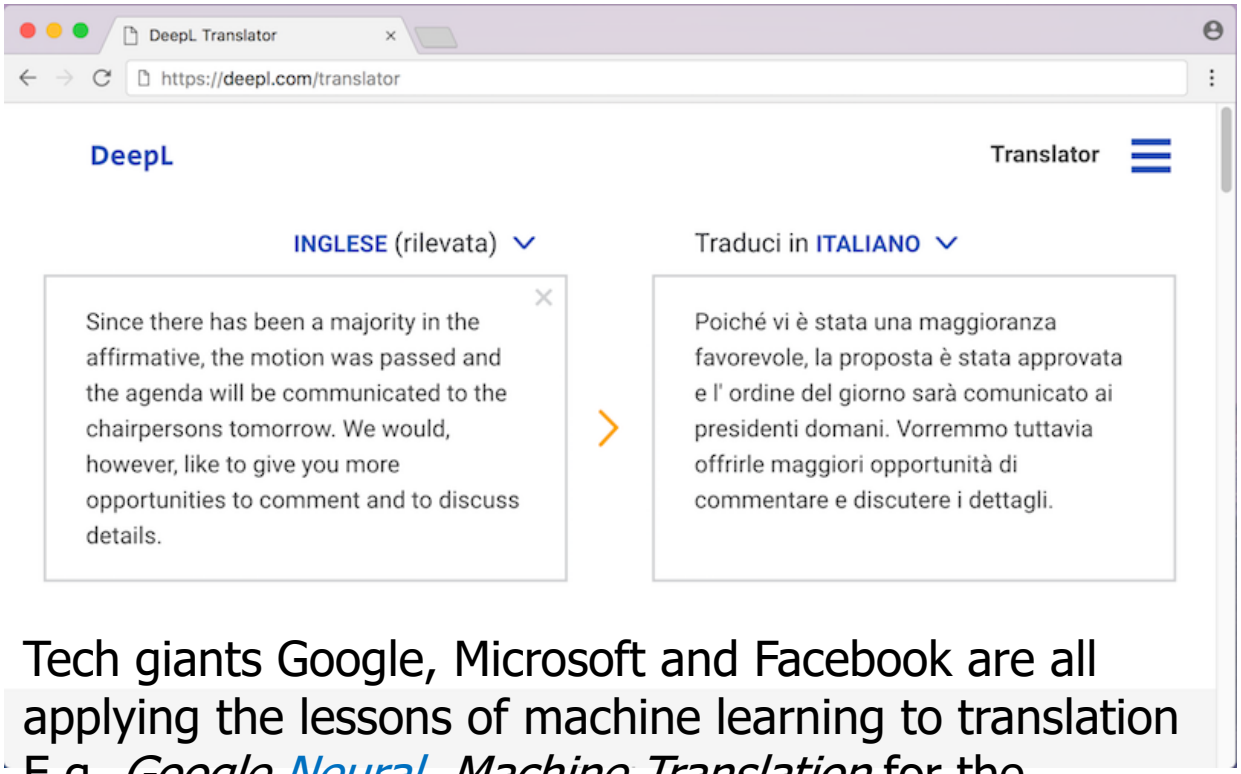
BrAID "Brugada syndrome and Artificial Intelligence applications to Diagnosis"
(PRJ @CIML- UniPi – Tuscany region) [2020- 2024]

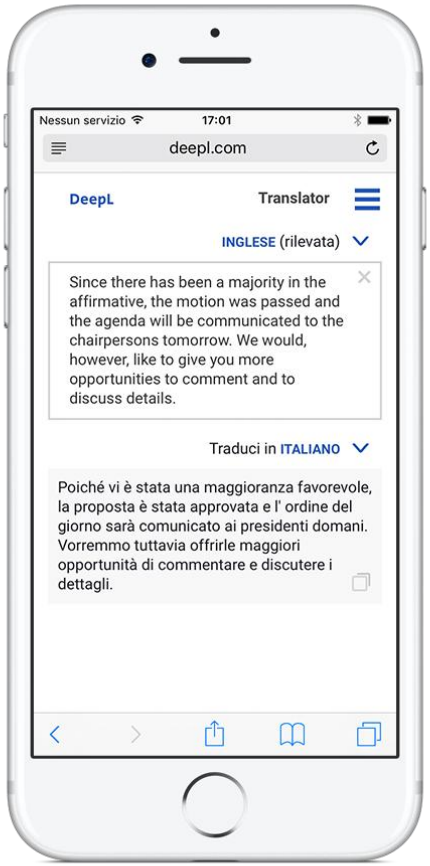


Regione Toscana braid

Machine Translation

Dip. Informatica
University of Pisa



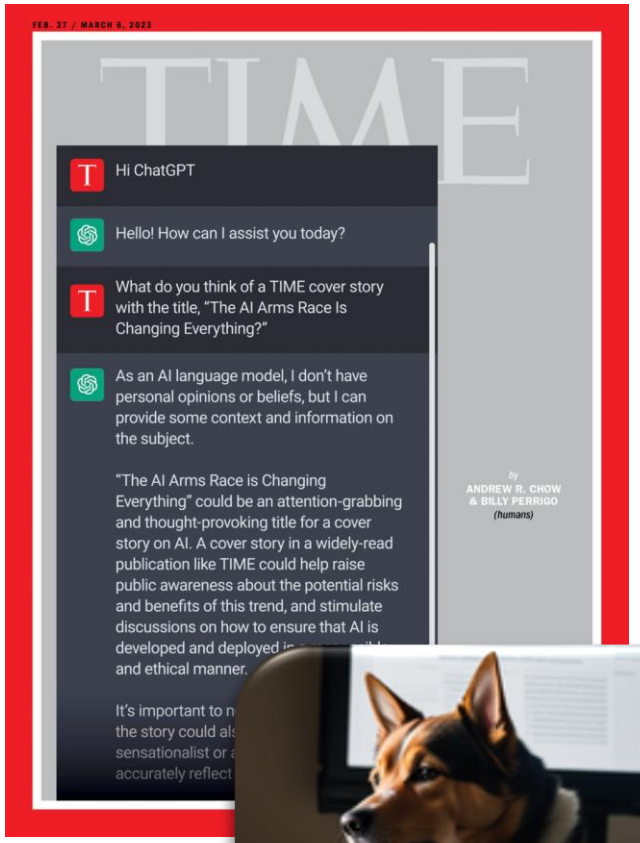


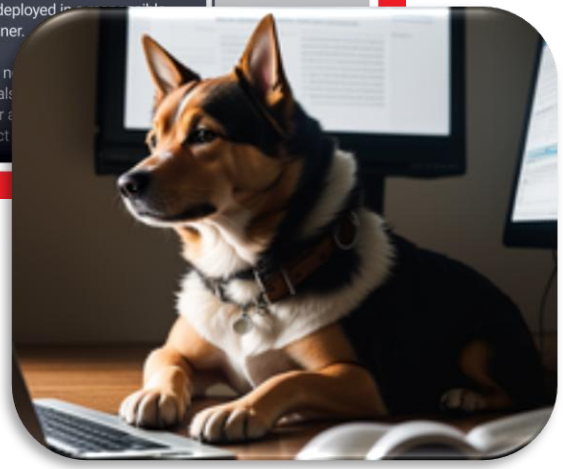
Tech giants Google, Microsoft and Facebook are all applying the lessons of machine learning to translation
E.g. *Google Neural Machine Translation* for the <Google Transate> tool since 2016,
but also small companies are making big progresses:
An Example by <DeepL>: since August 2017
a fully *Neural Network* based system.

