Introduction to the GDL course

Generative and Deep Learning (GDL)

Davide Bacciu (davide.bacciu@unipi.it)

Università di Pisa

Objectives

Train machine learning (ML) specialists capable of

- understanding the state-of-the-art in generative and deep learning models
- designing novel learning models
- developing applications using modern AI methods

Focus on:

- Methodological aspects and design choices of modern ML models (inductive biases et al)
- Solving learning tasks beyond predictive (distribution estimation, generation)
- Dealing with complex data (beyond i.i.d.)

Expected Outcomes

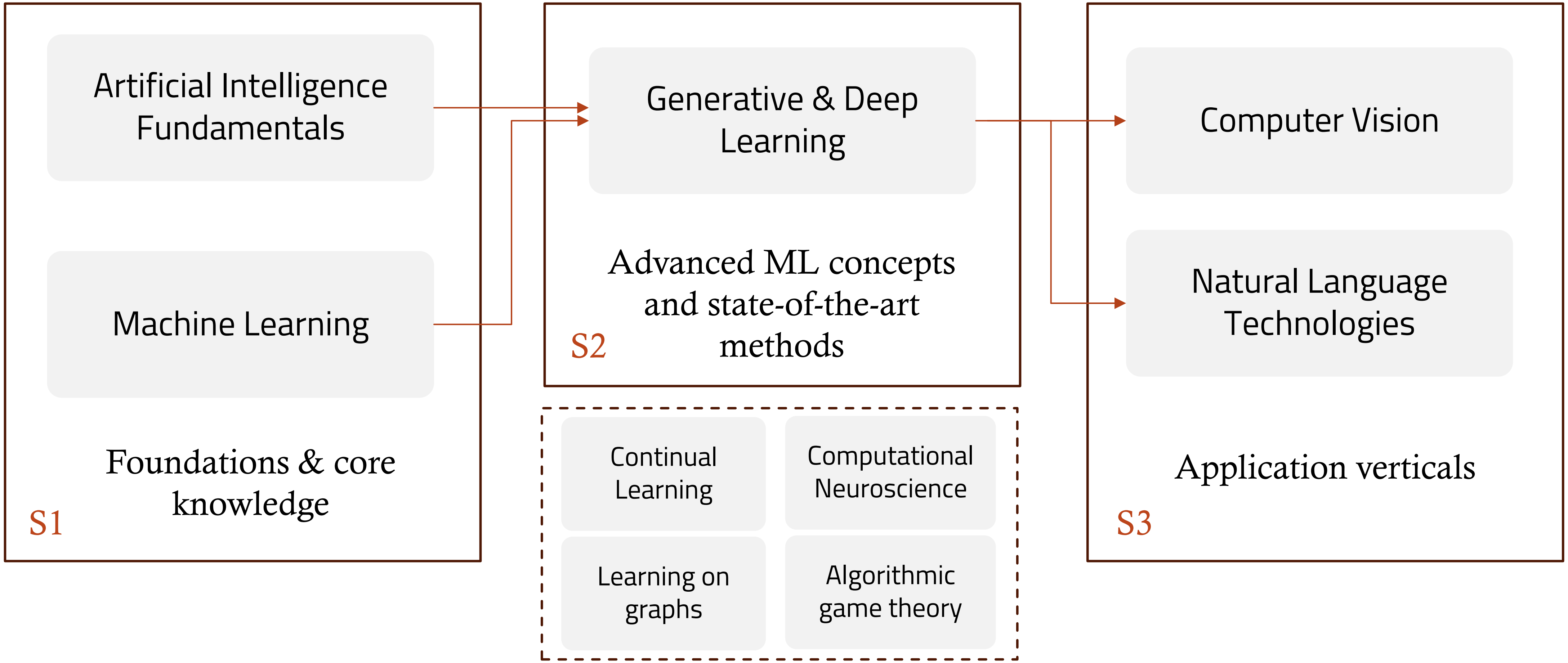
Gain in-depth knowledge of advanced machine learning models

Understand the underlying theory

Be able to individually read, understand and discuss research works in the field

Gain knowledge of modern ML/DL programming libraries and their use to design and train models

GDL in the AI Curriculum

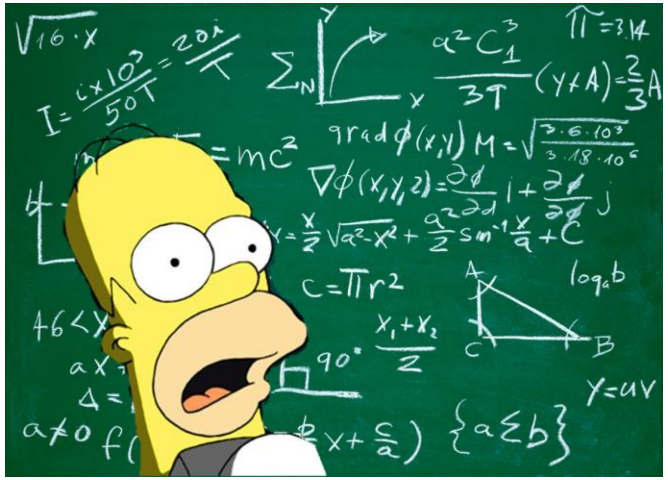


Prerequisites

- Knowledge of machine learning fundamentals
  - Pass the ML course or.. come discuss your ML skills with me
- Mathematical tools for ML
  - Algebra and calculus
  - Optimization
  - Probability and statistics
- Programming experience in Python (helpful)

In short: be comfortable with the knowledge in Chapters 2-5 of the [Deep Learning book]()

…and, above all, a disposition not to get (easily) scared by (some) math!



Organization

The course is organized in five modules

- Fundamentals of probabilistic models and causality
- (Generative) Learning in probabilistic models
- Deep Learning fundamentals
- Generative deep learning
- Advanced models and applications

Incremental approach, from fundamentals to advanced deep learning models with tightly interconnected module content

Topics (I)

- **Fundamentals of probabilistic models and causality**
  - Probability and statistics refresher
  - Graphical models representation
  - Causality
  - Inferring dependency from data
- **Learning in generative probabilistic models**
  - Learning as inference (ML, MAP, Bayesian)
  - Exact learning with fully observable models
  - Exact learning with latent variables
  - Approximated learning: Generalized Expectation Maximization, variational and sampling approaches
  - Probabilistic models: hidden Markov Models, Markov Random Fields, Latent Variable Models

Topics (II)

- **Deep Learning fundamentals**
  - Issues with information propagation in deep networks
  - Convolutional neural networks
  - Recurrent neural networks
  - Neural attention
  - Transformers
  - DL toolset: dropout, normalization layers, mini-batching, residual connections, …
  - DL libraries: Pytorch, Keras-TF, …

Topics (III)

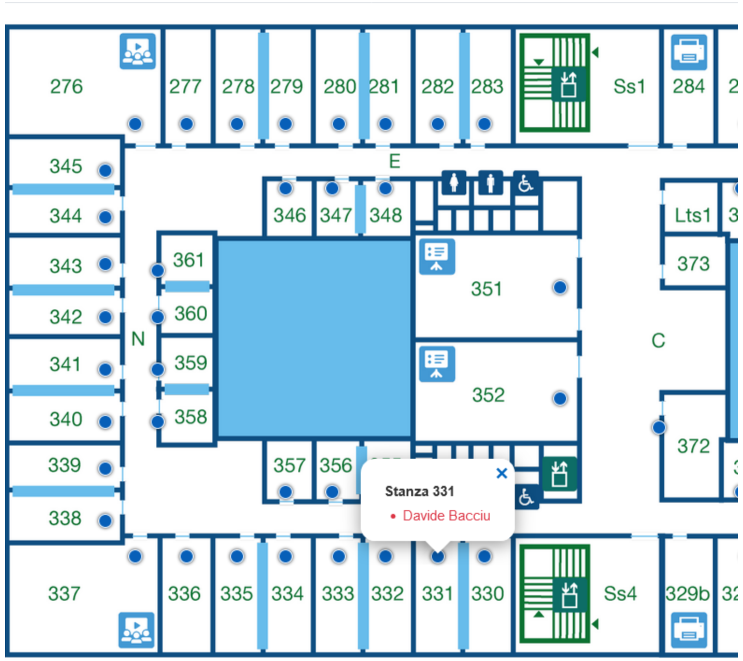
- **Generative Deep Learning**
  - Autoregressive generation
  - Generative adversarial networks
  - Autoencoder architectures
  - Diffusion models
  - Flow models
  - Energy, score and flow matching
  - Causal representation learning
- **Advanced topics and applications**
  - Deep learning for graphs
  - Reinforcement learning
  - ...guest lectures.. (when time allows)

Course Instructor

Davide Bacciu

- Email – davide.bacciu@unipi.it
- Office - Room 331O, Dipartimento di Informatica
- Office hours - Monday 16-18 (email me!)

