Introduction to the ISPR Course

INTELLIGENT SYSTEMS FOR PATTERN RECOGNITION (ISPR)

DAVIDE BACCIU – DIPARTIMENTO DI INFORMATICA - UNIVERSITA’ DI PISA

DAVIDE.BACCIU@UNIPI.IT

Objectives

Train machine learning (ML) specialists capable of

- designing novel learning models
- developing pattern recognition applications using ML
- developing intelligent agents using Reinforcement Learning (RL)

Focus on challenging and complex data

- Machine Vision: noisy, hard-to-interpret, semantically rich information
- Structured data: relational information (sequences, trees, graphs)

Lectures do not cover Natural Language Processing as there is a dedicated course

Expected Outcome

Methodology-oriented outcomes

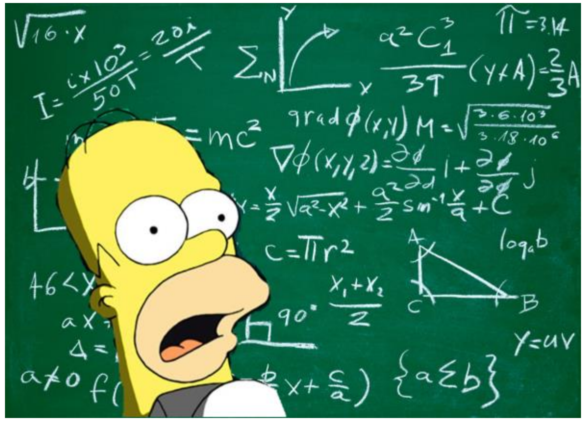
- Gain in-depth knowledge of advanced machine learning models
- Understand the underlying theory
- Be able to individually read, understand and discuss research works in the field

Application-oriented outcomes

- Learn to address modern pattern recognition problems
- Gain knowledge of ML, PR and RL libraries
- Be able to develop an application using ML and RL models

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)



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

Organization

The course covers five themes

- Introduction to Pattern Recognition
- Probabilistic (Generative) Models
- Deep Learning
- Generative Deep Learning
- Advanced models and applications

An incremental approach: from old school pattern recognition to state-of-the-art deep learning

Guest Lectures

Guest seminars by researchers and Ph.D. students on (tentative):

- Practical lectures on deep learning frameworks (PyTorch, TF/Keras, Ray)
- Reservoir computing
- Alternative to backprop
- Neural xDE framework
- Short seminars on hot research topics by guest lecturers
- …

Topics (I)

- Introduction to Pattern Recognition
  - Introduction to signal processing
  - Introduction to image processing
- Probabilistic (Generative) Models
  - Graphical models
  - Bayesian networks and causality
  - Hidden Markov Models
  - Markov Random Fields
  - Bayesian learning and variational inference
  - Sampling
  - Boltzmann machines

Topics (II)

- Deep Learning (DL) fundamentals
  - Deep autoencoders
  - Convolutional architectures
  - Gated recurrent networks
  - Transformers and encoder-decoder architectures
  - DL toolset: dropout, batch normalization, residual connections, attention
  - Neural memories
  - Deep learning with Pytorch and Keras-TF
- Generative deep learning
  - Exact likelihood models
  - Variational AE
  - Generative adversarial networks
  - Normalizing flow
  - Diffusion models