Large Languagee Models and Generative AI

Dip. Informatica
University of Pisa

- Core technology to develop the LLM
  - E.g **ChatGPT** Chat *Generative Pre-trained Transformer*, is a large language model-based chatbot developed by OpenAI and launched on November 30, *2022*, notable for enabling users to refine and steer a conversation towards a desired length, format, style, level of detail, and language used.
  - Similarly for other LLM models
  - Dolly x.0 (open source)
- Or systems to generate images, videos, music, etc.





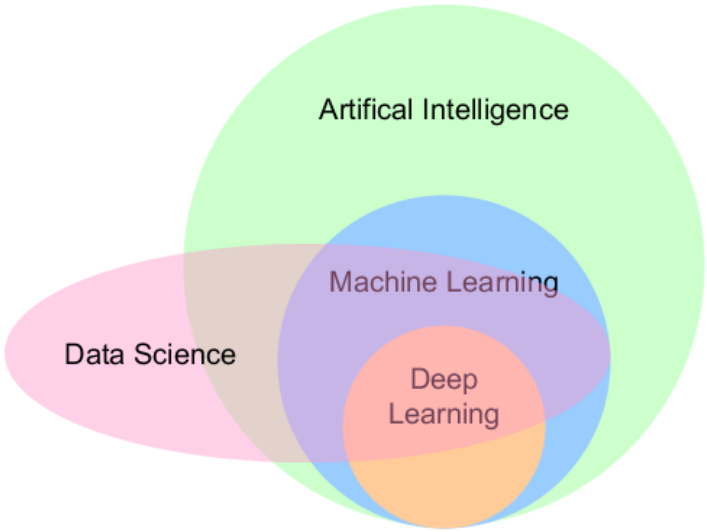
Stable Diffusion, prompt *Cinematic photo of a dog on the Internet editing Wikipedia*

And Related Areas

Dip. Informatica
University of Pisa

ML provides the basis (theories and methods) for new AI approaches (AI revolution)

- It opens huge application areas and opportunities
  - Large effort and investments in Industry
    - Latest **GPT** (General-purpose technologies: potential to drastically affect pre-existing economic and social structures)
  - Opportunity to be part of a big challenge for AI/CS/Society
  - "*ML scientist*" (top ranked in job growing demand, e.g. 2019)



Not all the story but nice view at a glance

Ultimate Aim

Dip. Informatica
University of Pisa

- Anything that is powerful may be dangerous (by misuses) but the ultimate aim of AI/ML is
  - To bring benefits to people by solving big and small problems
  - To accelerate human progress
  - To empower humans and to add intelligence for any other science field

- AI/ML to *augment our abilities enhancing our humanness in unprecedendt ways* (P. Domingo, Scientific American sept. 2018)

Turing award 2018

Dip. Informatica
University of Pisa

- On 27-th March 2019, the Association for Computing Machinery, the world's largest society of computing professionals, announced that Drs. Hinton, LeCun and Bengio had won this year's Turing Award for their work on **neural networks**. The Turing Award, which was introduced in 1966, is often called the *Nobel Prize of computing*, and it includes a $1 million prize …

- "For conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing"

Y. Bengio

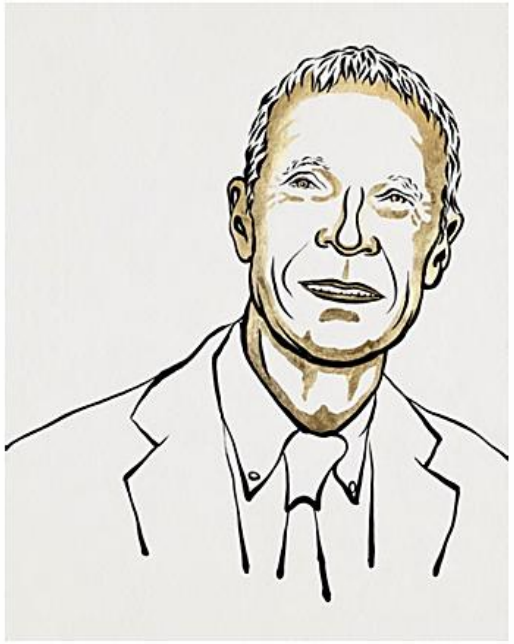
G. Hinton

Y. LeCun

Nobel Prize – Phyisic 2024

Dip. Informatica
University of Pisa

"for foundational discoveries and inventions that enable machine learning with artificial neural networks"



John Hopfield
Born: 1933, USA

Geoffrey E. Hinton
Born: 1947, UK

- https://www.nobelprize.org/uploads/2024/11/Slideshow_Physics_2024_NobelPrizeLessons.pdf

Nobel Prize – Chemistry 2024

Dip. Informatica
University of Pisa

David Baker is awarded *"for computational protein design"* and Demis Hassabis and John Jumper *"for protein structure prediction"*.

David Baker
Born: 1962, USA

Demis Hassabis
Born: 1976, UK

John Jumper
Born: 1985, USA

- "I've always thought if we could build AI in the right way, it could be the ultimate tool to help scientists, help us explore the universe around us. *Demis Hassabis* (**Alpha Fold** - Deep Mind)
- **https://www.nobelprize.org/uploads/2024/10/Slideshow_Chemistry_2024_NobelPrizeLessons.pdf**

Why study ML?

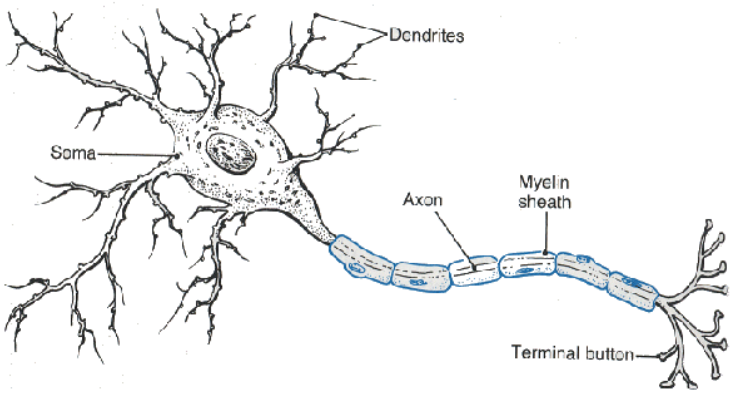
Machine learning in Master Degree

Dip. Informatica
University of Pisa

Why study *Machine Learning* ?

- To know the basic principles of learning processes (computational aspects)
- To know new computing paradigms

  e.g. natural inspired models: **Neural Networks**
  - Studied as computing paradigms since the 40 ' (more than 80 years ago)
  - Neurobiological inspiration
  - Nowadays: set of powerful computing models for *function approximation* and with predictive capabilities supported by a rigorous theoretical ground (*learning theory*)
  - Models for Deep Learning
- To be able to apply these models with a *rigorous approach*
- Basic for all in subsequent AI courses



Modern ML

Dip. Informatica
University of Pisa

- From a collection of heuristic methods (often collections of tricks originating from the contributing fields of biology, statistics, neural networks, fuzzy logics, …) to the construction of *general conceptual frameworks*
- *Fundamental concepts and principles* that govern the learning processes: a new understanding has emerged:
  - Expression of different models in different (math/C.S.) "languages" does not change the principles
  - Different applications fields do not change the principles
  - New technological advancements do not change the principles
- The core of the *principles* and the *methods* to build flexible/powerful computational models from data is "*the Machine Learning*"

ML aims: points of view

Dip. Informatica
University of Pisa

- As AI methodology → Build Adaptive/Intelligent Systems
- As statistical learning (inference of hypothesis within math. principles)
  → Build powerful predictive system for Intelligent Data Analysis
- As computer science method for innovative application areas
  → Using models as a tool for complex (interdisciplinary) problems

- All these points of view will be considered in our course,
- following a critical approach to discus the different ML methods with evaluation of their *limitations* and appropriateness of use!
- Attention to the principles underlying the methodologies.