Reference person for the course (admin, exams and the like)



Course Co-lecturer

Riccardo Massidda – (riccardo.massidda@di.unipi.it)

- Postdoc @ University of Pisa
- Specializes on causal learning
- Course coverage
  - Module of 4 lectures on graphical models and causality fundamentals
  - Coding tutorials
  - Causal representation learning

Course Schedule

Weekly timetable:

Day	Time
Tuesday	11.15-12.45
Wednesday	16.15-18.00
Thursday	11.15-12.45

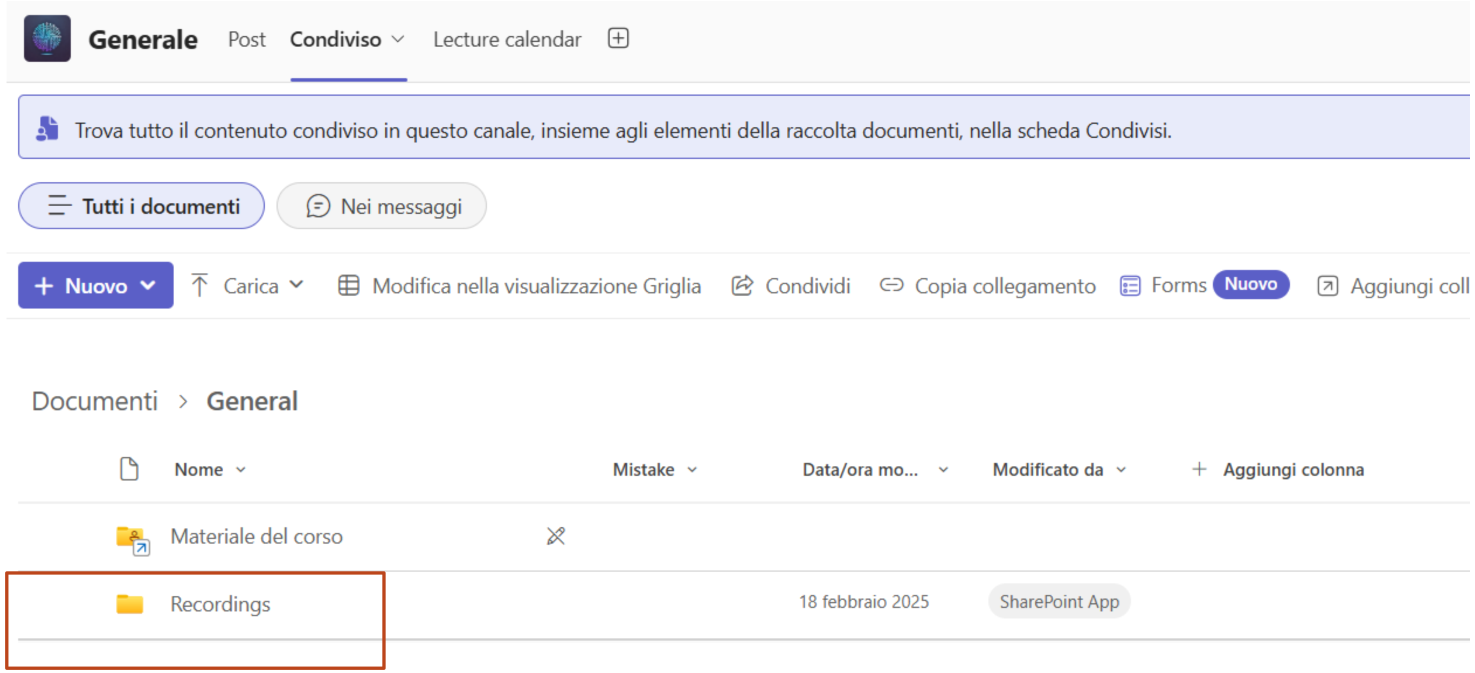
Talk now if there are severely under-looked issues with the schedule!

Course comprises 36-37 lectures

- Course will be given in-person
- Video recording of the lectures will be made available (best effort quality and timeliness)

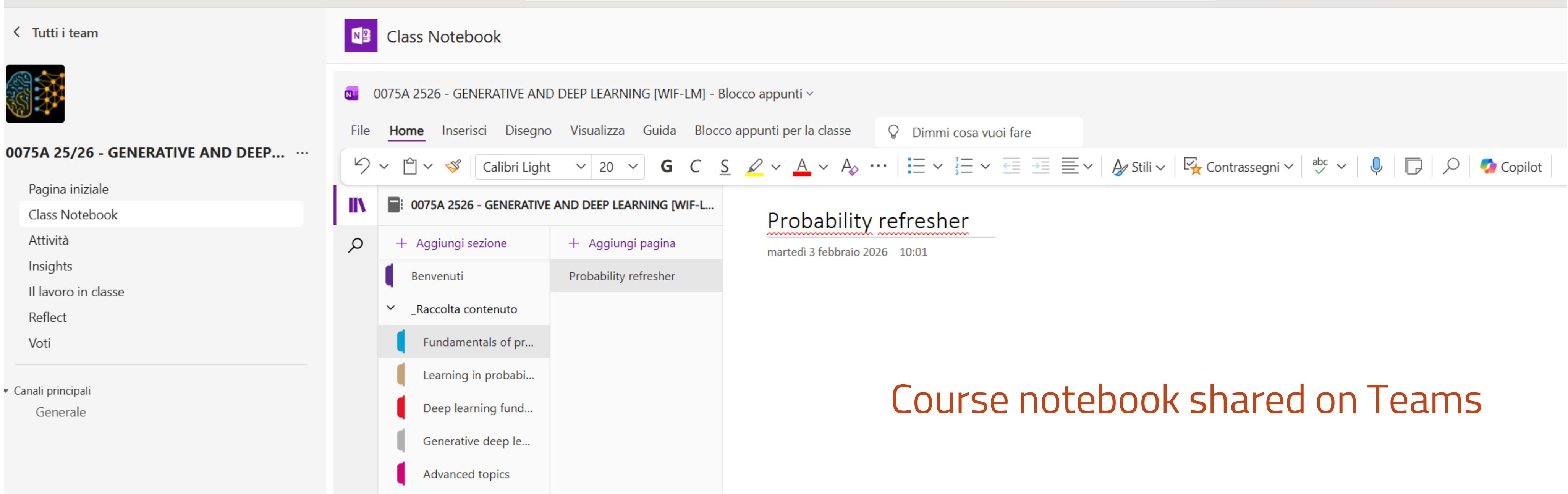
Course MS Team

- The official MS Teams for the course is here (team code: amr96zj)
- Lectures will not be streamed: Team only used for video recording and sharing



Course MS Team

- The official MS Teams for the course is [here](#) (team code: amr96zj)
- Lectures will not be streamed: Team only used for video recording and sharing



Course notebook shared on Teams

Course Homepage

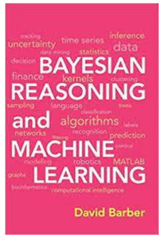
- Reference webpage on Moodle:

  https://elearning.di.unipi.it/course/view.php?id=1128

- Here you will find
  - Course information & news
  - Lecture slides
  - References to the chapters of the official textbooks
  - Additional course materials: scientific papers, interesting code and libraries, …
  - Midterms and final project assignments

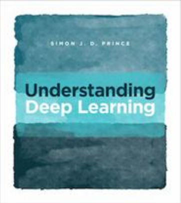
Subscribe to the Moodle to receive **feeds and news**

Reference Books



Probabilistic models reference (free pdf, with code):

David Barber, **Bayesian Reasoning and Machine Learning**, Cambridge University Press (2012)



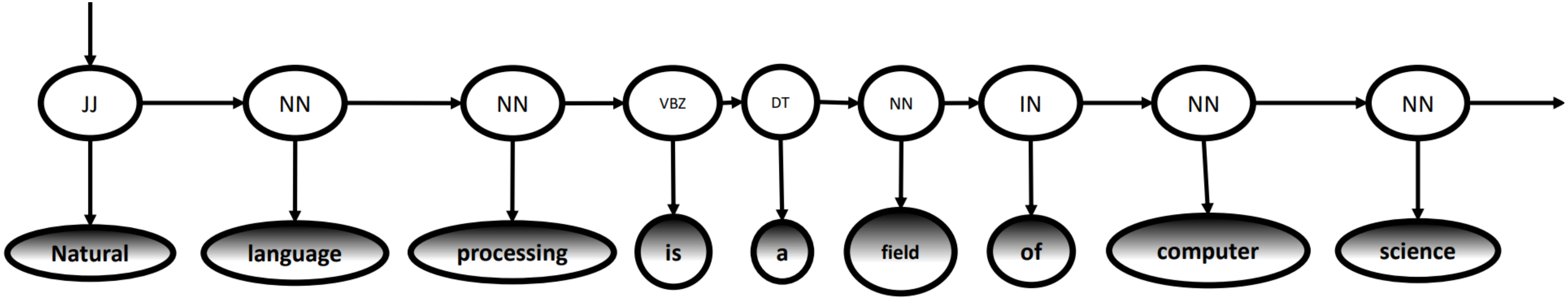
Deep learning reference (free pdf, with additional materials):