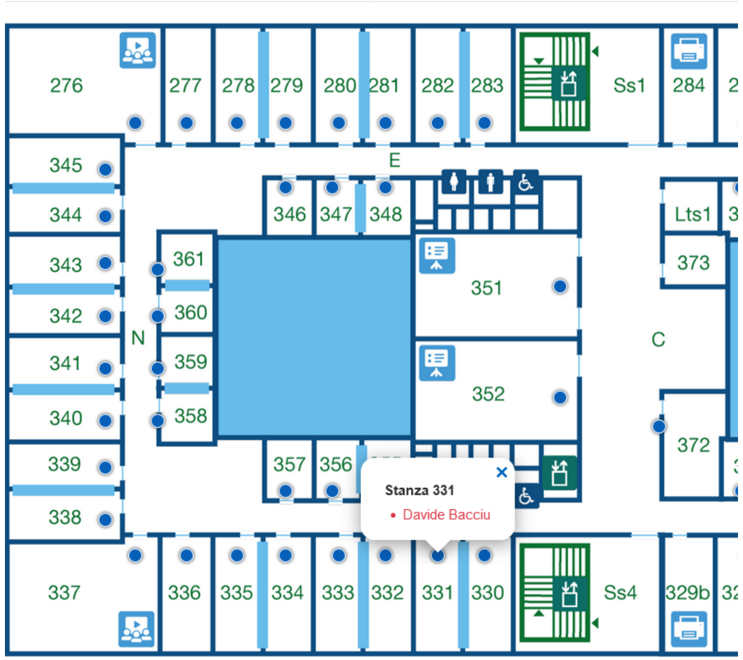
Topics (III)

- Advanced topics and applications
  - Reservoir computing
  - Dynamical systems and neural networks
  - Alternatives to backprop
  - Deep learning for graphs
  - Reinforcement learning
  - Machine vision, multimodal learning, BioInformatics, robotics,...
  - ...

Course Instructor

Davide Bacciu

- Email – davide.bacciu@unipi.it
- Tel - 050 2212749
- Office - Room 331O, Dipartimento di Informatica
- Office hours - Thursday 16-18 (email me!)



UNIVERSITÀ DI PISA

Course Schedule

Weekly Timetable:

Day	Time
Tuesday	11.00-12.45
Wednsday	16.15-18.00
Thursday	14.15-16.00

Talk now if there are incredibly worrisome issues with the schedule!

Course comprises 35-36 lectures

- Course will be given in-person and streamed online on Teams for Ph.D. students
- Video recording of the lectures will be available (to everybody) on Teams

Course Homepage

Reference Webpage on Moodle:

https://elearning.di.unipi.it/course/view.php?id=278

Here you can find

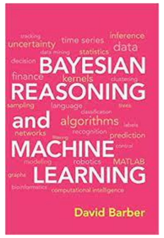
- Course information
- Lecture slides
- Articles and course materials
- Midterms and final project assignments

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

Reference Books (changing this year)

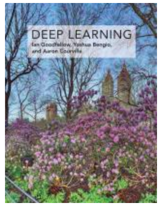
For pattern recognition refer to slides (and additional material)

Previous years



Probabilistic learning reference (free pdf):

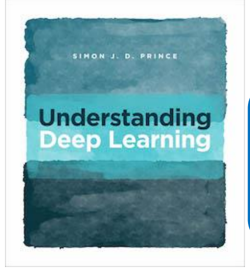
David Barber, **Bayesian Reasoning and Machine Learning**, Cambridge University Press (2012)



Deep learning reference (free pdf):

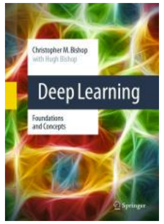
I. Goodfellow, Y. Bengio and A. Courville, **Deep Learning**, MIT Press (2016)

Starting this year



Deep learning (free pdf):

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



Probabilistic & deep learning (free pdf):

Chris Bishop, Hugh Bishop, **Deep Learning Foundations and Concepts** , Springer (2024)

I will keep reference to both sets of books for this year

UNIVERSITÀ DI PISA

The Origins of Pattern Recognition (PR)

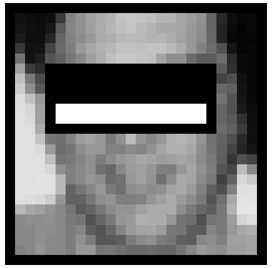
Duda and Hart, 1973
Machine recognition of meaningful regularities in noisy or complex environments

A variety of approaches to realize it

- Statistical PR
- Clustering
- Rule-based systems (fuzzy)
- Signal processing
- Logic and reasoning
- Structural and syntactic PR
- ...and of course, machine learning!