Practical aims

Dip. Informatica
University of Pisa

Two concrete use of the course:

- Learn how to *develop new models* in the ML field
- Learn to *apply* current state-of the-art methodologies for problems in other (interdisciplinary) fields

Great effort: why?

Dip. Informatica
University of Pisa

- There are many topics and it is a continuously evolving field: we have to cover them
- ML course to provide solid ground for Data Science and modern AI (and all the following courses)
- For your professional activity, it is expected from **you** to be the ML expert !!!
  - Keep it in mind (to motivate your effort)! But also:
  - Interesting: *ML: the driving force of AI* (N. Cesa-Bianchi 2019).
  - A lot of fun: working with learning machine is intriguing
  - A cooperative context for this class will help to work hard and better

Intrerlude: ML in the curricula

Dip. Informatica
University of Pisa

Since 2017 organization by curricula of the CS master programme

PLEASE *read by yourself* the slides: **INTRO-curricula25-v0.1.pdf**

And ask if something is not clear. Thanks.

- Welcome to the new students in CS
- The curricula
- ML in the curricula
- Other master degree students
- Notes for UniPi IIA students
- AA1 (past course) and ML course

PLEASE FILL THE FOLLOWING FORM,
help me to know you and to help you:

https://forms.gle/B5CvwY84vHcCZ5kk7

Using the UNIPI account

How to study this course

Prerequisites

Dip. Informatica
University of Pisa

- No other course is strictly needed (we start from scratch)

- General prerequisites (typically from a First Cycle «Laurea» Degree Programme in CS/Math/Physics/Eng./…):
  - mathematical analysis (functions, differential calculus)
  - elements of matrix notation and calculus
  - elements of probability and statistics
  - algorithms

  See few slides at the end of this lecture for math background references!

Objectives

Dip. Informatica
University of Pisa

- We introduce the principles and the critical analysis of the main paradigms (models and algorithms, with a focus on *neural networks*) for learning from data.
- Methods:
  - The concepts are progressively introduced *starting from simpler approaches up to the state-of-the-art models* in the general conceptual *framework* of modern machine learning (learning of functions from examples).

The course focuses on the

- critical analysis of the characteristics for the design and use of ML
- rigorous experimental evaluation (use of ML tools *versus* correct/good use of ML)
- computational aspects of learning systems (CS!)

Course structure - preview

Dip. Informatica
University of Pisa

- There is a strong structure in the course:
  - *starting from simpler approaches up to the state-of-the-art models*
  - *the fundamental concepts will be introduced by the means of different models*
  - If you follow the *file rouge* (if you recognize that the main concepts do not change) you gain an easer approach to study the ML course!!!

  Practical: **1)** Study the singular methods

  2) Consider the connections/relationships/comparison among them (also to abstract the concepts)

- We will use different "language" building in parallel different approaches to learn (to infer hypothesis) from data

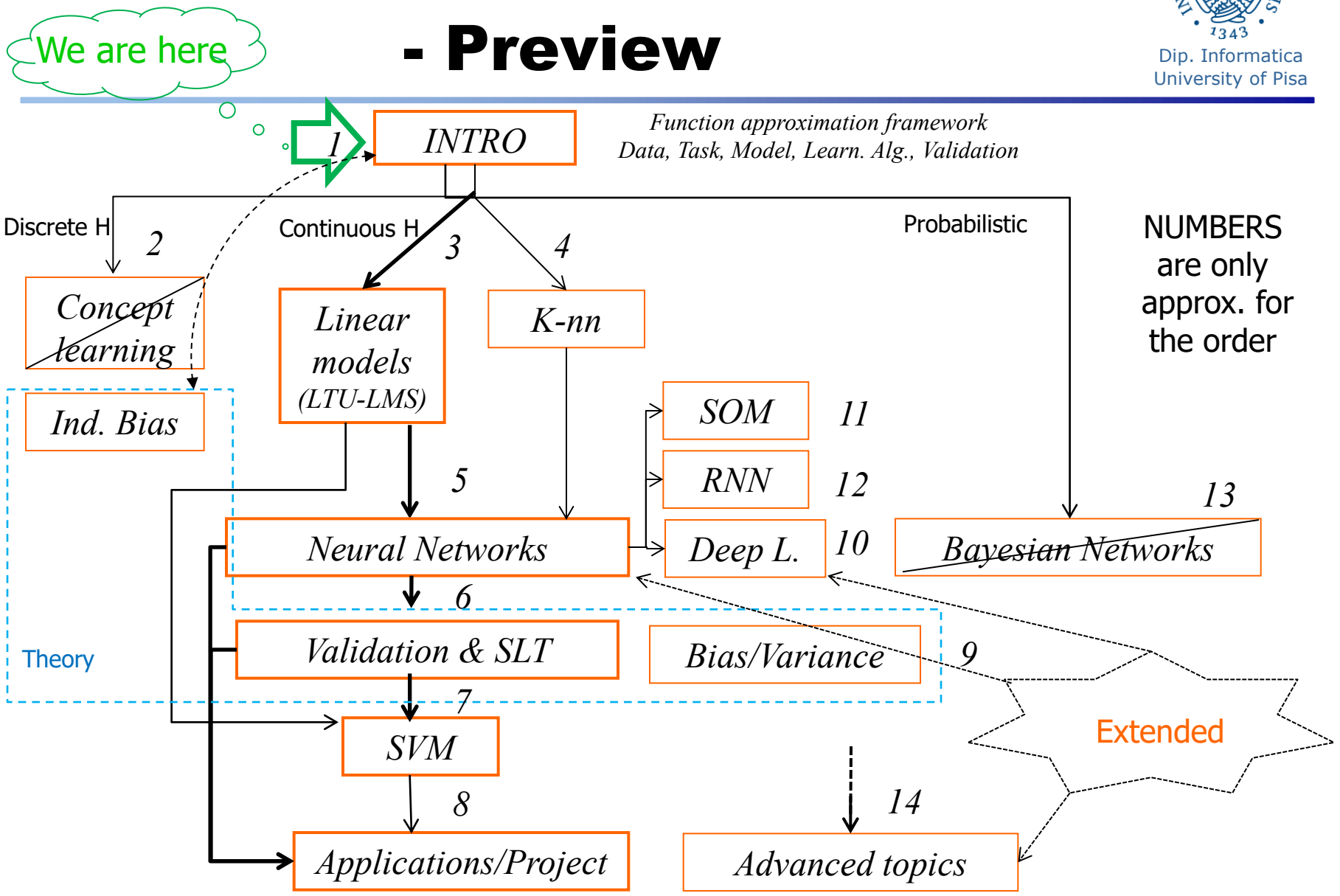
  - From simpler to more flexible models, along with the support of theoretical results
  - Toward the understanding of the relationship between *flexibility* and *complexity control* as a fundamental concept for the model generalization capability

- A graphical overview will help us to collocate the different lectures in the course structure
  - where we are, done, where we are going to, and why

(speech)