Simon J.D. Prince, **Understanding Deep Learning**, MIT Press (2023)

For reinforcement learning and deep learning for graphs refer to lecture slides (and additional materials)

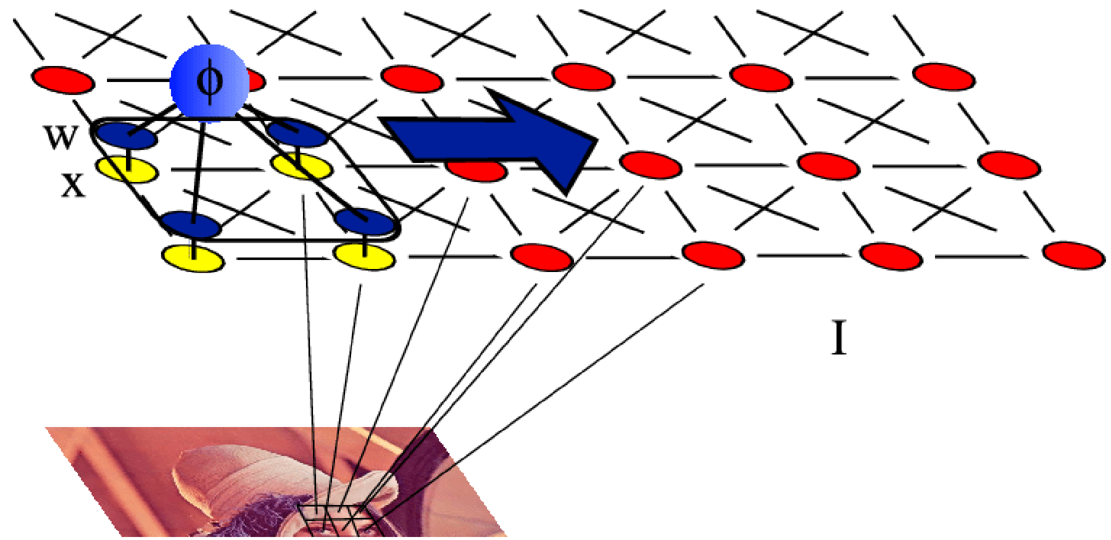
At the roots of probabilistic learning

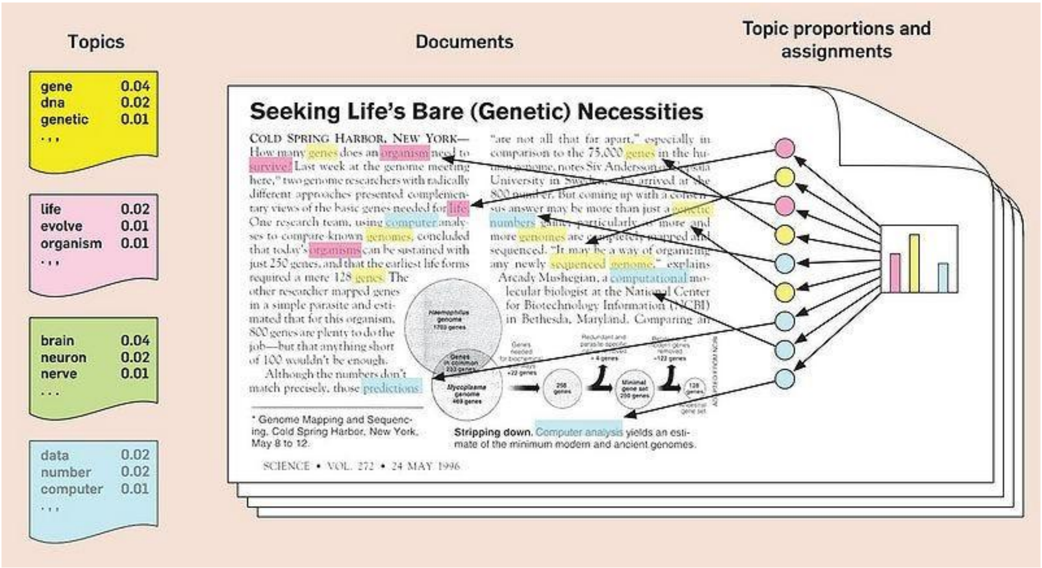
The dawn of language modelling with Hidden Markov Models



Probabilistic learning in information understanding

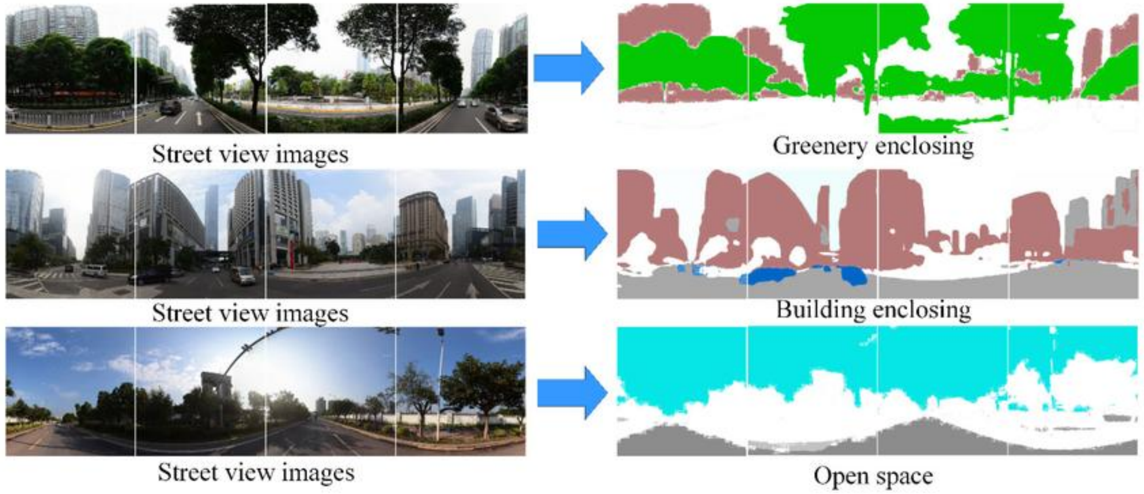
Markov Random Fields for image segmentation





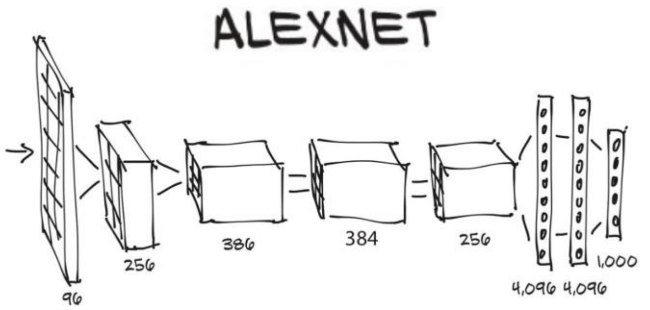
Latent variable models for text understanding