ML Course structure - Preview

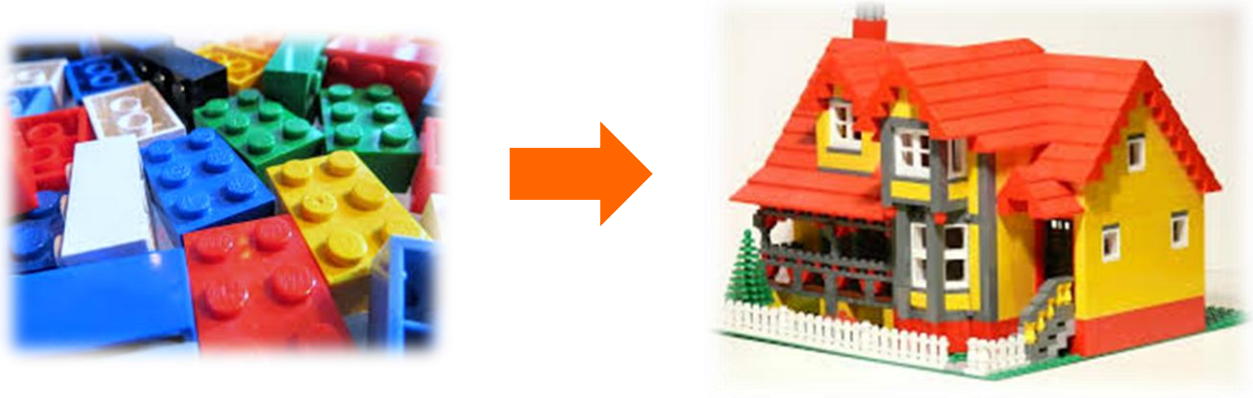
Dip. Informatica
University of Pisa



Again emphasis on the course structure

Dip. Informatica
University of Pisa

- This is NOT just a diagram on the order of the lectures
- It describes our path and the approach: we introduce some «*bricks*» and then will use the bricks to build ML models (like NN from linear models or a ML systems from validation techniques, etc.).
- In this way things will be simple at the end (by reusing what we did before)
- So keep in mind this: the intermediate models/concepts are often a piece of a puzzle/building that we will compose later: you have to look back at the end
  - Always follow the structure and please ask to clarify if needed



many bricks vs just a unique building (our home)

Symbol Legend

Dip. Informatica
University of Pisa

- Fast in class, often secondary, it will be treated later or useful for your successive re-reading (e.g. to connect arguments)

- Tech argument, fast in class, *important topic* but not strictly need in order to follow the course main line at a first glance, but useful for your successive deeper reading (e.g. because it is easier to be understood *later*, or useful to develop the *project* etc.)

Exercise

- "*Exercise*": something *useful* to do by yourself

Course contents (skip now)
(Indicative synthetic programme)

Dip. Informatica
University of Pisa

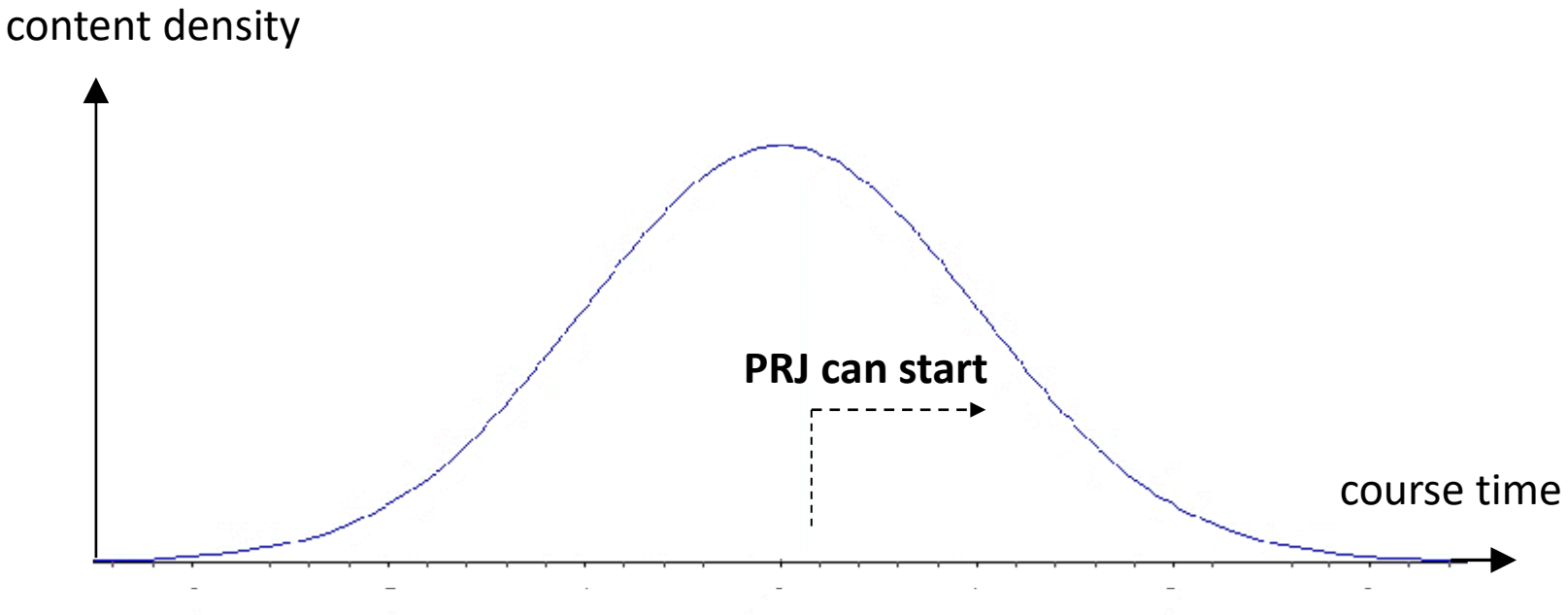
WARNING: See the final program provided at the end of the course

- **Introduction:** Computationl learning tasks, prediction, discrete and continuous spaces, inductive bias, generalization
- **Basic concepts and models**: linear models (models and properties), nearest neighbor (models and properties), propositional and rule based models,.
- **Neural Networks (NN):** Perceptron and computational properties. Introduction to multilayer feedforward NN: architectures and learning algorithms*. Regularization. Recurrent NN*. *Intro to recent NN paradigms**: Deep Learning, Randomized NN.
- **Principles of learning processes and general practical aspects:** Validation: model selection and model assessment, Bias-Variance analysis. Elements of Statistical Learning Theory, VC-dimension.
- **Support Vector Machines:** linear case, kernel-based models.
- **Bayesian and Probabilistic/Graphical models** [removed]
- **Unsupervised learning:** vector quantization, self-organizing map.
- **Introduction to Benchmarks and Applications** *
- **Advanced approaches**: Learning for Structured data (intro)*

*** These parts have been progressively extended**

Course Intensity/Speed Distribution through Time

Dip. Informatica
University of Pisa



Introduction: slow and soft

Core: Main models (NN, SVM) and Validation: dense (faster)

Advances: fast but consequential (not dense) and stimulating

Assessment methods

Dip. Informatica
University of Pisa

Exam:

- **Project** (Written exam)
  - Students have the opportunity to develop a project realizing/applying a learning system simulator (typically a simple neural network) and to validate it through benchmarks. A written report (slides) will show the results.

    NEW
    - Great *opportunity* to apply the concepts by yourself
    - Great *opportunity* to show your concrete understanding and effort for the exam
  - Deadline: ~(around/ or more than) 10-14 working days before the oral exam session (see the Moodle folder of your session for the exact deadline)
  - **See details in the lecture for project presentation**
  - Last years: *competition* with *blind-test*
    - which is part of the benchmark results in the prj
    - also some joint proposals with CM course

- **Oral exam** (dates according to the exams sessions)

Exam (II): Some details on PRJ

Dip. Informatica
University of Pisa

- **Project**: we will discuss in *a specific lecture* all the details but from now it useful to know that: NEW
  - It is made by a ***team*** of 2 or 3 students (exceptions are *very rare*, subject to restrictions, MUST be justified by serious impairments and authorized in advance)
    - A shared document *will be open* to find partners (see Moodle)
  - It is made just one time, i.e. with 1 delivery
  - It is corrected/discussed jointly with the oral exam (at the exam session)
  - It is an implementation of a NN or an application of ML SW (type A or B)
    - It includes code made by you (type A) + results (also for the competition) + report (type A and B) in the form of slides within a basic template NEW
    - *Type A*: if you are self-motivated to develop a model by and with good programming skills, programming language is free (we will discuss later)
    - *Type B,* using ML libraries or *code made by LLM* (NEW): more effort in the experimental and comparison part
    - There is no difference in the grade, it depends on the quality
    - Neither A nor B: ask to me (very very rare case, e.g. PhD students)

Exam (III): Details on Oral and Tests

Dip. Informatica
University of Pisa

- **Oral**: Prj discussion + questions on *all the course content*
  - Including written questions [in class **general (pre)-test**] for all the candidates at the date of the official exam session
  - **Intermediate tests are scheduled during the semester**: a set of (easy) intermediate quizzes will be provided during the course through Moodle, at the end of each cycles of lectures: 5/10 minutes ***in class*** tests (**only** for **in presence** students, i.e. physically present in the classroom), using your devices (laptop, tablet, or even a smartphone), BYOD policy: check detailed HOW TO INFO in the Moodle FAQ section
    - If successful passed (min. 80% of tests, and score mean > 6/10 on all the completed tests) you can skip the winter session written general pre-test (Jan. or Feb.)
    - It means that the intermediate tests are successful passed if (and only if):
      - you participated to a minimum of 80% of the in-class tests (no threshold is needed on the score of each test) AND
      - you achieved a mean score over 6, over all the completed tests
    - **Note**: intermediate tests are strongly suggested but not mandatory, see later

Exam (IV): Details on Oral Finalization

Dip. Informatica
University of Pisa

- And then (at each session) we distribute the dates of the **prj discussion and oral finalization** in the following days/weeks (for each group)
  - The calendar of oral slots depends on the season. The first sessions (first session and especially second session may require a greater interval to complete all the orals, we will discuss these aspects later)
  - ALL the students of the prj group must be present for the exam, i.e. making the exam in the same session
  - Style: I'll ask answer first by math language (e.g. in the form of *equations* due to clarity, synthesis, …; then we can discuss on them)

- Take care to register your name in the official UNIPI (esami.unipi.it) portal for exams (check the **deadline** in large advance): It is a must, only registered student can be admitted to the exam (all students rule)

Exam: schematic synthesis

Dip. Informatica
University of Pisa

In the following order:

1. **Course lectures** (stay on-line with them!)
   - And perform the **Intermediate tests** during the course
2. **Project work** (you can start around in the middle of the semester)
3. **Project package delivery:** results, slides report etc. at the date specified in the Moodle (around or more than 10/14 working days before the official exam session, check Moodle deadline)
4. **Oral:** start at the date of the official exam session, with questions in written form [in class **general (pre)-test**] (it is in the same session of the prj delivery).
5. **Project discussion and oral finalization** according to a specific calendar for each group, i.e. as soon as possible in one of the following days/weeks (all the partner students must be present).

The main hints (I)

Dip. Informatica
University of Pisa

- Follow the lectures and slides as a guide, studying ***progressively*** during the course
  - Special interactive classes will be used to assess activities and make a ML discussion forum (**Q&A classes**). This is a class, not a set of records*.
  - Within a **cooperative** and collaborative context!
  - The **intermediate tests** are also meant to stimulate you to **stay "*online*"** and to have constant ***self-evaluations***, hence useful for you!

- A major hint *from past students*:

1. **FIRST study the course content**
2. **THEN apply for the project**
   - This help to boost efficacy and efficiency of your work

The main hints (II): Help!

Dip. Informatica
University of Pisa

If something is *not* clear to you:

- Don't worry
- Ask me → see later *Q&A* modality
- Collaborate with your teammates/colleagues
- Use the books/papers : *bibliography* + personal search → see later for bib
- Study, it's ML!!! It's a constantly evolving field, …
  - go back to slide "Great effort: why?"

Good news

Dip. Informatica
University of Pisa

- This year we will have a ML course assistant to help you for the *project development*
- Take the opportunity to interact with him during the course

Other info

Dip. Informatica
University of Pisa

- **Lectures schedule:** (beginning hour is sharp to have a break)
  - Tuesday 11-13, Aula E
  - Thursday 16-18, Aula E
  - Friday 14-16, Aula E

- **Q&A (with the teacher)**: Thursday **after lecture***+ *special Q&A classes*.
  - * NOTE: the time after the lecture IS NOT mandatory, it is not part of the lecture, it is for singular/*collective* student Q&A reception
  - Or write to me an email for very personal questions, using the TAG **[ML-25-Q]** in the subject (**this is a must**!)

- **ML mailing list**: I'll use Teams and/or Moodle news
  - (take care to have on it the email address that you really use)

- **Make up for lost lectures:**
  - Monday afternoon (they will be announced later, on the need)
  - Lesson recovery *09/12 – 19/12 2025* by academic calendar (consider this period before programming your Christmas vacations travels!)*****

Textbook and material (1)

Dip. Informatica
University of Pisa

- **Course notes (slides copy)**

  with material from time to time indicated in the classroom:
  bibliographic references at the end of each specific topic via books chapters and/or online material.

Hints:

1. *Course notes are a very useful guide to the selected topics!*
2. *Have a look in advance (I typically provide them in advance in the Moodle)*
3. *Take your notes [past student suggestion]: make connections among slide phrases, and slide themselves can be not sufficient to reconstruct the lecture line*
4. *I would like to share in our Moodle your note of the course: please, contact me if you like to share your note (in electronic form)*

5. *Learn to take advantages reading different sources (**master degree level**): see next page for the books and the refences in the slides*

How to get the slides?

Dip. Informatica
University of Pisa

The electronic version of the slides is intended only for authorized student, **for their personal use.**

- The link to the electronic slides is reserved (draft material). It is **not** a public site!

- **Material: Moodle: https://elearning.di.unipi.it/**
  - See **Machine Learning 2025**
  - See also the **FaQ and general information** section therein

- Please do not use the link in any web site/social media
- Please do not repost slides in any form
- Take care of last version file ML-25…–v* (sometime after the lecture!!!)