From Probabilistic to Deep Learning



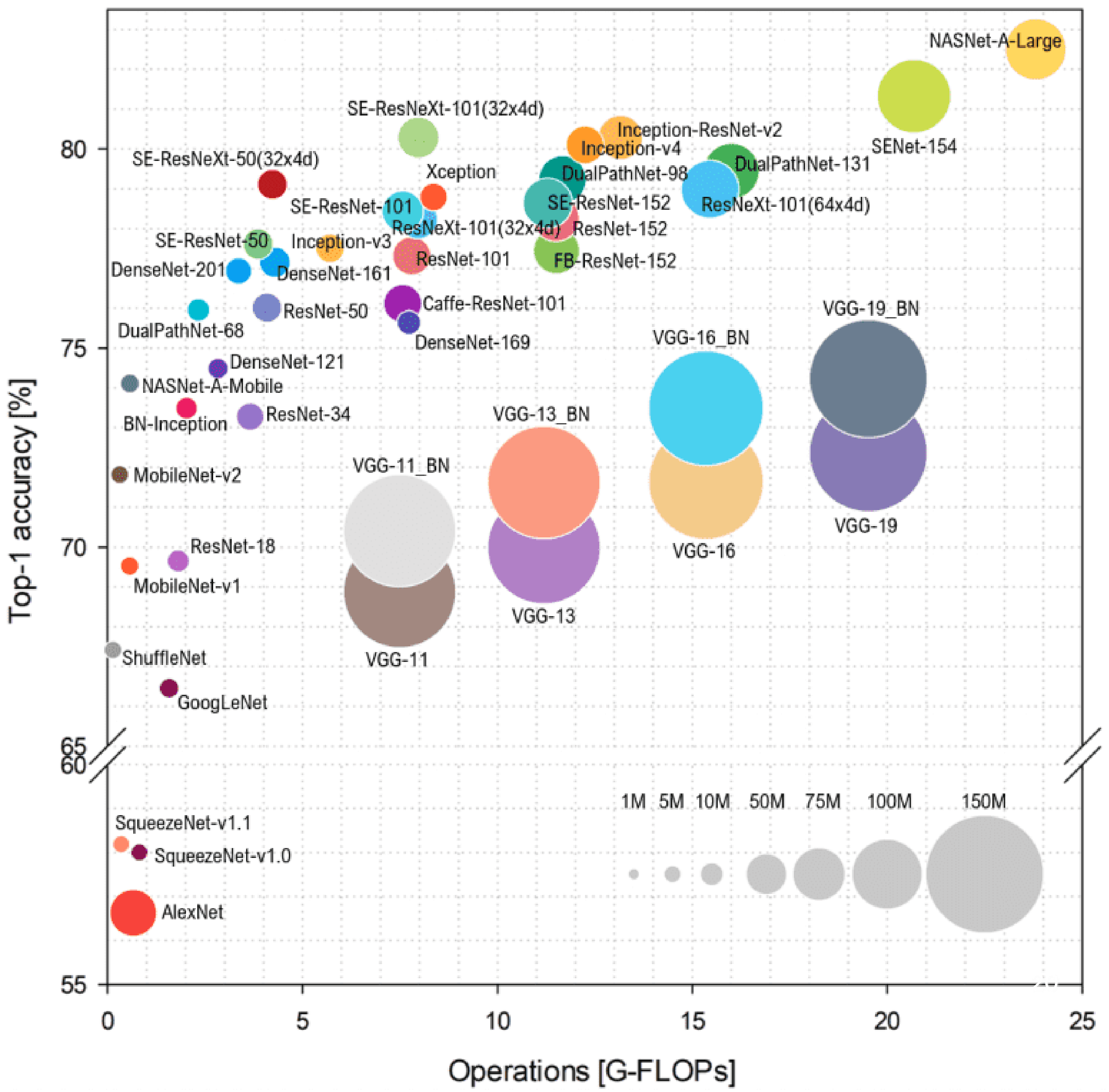
Latent variable models for image semantic segmentation