How to get the slides? (II)

Dip. Informatica
University of Pisa

- If you not have a UNIPI student account yet:
- The didactics offices typically provide the possibility to connect to the Unipi telematic services (Moodle, Teams, etc …), your credentials, even before formally enrolling, please inquire by yourself
- Try connecting to the ALICE service
  - If really nothing works, please ask to me by email for a TEMPORARY guest account *for Moodle*

Textbook and material (2): Books

Dip. Informatica
University of Pisa

- **General Bibliography** (library or on-line):
  - S. Haykin: ***Neural Networks: a comprehensive foundation***, IEEE Press; (2nd. Edition, 1998) → OR better

    S. Haykin: ***Neural Networks and Learning Machines,*** Prentice Hall; *(3rd Edition, 2008)*
  - T. M. Mitchell: ***Machine learning***, McGraw-Hill, 1997
  - I. Goodfellow, Y. Bengio, A. Courville: ***Deep Learning,*** MIT Press, 2016

    On-line: http://www.deeplearningbook.org/

Textbook and material (3)

Dip. Informatica
University of Pisa

- **Other references (examples):**
  - (AIMA) Russell, Norvig: *Artificial Intelligence: A Modern Approach* Prentice Hall (3 edition, 2009)
  - Hastie, Tibshirani, Friedman: ***The Elements of Statistical Learning,*** Springer Verlag, 2001 (new eds. up to 2017) → 2017 *Free on-line copy*
  - Cherkassky, Mulier: *Learning from data: concepts, theory, and methods*, Wiley, 1998 (new: 2nd Ed., 2007)
  - C.M. Bishop: *Pattern Recognition and Machine Learning*, Springer 2006 – free available from the web site
  - Bishop: *Neural Networks for Pattern Recognition*, 1995
  - Duda, Hart, Stork: *Pattern Classification*, 2nd. ed. J. Wiley & Sons, 2001
  - Shalev-Shwartz, Ben-David: ***Understanding Machine Learning: From Theory to Algorithms***, Cambridge University Press, 2014, *Free online copy*

Many technical books/blogs exists, take care of the scientific quality!

Textbook and material (4)

Dip. Informatica
University of Pisa

For software libraries:

- Books that describe *Keras* | *Scikit-Learn*
  - "Deep Learning with Python" (F. Chollet)
  - "Hands-on Machine Learning with Scikit-Learn, Keras & Tensorflow" (A. Geron).
- Using PyTorch, NumPy/MXNet, JAX, and TensorFlow

```
@d2l.add_to_class(LinearRegressionScratch)
def loss(self, y_hat, y):
    l = (y_hat - y) ** 2 / 2
    return l.mean()
```

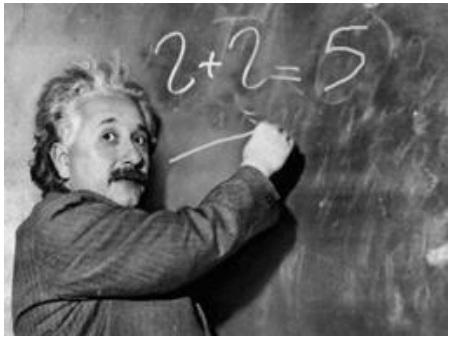
  - "Dive into Deep Learning" (A. Zhang, Z. Lipton, M. Li, and A. Smola) (2023). Cambridge University Press. Freely available from the website.
- We will have an introduction to the main libraries later (by the course assistant)

Many technical books/blogs exists, take care of the scientific quality!

Background

Background 1

Dip. Informatica
University of Pisa



Very informal math preliminaries:

- E.g., AIMA appendix A.2 and A.3 mathematical background) free available at http://aima.cs.berkeley.edu/newchapa.pdf
- Or Deep Learning book chapters I.2, I.3, I.4 (i.e. part I: 2,3,4) http://www.deeplearningbook.org/
- Or (much more) *Mathematics for Machine Learning* (2020) https://mml-book.github.io/

- And feel free to ask for any doubt!!!!
- Don't worry: AI/ML are multidisciplinary fields from the beginning!!!!!

Background 2

Dip. Informatica
University of Pisa

NEW since 2017:

- In parallel we have the course of CM (*Computational mathematics for learning and data analysis*) or *Optimization for Data Science* (@DSBI)
- Great synergy with ML

- Its nature and aim is to provide a deep mathematical background for learning methods. You can consider such lectures
  - Time table: please, check by yourself

- Anyway, the course is *not* mandatory to follow ML, as many students in the past from other Master Degree or curricula without CM
  (I'm aware of this)

Notation

Dip. Informatica
University of Pisa

Generally, we will use consistent notation among the most important parts. However:

- Sometime *locally defined* notation will be used to *simplify* the concepts introduction and/or only if needed to be consistent with *different books*
  - Anyway, the equivalent symbols for the same concept will be indicated
- A document with the "**notation summary**" of the course will provided (after the models discussion, to resume).
- In any case, don't go crazy for notation issues: please ask to me if you have some issues, I will be happy to help (better immediately than later) and to correct if needed

Basic background (references)

Dip. Informatica
University of Pisa

- Multivariable/Multivariate calculus

  Functions with multiple inputs: $f(x_1,x_2)$ or $f(\mathbf{x})$, where $\mathbf{x}=(x_1,x_2,\ldots)$
  - *Partial derivative, gradient.*

- Probability:
  - *probability density function, mean and variance, Normal random variable*
  - *p(x) or P(X=x), conditional probability: p(x|y), joint probabilities, …*

- Matrix calculations and notations: *x (scalar)*, $\mathbf{x}$ *(vector)*, $\mathbf{X}$ *(matrix)*
  - *inner (dot, scalar) product, inverse, norms, …*

- *Notions that can be studied in parallel(e.g. Numerical Methods and Optimization courses and/or the mentioned CM course):*
  - *The linear least-squares problem: use of the Singular Value Decomposition*
  - *Optimization: Lagrange multipliers*

(it) *Prodotto Scalare*
(en) Inner (dot, scalar) product

Dip. Informatica
University of Pisa

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i \qquad \mathbf{a} \cdot \mathbf{b} = |\mathbf{a}|\,|\mathbf{b}| \cos\theta$$

- Other notations $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{a}^T \boldsymbol{b} = \boldsymbol{a}^t \boldsymbol{b} = <\boldsymbol{a}, \boldsymbol{b}>$ and even $\boldsymbol{ab}$ if the context is clear
- Magnitude/size (it '*Modulo*') or length or *Euclidean norm* of a vector $\boldsymbol{x}$

$$\sqrt{\boldsymbol{x}^T \boldsymbol{x}} = \sqrt{\boldsymbol{x}^T \cdot \boldsymbol{x}} = \sqrt{\sum_i {x_i}^2} = d(\boldsymbol{x}, 0) = \|\boldsymbol{x}\|_2 = \|\boldsymbol{x}\| = |\mathbf{x}|$$

Simplified

- Generalization: function that relate a couple of vectors to a number (a scalar): (bilinear symmetric form, by the notation <,>)

$$\langle v, w \rangle = \langle w, v \rangle$$
$$\langle v + w, u \rangle = \langle v, u \rangle + \langle w, u \rangle$$
$$\langle kv, w \rangle = k\langle v, w \rangle$$

- Cauchy-Schwarz inequality $|\langle x, y \rangle| \leq \|x\| \cdot \|y\| \quad \forall x, y \in V$

Tensors & other defs.

Dip. Informatica
University of Pisa

Tensor

- Simple view: an array of numbers arranged on a regular grid with a variable number of axes is known as a tensor.
- i.e. A multi-dimensional array of numerical values.
- E.g. $X_{i,j,k}$ (3 indices)
- Used trivially in ML just for notation advantage,
  but in general with more properties in math useful dealing with change of basis, basis independence etc. in linear algebra, geometry, physics……
- In ML help focus when using GPU etc. (we will see later)

Other defs.

$$\|\mathbf{x}\|_1 = \sum_i |x_i| \qquad L^1 \text{ norm}$$

$$\|\mathbf{x}\|_\infty = \max_i |x_i| \qquad \text{Max norm}$$

Partial derivative: calculus

Dip. Informatica
University of Pisa

- The **partial derivative** generalizes the notion of the derivative to higher dimensions.
  A partial derivative of a multivariable function is a derivative with respect to one variable (e.g. x_1) with all other variables held constant.