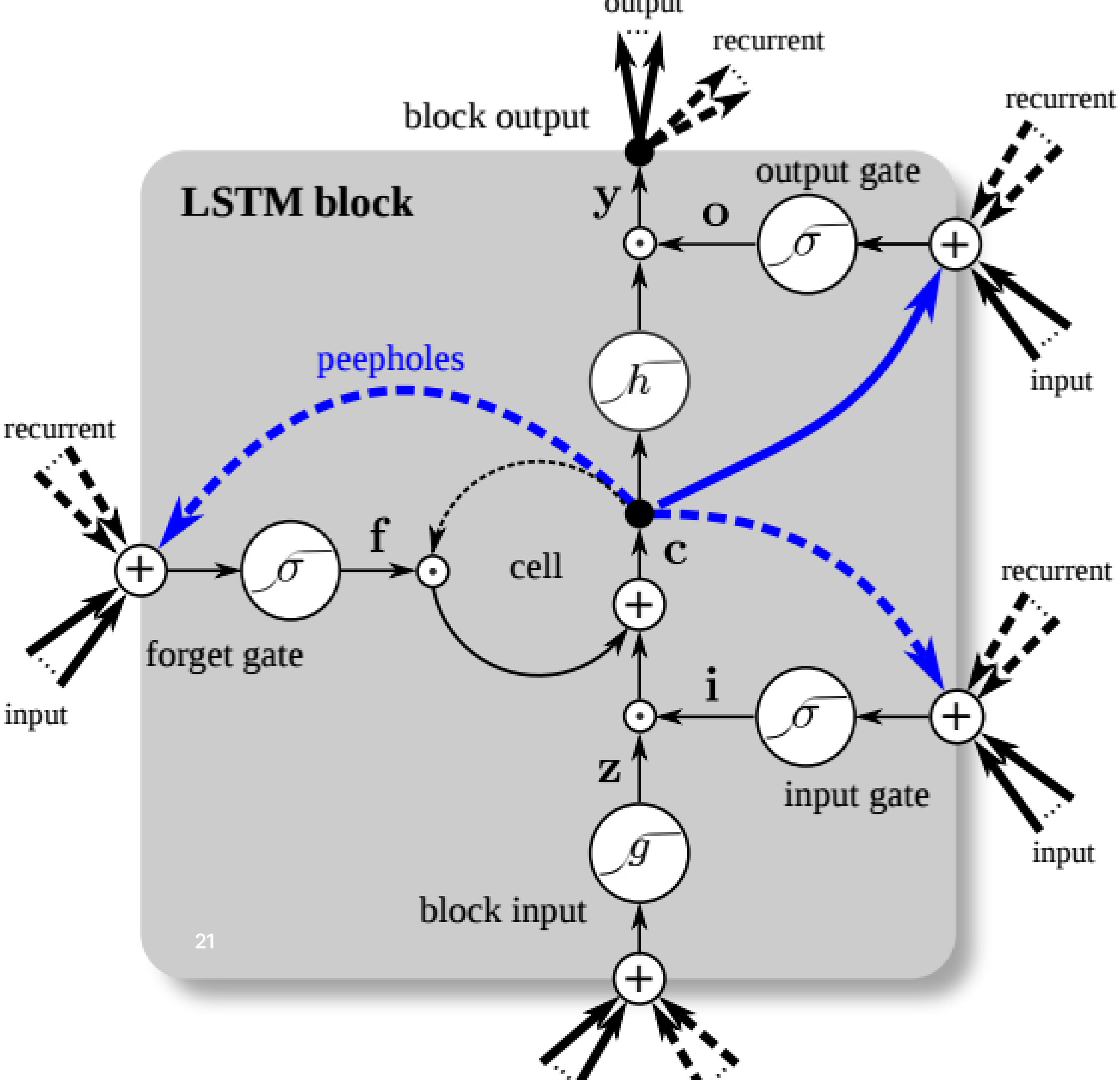
Prof. Fei Fei Li



Convolutional neural networks winning image classification competitions

Then.. things started going offhand



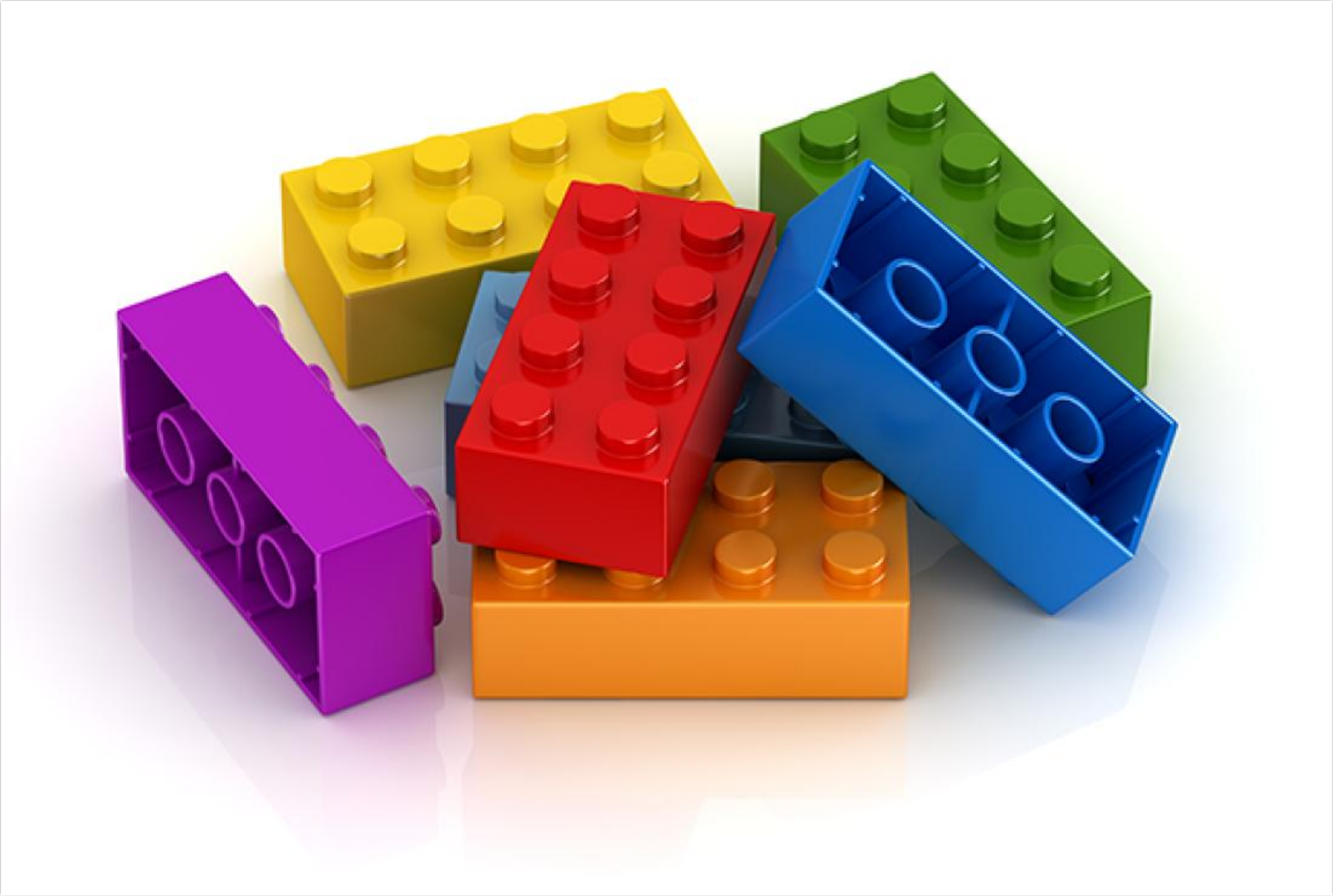


Early language modelling with deep learning

Long Short Term Memory (LSTM): processing sequences and attempting at rescuing gradients since early '90s

The Deep Learning Lego

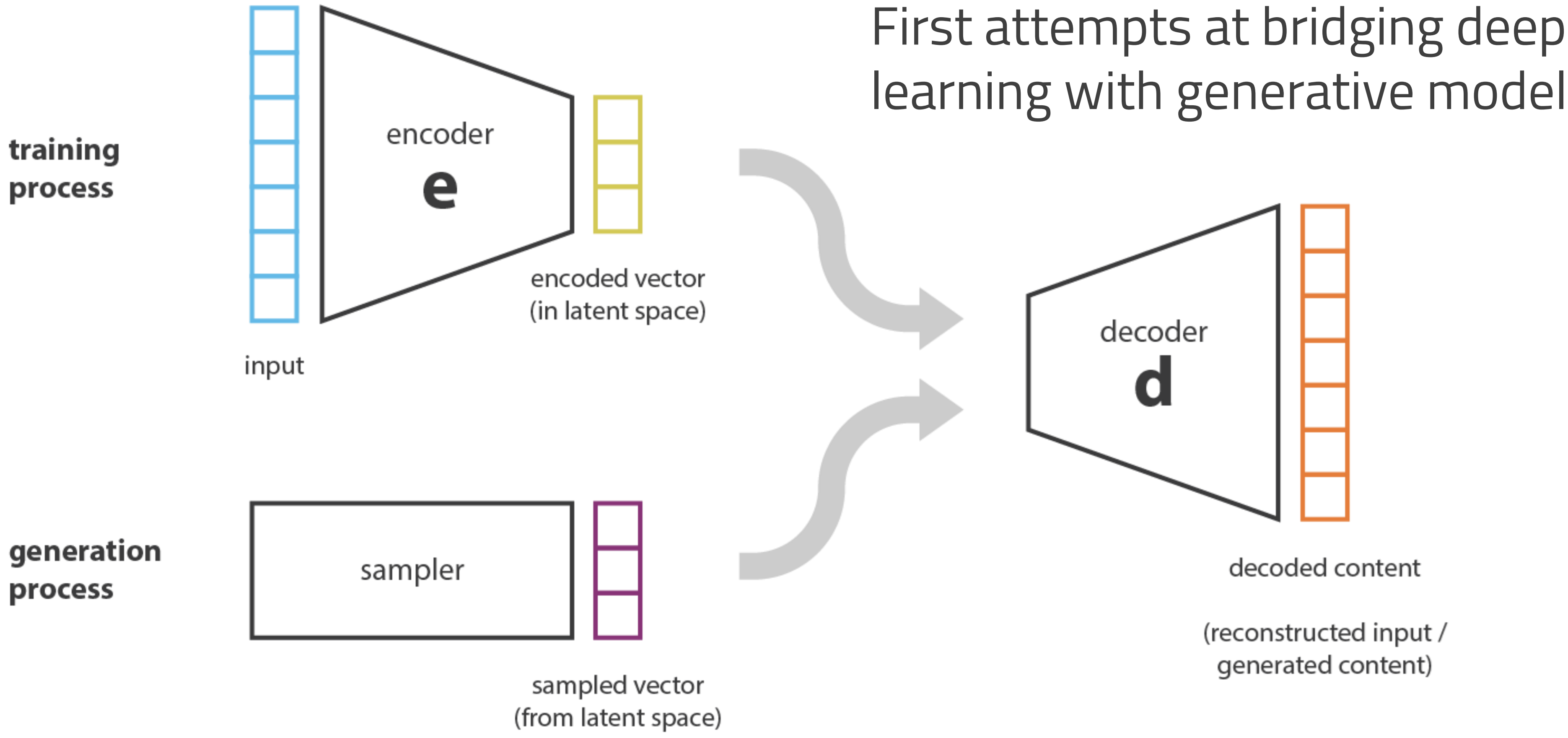
Creating application by putting together various combinations of learning modules



Composing neural and probabilistic models

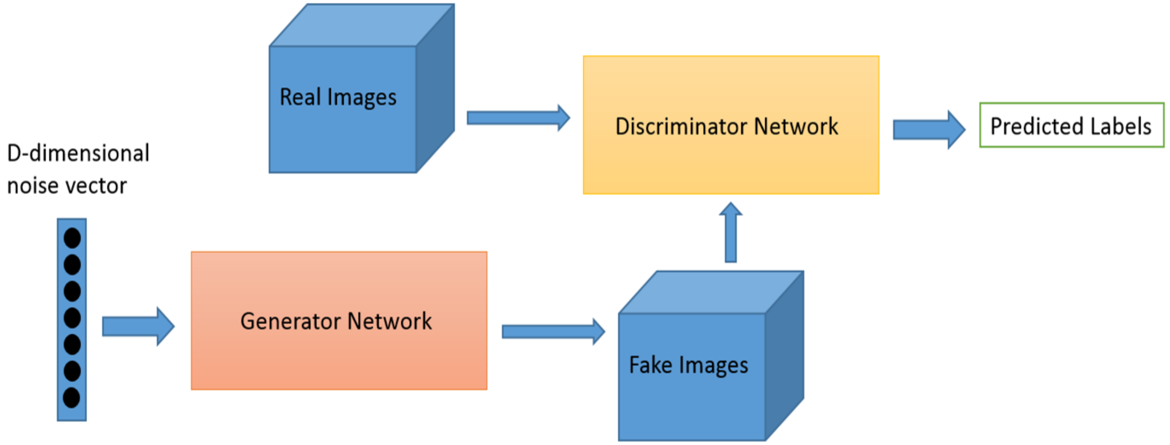
Variational Deep Learning

First attempts at bridging deep learning with generative models



The rising of generative deep learning

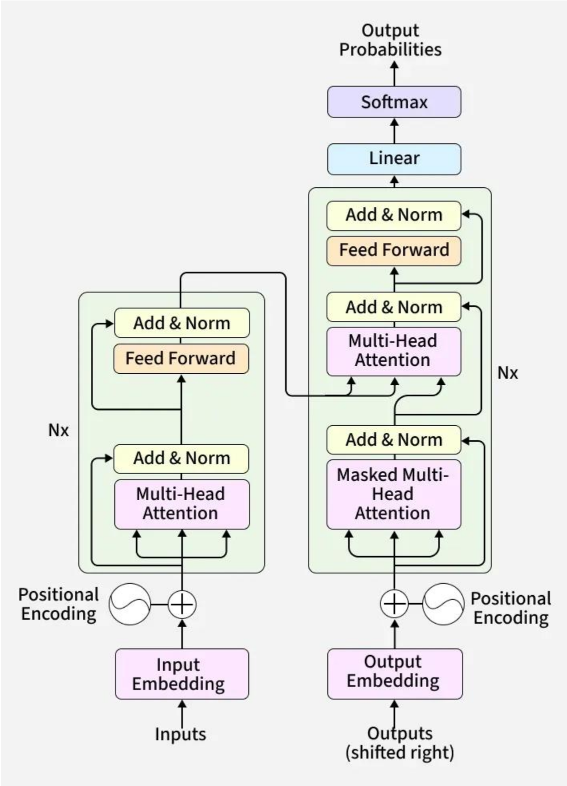
Generative Adversarial Networks…



…and their bedrooms

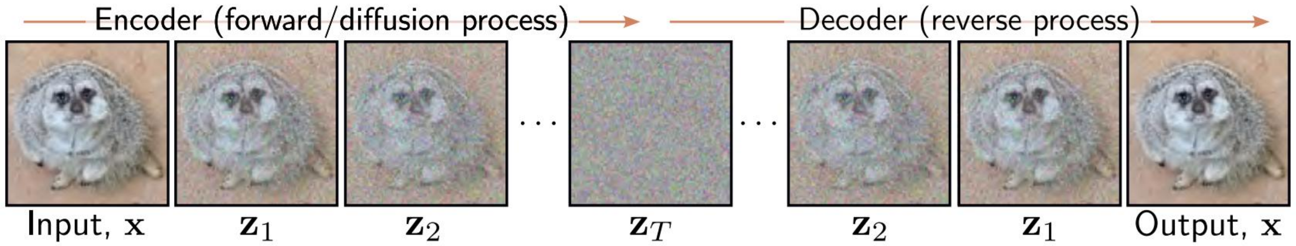
GANs starting to get better at face generation

But nowadays nobody cares because we have…



Transformers and…

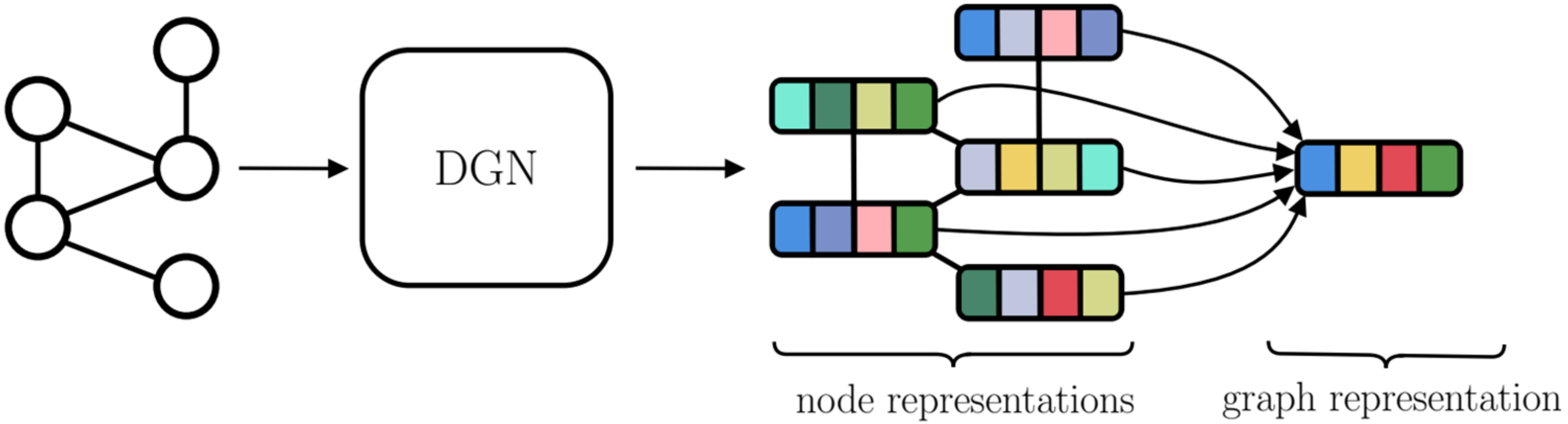
…diffusion models!