- Hence, if $Df(x) = f'(x) = df/dx$
- for $f(x_1, x_2, x_3)$ we can compute: df/dx_1, df/dx_2, df/dx_3

$$\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \frac{\partial f}{\partial x_3}$$

Gradient

Dip. Informatica
University of Pisa

- When a function of two variables $f(x_1, x_2)$ have partial derivatives at each point (x_1, x_2) we can associate the **vector** of the two partial derivatives $\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right)$ called the **gradient** of f, often denoted ∇f or *grad f.*

- For n variables: the gradient is the vector field whose components are the partial derivatives of f:

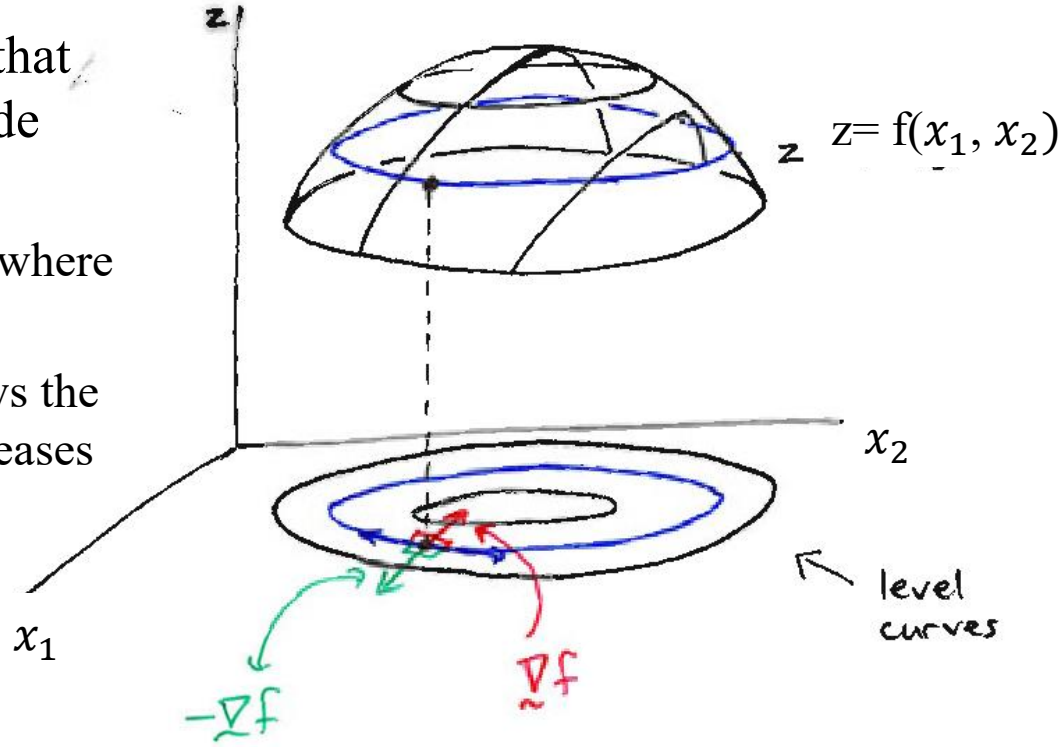
$$grad\, f = \mathbf{e}_1 \frac{\partial f(x_1, \ldots, x_n)}{\partial x_1} + \cdots + \mathbf{e}_n \frac{\partial f(x_1, \ldots, x_n)}{\partial x_n}$$

where the $\mathbf{e}_i$ are the orthogonal unit vectors pointing in the coordinate directions *(e.g. 001, 010, 100 in 3D).*

Gradient on a surface

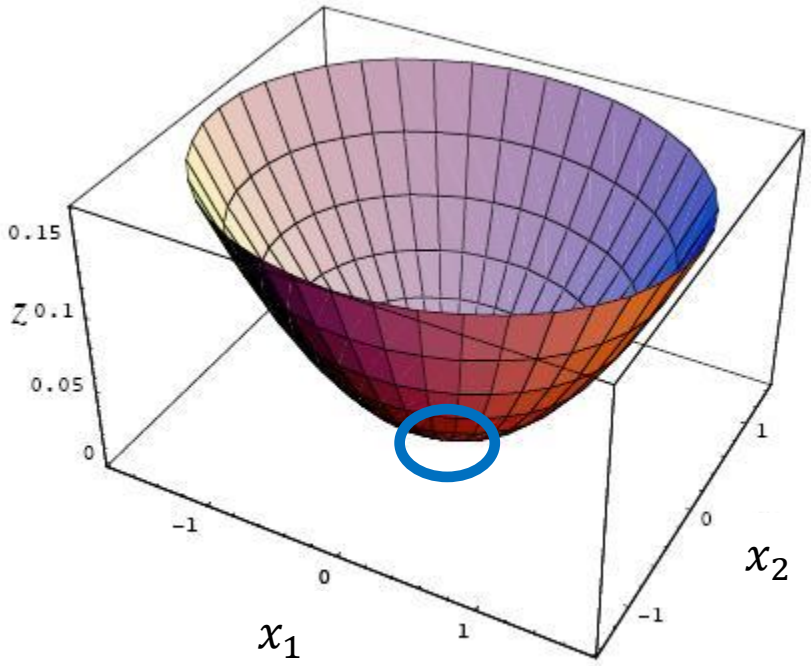
Dip. Informatica
University of Pisa

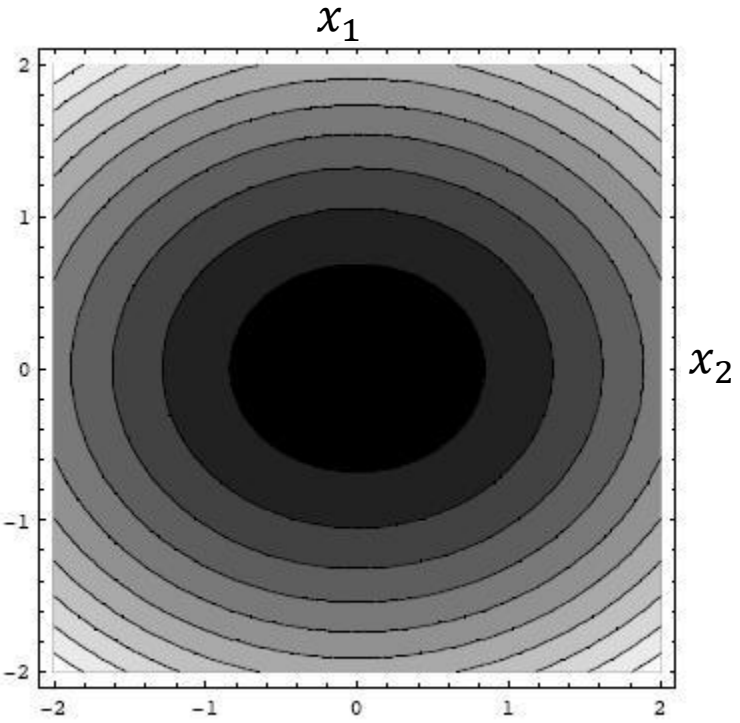
- **The gradient at a point** is a **vector** pointing in the **direction of the steepest slope** at that point**.**
- The steepness of the slope at that point is given by the magnitude of the gradient vector
  - The gradient shows the direction where the function grows
  - The negative of the gradient shows the direction where the function decreases