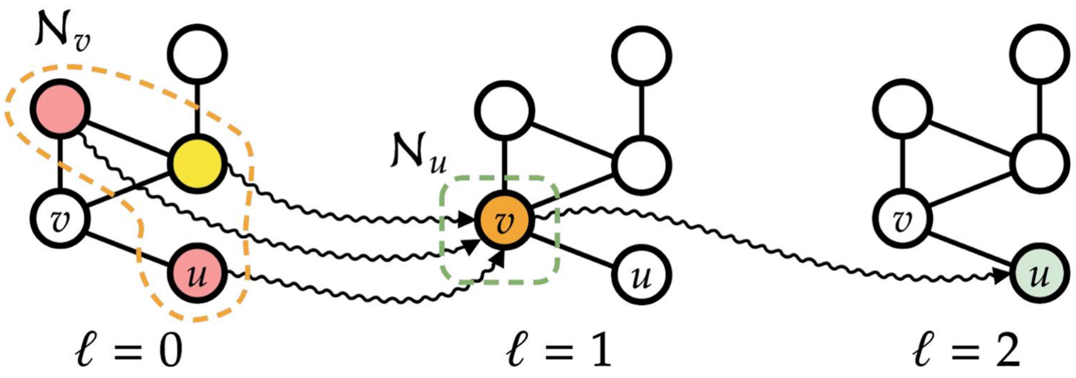
Also for videos

With a somewhat weird passion for bears and guitars…

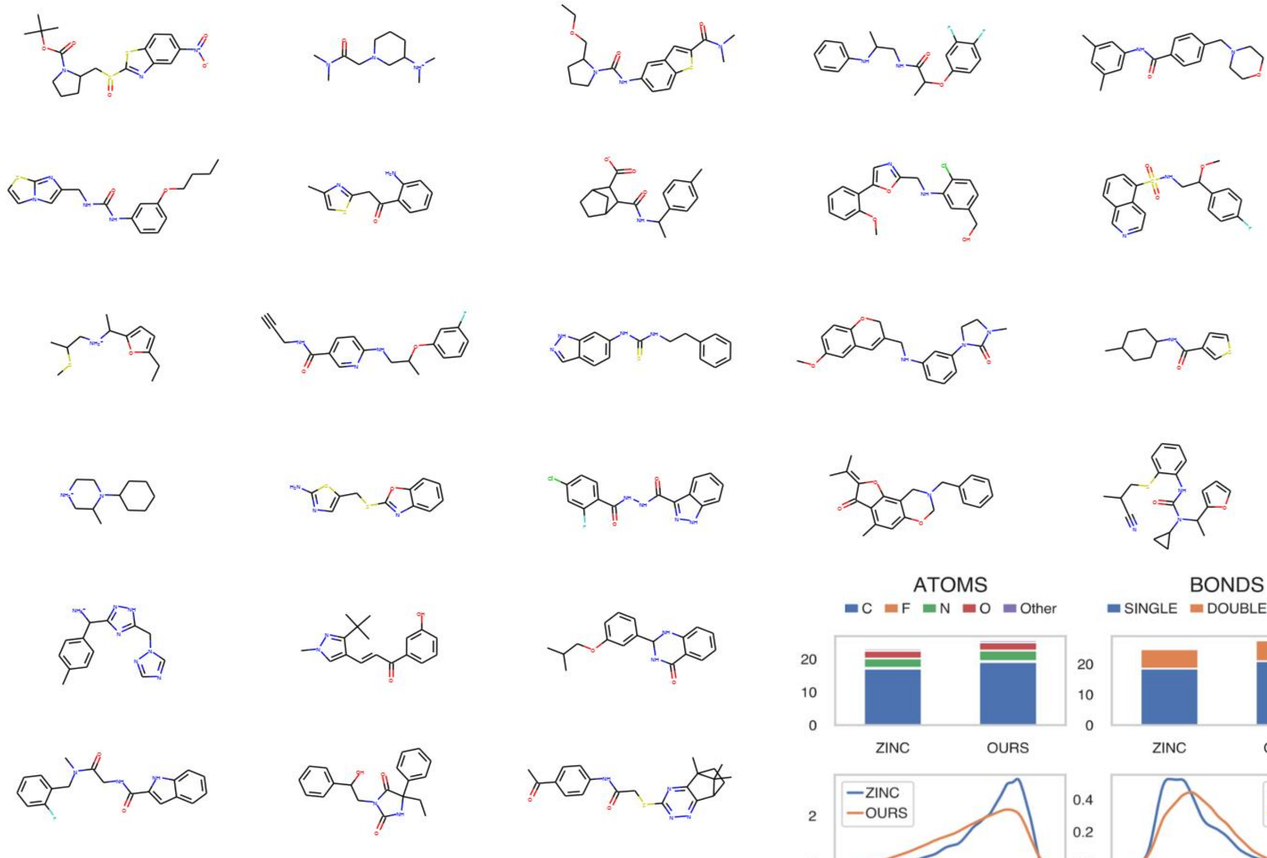
Graph Neural Networks

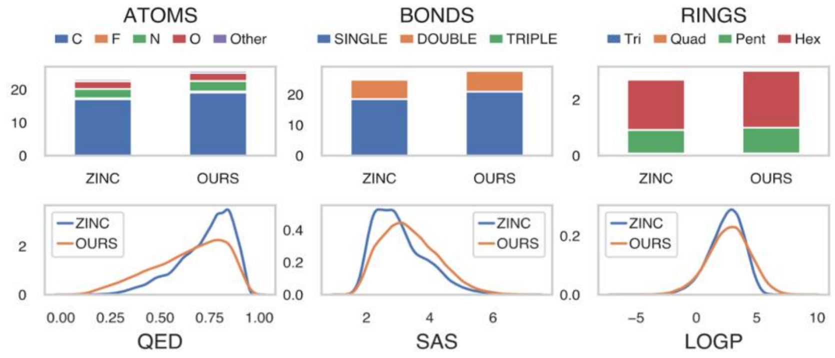


A growing deep learning field reaching maturity



Comes with its own generative challenges

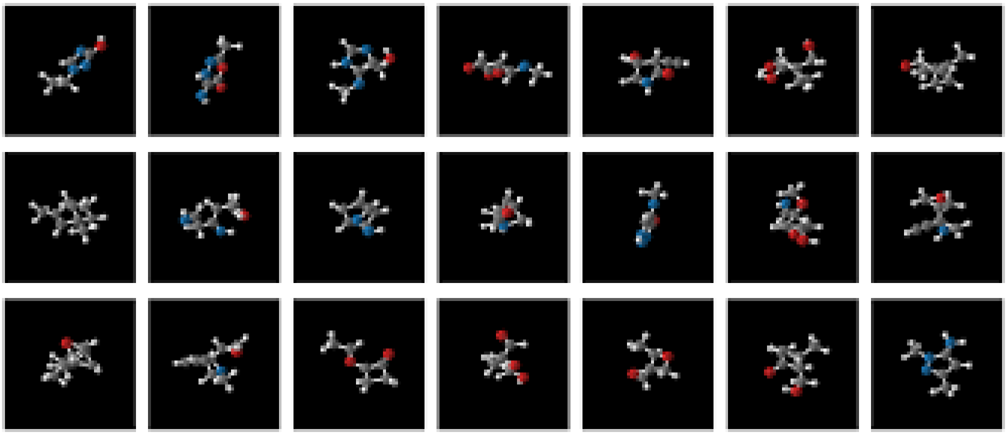




Clock is ticking also for diffusion models

Credit to animation here

Because flow-based models are knocking at the door



The Course Philosophy

- Start from traditional probabilistic graphical models
  - How to reason in terms of probabilities
- Introduce learning as an inference problem
  - Learning density functions
  - Learn some useful approximation techniques
- Introduce powerful adaptive models (deep neural networks)
  - Choosing the right inductive bias for the task and the data
  - How to make them work despite the nasty numerical issues
- Using deep models to approximate densities (generative deep learning)
  - Understand how probability is fundamental to modern machine learning
  - Connect the dots between traditional generative models and deep learning

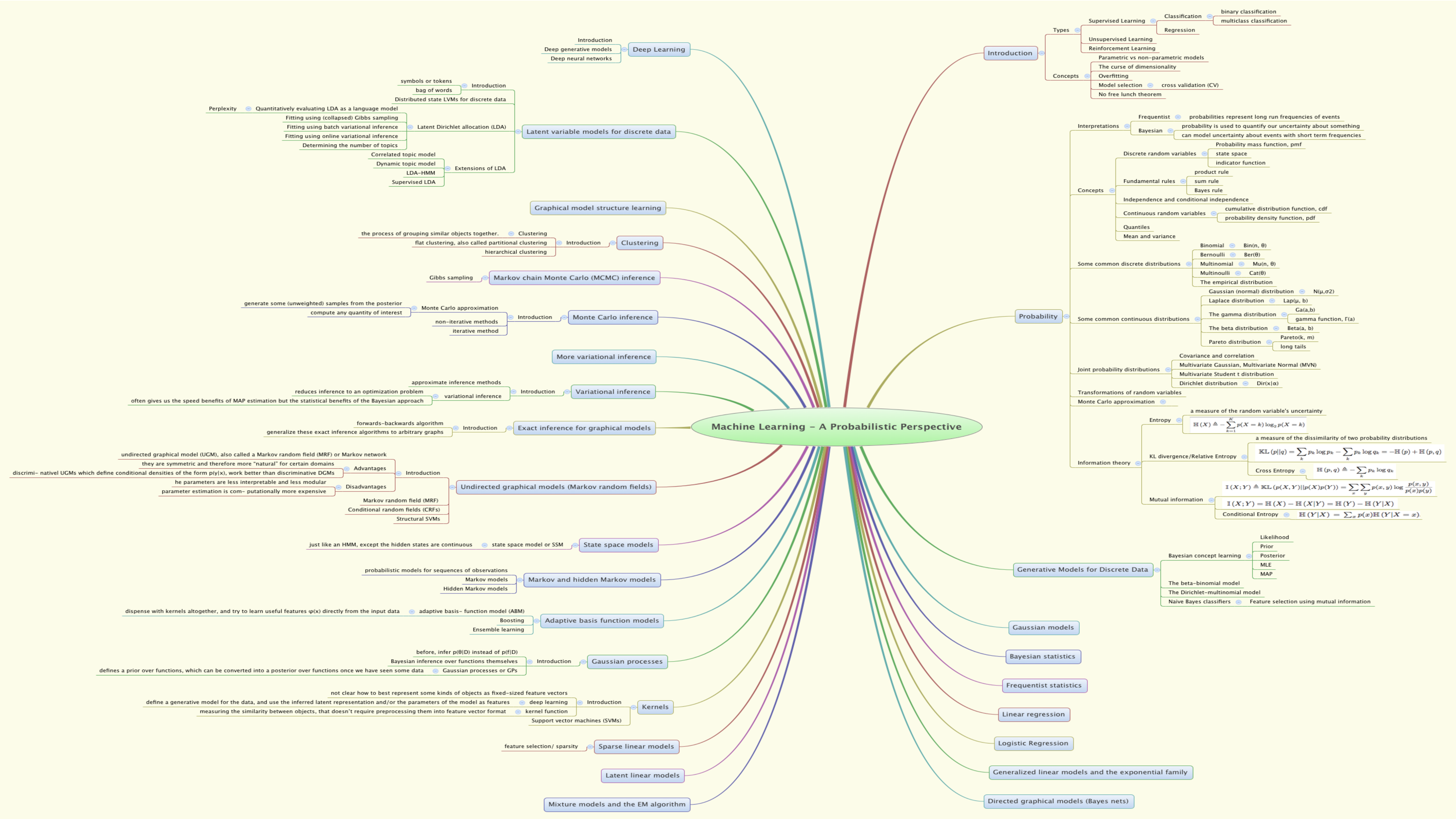
A practical approach with code complementing theory when possible

On probabilities and pains of the sort

From student anonymous advices (ISPR some 4-6 years ago)

- The course should only briefly mention probabilistic models and focus on state-of-the-art models (convolutional neural networks, gated recurrent networks, …)
- Too much attention on generative models, which are not state of the art….
- …more lectures on GANs…

Don’t obsess on models: focus on first principles!


Machine Learning - A Probabilistic Perspective
Introduction
Types
Supervised Learning
Classification
binary classification
multiclass classification
Regression
Unsupervised Learning
Reinforcement Learning
Concepts
Parametric vs non-parametric models
The curse of dimensionality
Overfitting
Model selection
cross validation (CV)
No free lunch theorem
Probability
Interpretations
Frequentist
probabilities represent long run frequencies of events
Bayesian
probability is used to quantify our uncertainty about something
can model uncertainty about events with short term frequencies
Concepts
Discrete random variables
Probability mass function, pmf
state space
indicator function
Fundamental rules
product rule
sum rule
Bayes rule
Independence and conditional independence
Continuous random variables
cumulative distribution function, cdf
probability density function, pdf
Quantiles
Mean and variance
Some common discrete distributions
Binomial
Bin(n, θ)
Bernoulli
Ber(θ)
Multinomial
Mu(n, θ)
Multinoulli
Cat(θ)
The empirical distribution
Some common continuous distributions
Gaussian (normal) distribution
N(μ,σ2)
Laplace distribution
Lap(μ, b)
The gamma distribution
Ga(a,b)
gamma function, Γ(a)
The beta distribution
Beta(a, b)
Pareto distribution
Pareto(k, m)
long tails
Joint probability distributions
Covariance and correlation
Multivariate Gaussian, Multivariate Normal (MVN)
Multivariate Student t distribution
Dirichlet distribution
Dir(x|α)
Transformations of random variables
Monte Carlo approximation
Information theory
Entropy
a measure of the random variable's uncertainty
$\mathbb{H}(X) \triangleq -\sum_{k=1}^{K} p(X=k)\log_2 p(X=k)$
KL divergence/Relative Entropy
a measure of the dissimilarity of two probability distributions
$\mathbb{KL}(p||q) = \sum_k p_k \log p_k - \sum_k p_k \log q_k = -\mathbb{H}(p) + \mathbb{H}(p,q)$
Cross Entropy
$\mathbb{H}(p,q) \triangleq -\sum_k p_k \log q_k$
Mutual information
$\mathbb{I}(X;Y) \triangleq \mathbb{KL}(p(X,Y)||p(X)p(Y)) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$
$\mathbb{I}(X;Y) = \mathbb{H}(X) - \mathbb{H}(X|Y) = \mathbb{H}(Y) - \mathbb{H}(Y|X)$
Conditional Entropy
$\mathbb{H}(Y|X) = \sum_x p(x)\mathbb{H}(Y|X=x)$
Generative Models for Discrete Data
Bayesian concept learning
Likelihood
Prior
Posterior
MLE
MAP
The beta-binomial model
The Dirichlet-multinomial model
Naive Bayes classifiers
Feature selection using mutual information
Gaussian models
Bayesian statistics
Frequentist statistics
Linear regression
Logistic Regression
Generalized linear models and the exponential family
Directed graphical models (Bayes nets)
Deep Learning
Introduction
Deep generative models
Deep neural networks
Latent variable models for discrete data
Introduction
symbols or tokens
bag of words
Distributed state LVMs for discrete data
Latent Dirichlet allocation (LDA)
Quantitatively evaluating LDA as a language model
Perplexity
Fitting using (collapsed) Gibbs sampling
Fitting using batch variational inference
Fitting using online variational inference
Determining the number of topics
Extensions of LDA
Correlated topic model
Dynamic topic model
LDA-HMM
Supervised LDA
Graphical model structure learning
Clustering
Introduction
Clustering
the process of grouping similar objects together.
flat clustering, also called partitional clustering
hierarchical clustering
Markov chain Monte Carlo (MCMC) inference
Gibbs sampling
Monte Carlo inference
Introduction
Monte Carlo approximation
generate some (unweighted) samples from the posterior
compute any quantity of interest
non-iterative methods
iterative method
More variational inference
Variational inference
Introduction
approximate inference methods
variational inference
reduces inference to an optimization problem
often gives us the speed benefits of MAP estimation but the statistical benefits of the Bayesian approach
Exact inference for graphical models
Introduction
forwards-backwards algorithm
generalize these exact inference algorithms to arbitrary graphs
Undirected graphical models (Markov random fields)
Introduction
undirected graphical model (UGM), also called a Markov random field (MRF) or Markov network
Advantages
they are symmetric and therefore more "natural" for certain domains
discrimi- nativel UGMs which define conditional densities of the form p(y|x), work better than discriminative DGMs
Disadvantages
he parameters are less interpretable and less modular
parameter estimation is com- putationally more expensive
Markov random field (MRF)
Conditional random fields (CRFs)
Structural SVMs
State space models
state space model or SSM
just like an HMM, except the hidden states are continuous
Markov and hidden Markov models
probabilistic models for sequences of observations
Markov models
Hidden Markov models
Adaptive basis function models
adaptive basis- function model (ABM)
dispense with kernels altogether, and try to learn useful features φ(x) directly from the input data
Boosting
Ensemble learning
Gaussian processes
Introduction
before, infer p(θ|D) instead of p(f|D)
Bayesian inference over functions themselves
Gaussian processes or GPs
defines a prior over functions, which can be converted into a posterior over functions once we have seen some data
Kernels
Introduction
not clear how to best represent some kinds of objects as fixed-sized feature vectors
deep learning
define a generative model for the data, and use the inferred latent representation and/or the parameters of the model as features
kernel function
measuring the similarity between objects, that doesn't require preprocessing them into feature vector format
Support vector machines (SVMs)
Sparse linear models
feature selection/ sparsity
Latent linear models
Mixture models and the EM algorithm

Reference (Artificial) Languages

Reference language for the course is Python (needed for midterms!)

- Students of the AI curriculum should be already familiar with
- Easy-to-learn language enhanced by reasonable editors and graphical environments
- Lots of library support for deep learning

For the final project there is some reasonable flexibility in which language you can use (no deep learning in Pascal, please!)

Exams – M.Sc. Students

M.Sc. students following the course lectures can complete the exam by

- **Midterm Assignments** - A total of 4 short assignments on experiences related to course topics
- **Oral Exam** - An examination on the course program

The traditional way (for non-attending students, those who fail or don't like the other way)

- **Final Project** - A written report on the implementation of a non-trivial learning model and/or an AI-based application relevant for the course
- **Oral Exam** - A 15 minutes presentation of the final project plus examination on the course program

Exams – Ph.D. Students

Let's find a topic that is of interest for you, maybe part of your research project, and that is consistent with the course topics.

Several options possible:

- **Essay** – A research technical report on the topic of interest
- **Code** – A software exploring/implementing some research model/experiment/benchmark
- **Anything else that makes sense for research...**

No oral exam needed

Midterm Assignments

Assignments 1 to 3 – Individual midterms

- Delivery of a Python notebook about the implementation of a model/method of interest for the course
- Tested for correctness on hidden test set

Assignment 4 – Group midterms (team of 5)

- A poster on a recent development on generative and deep learning
- Presented during a poster session the final week of May

Timeline

- One midterm every 3-4 weeks
- Should be doable with a couple of afternoons' work
- Midterm published: early March, late March, late April, mid May
- Midterm delivered: late March, mid April, mid May, late May

One important note

- Midterms are for students who regularly participate to classes (otherwise you have the traditional final project + oral way)
- There will be processes in place to monitor attendance for midterm submission

Midterm ID	Student ID (Matricola)
1	XXX
2	YYY
…	…

Midterm IDs assigned in person during random lectures before each midterm

- When submitting your midterm, you will name your file with the provided Midterm ID
- Midterm ID will be matched to the Student ID
- Submissions without matching ID will be marked as fail

Final Project

Develop a tested and commented software implementing a non-trivial learning model and/or a DL-based application relevant for the course. Choose from a set of suggested topics or propose your own topic of interest

Timeline

- Suggested topics list published: mid May
- Choose project: email me to arrange a topic
- Deliver: report describing the model and its validation (~10 pages), source code (zipped, no dataset) and presentation (15 slides max)
  - When: by the standard exam date (appello) (strict)

Oral Exam

(Give your presentation on the final project (15 minutes))

- Discuss it in front of me and anybody interested
- Be prepared to answer my questions on the presentation

Only for those who did not do the midterms

An oral exam with questions covering the course contents

- Lectures whose content is not relevant for the final exam will be clearly marked as such

Remember to upload the presentation/report/code on Moodle by the appello deadline

Expectations about the oral exam

Let’s set our mutual expectations right and live happily ever after

- Yes, the oral exam is going to be about methods & models, also from a theoretical perspective
- Yes, I can ask you to derive statements and describe the models/methods formally
- Focus on reasoning not memorization: I will ask questions in ways that will show if you have learned things by hearth
- Structure of the oral discussion:
  - Expect roughly 3 questions of increasing level of complexity (higher complexity==higher grade)
  - If you cannot answer the current question in reasonable time (i.e. some 5-10 minutes) we conclude the oral and you get the grade related to the question(s) you have managed to answer
  - I don’t do on-request random questioning until you get one you can answer
- **Minimum threshold to pass the course requires being able to formalize models seen as part of the course and being able to convincingly discuss their properties and use**

How to get past this course?

Grading (with midterms)

- Midterms only wave the final project and oral presentation: there is no vote for them, only pass/fail
- The exam grade is given by the oral examination grade

Grading (traditional way): $\frac{(G_P+G_O)}{2}$

- $G_P \in [1,30]$ is the project grade
- $G_O \in [1,32]$ is the oral grade

Policies

- You can discuss assignments/projects with others, but we ask that you write your solutions individually
- We support use of genAI (LLMs, coding copilots, etc):
  - **Responsibility for content**: Students who use genAI in their assignments take full responsibility for the content they submit
  - **Acknowledgment of AI use**: Clearly acknowledge any use of genAI, specifying the nature and extent of assistance received from AI (by having a statement at the beginning of the notebook)
  - **Ethical use and originality**: Do not use AI to plagiarize, misrepresent original work, or fabricate data
  - **Instructor discretion**: We may specify assignments for which genAI use is prohibited

Upcoming...

Fundamentals of probabilistic models and causality

Refresh useful knowledge from probability and statistics, and introduce fundamental concepts for working with probabilistic models, including the graphical model's formalism.

Module Topics

- Rules to manipulate probability
- Relevant density functions and their use
- Graphical models representation
- Conditional independence and d-separation
- Causality
- Structure learning from data

Next Lecture

Introduction to probabilistic models

- Probability and statistics refresher
- Notable distributions
- Rules to manipulate probability
- Probabilistic inference in machine learning

Next lecture will be tomorrow Friday 20/02/2026 h. 16-18 in Room E

Recovery lecture outside of standard course time

Onboarding

Remember to register on the course Moodle

https://elearning.di.unipi.it/course/view.php?id=1128

When you send me an email about the course include tag [GDL] in the subject (it may as well end up in thrash, but then it would be due to UNIPI's psychotic spam filter not to me)

Or you may not receive answer if you ask questions on administrative and procedural aspects that I have already explained during the lectures and/or on the course website and materials.

Questions?