Local Minimum

Dip. Informatica
University of Pisa





- Example of a *stationary point*: the gradient is zero

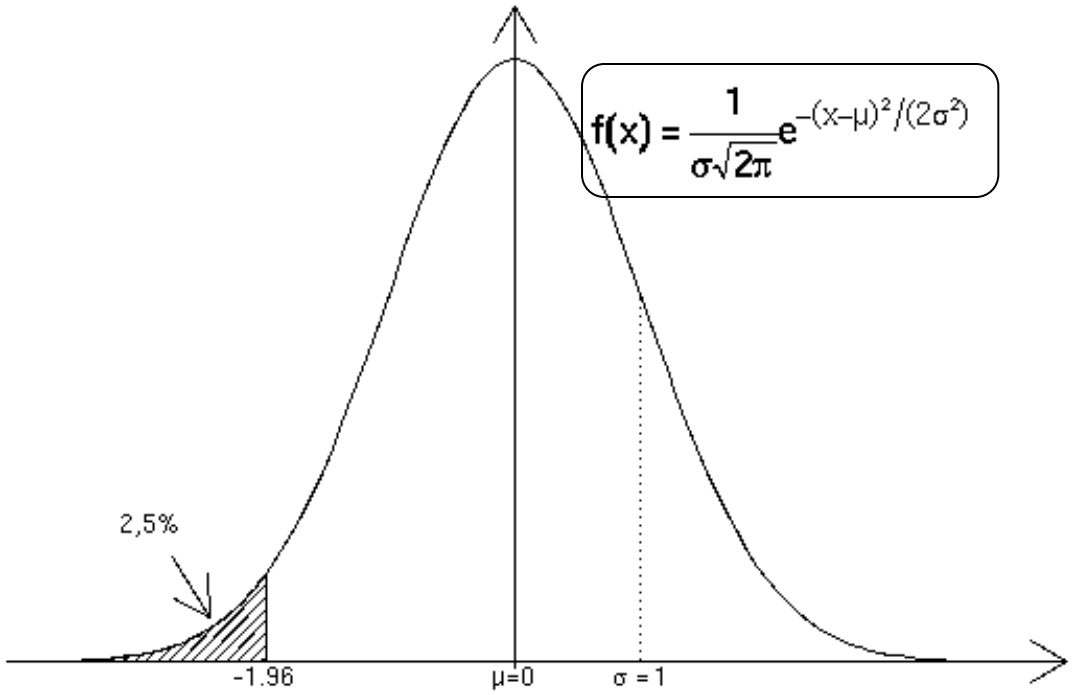
(while are also stationary saddle points or local min/max)

Density (example)

Dip. Informatica
University of Pisa

- The density function of the normal random variable with mean 0 and variance 1 (called standard normal distribution, see the figure) and the analytical expression of the corresponding density in the generic case *(mean μ and variance σ^2).*

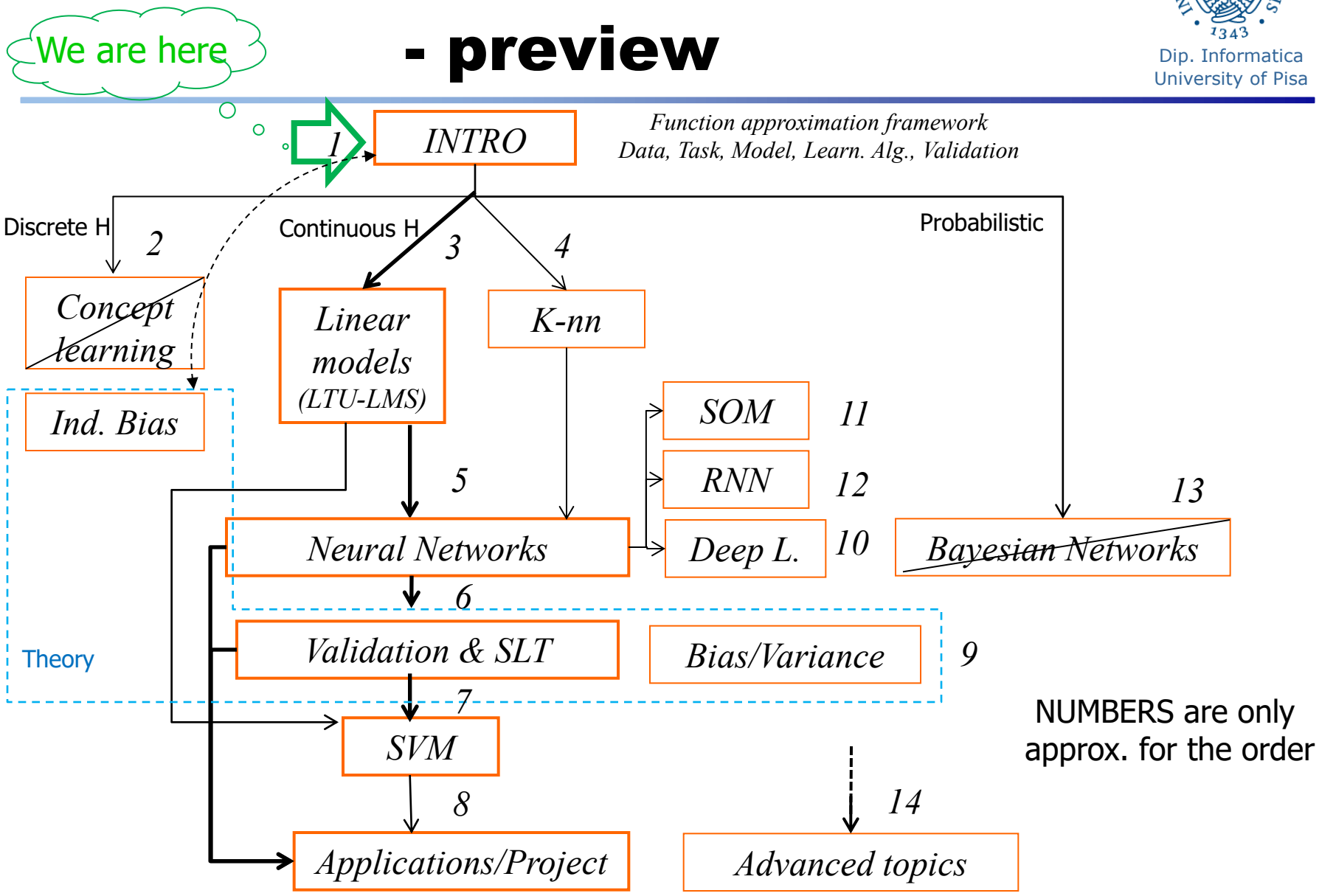
The probability density function (pdf) for a **normal distribution:** the "Gaussian function"



Next steps

ML Course structure - preview

Dip. Informatica
University of Pisa



Summary of the Intro to ML

Dip. Informatica
University of Pisa

- Part I (now)
  - Motivations, contextualization in CS
  - Course info
- Part II (in Lect.s 2 and 3)
  - Utility of ML
  - Learning as function approximation (pilot example)
  - Design components of a ML system, including
    - Learning tasks
    - Hypothesis space (and first overview)
    - Inductive bias (examples in discrete hypothesis spaces)
    - Loss and learning tasks
    - Generalization (first part)
- Part III (in Lect. 4)
  - Generalization and Validation

Aim: *overview and terminology*
before starting to study models and learning algorithms

PROLOGUE END

Alessio Micheli
micheli@di.unipi.it

Computational Intelligence & Machine Learning Group

Dipartimento di Informatica
Università di Pisa - Italy