The Viola-Jones Algorithm

Consider the following two hand drawn pixel masks



VJ1

VJ2

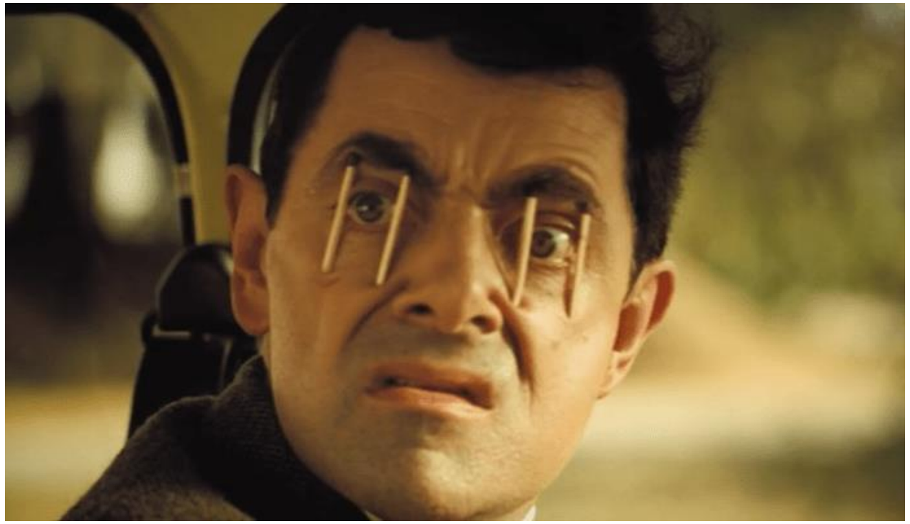
Sum pixels in the white area and subtract those in the black portion

- VJ1 is large in the eye region
- VJ2 is large on the nose stripe

VJ algorithm positions the masks on the image and combines the responses (≈ 5K hand aligned examples)

PR Stages – An historical View

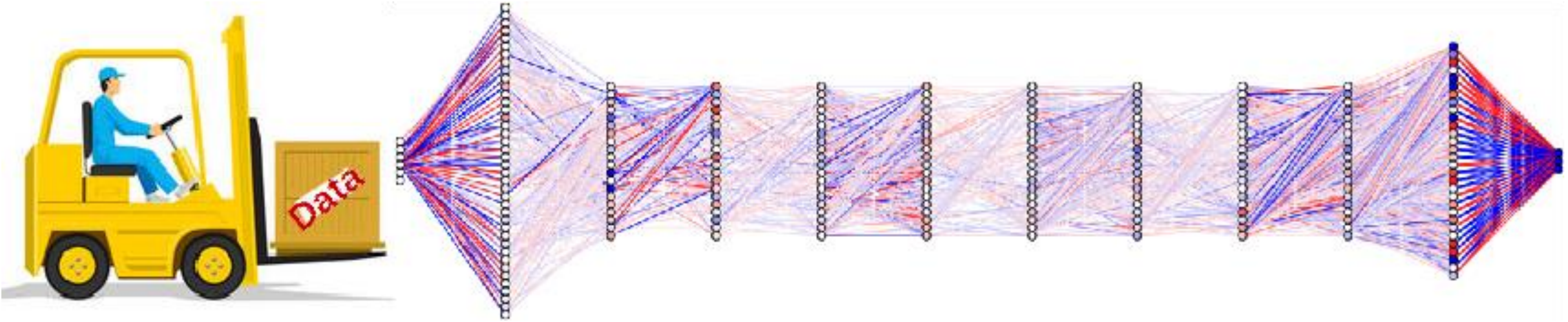
1. Identification of distinguishing attributes of the object/entity (feature detection)
2. Extraction of features for the defining attributes (feature extraction)
3. Comparison with known patterns (matching)



Basically, lots of sleepless nights hand-engineering the best data features

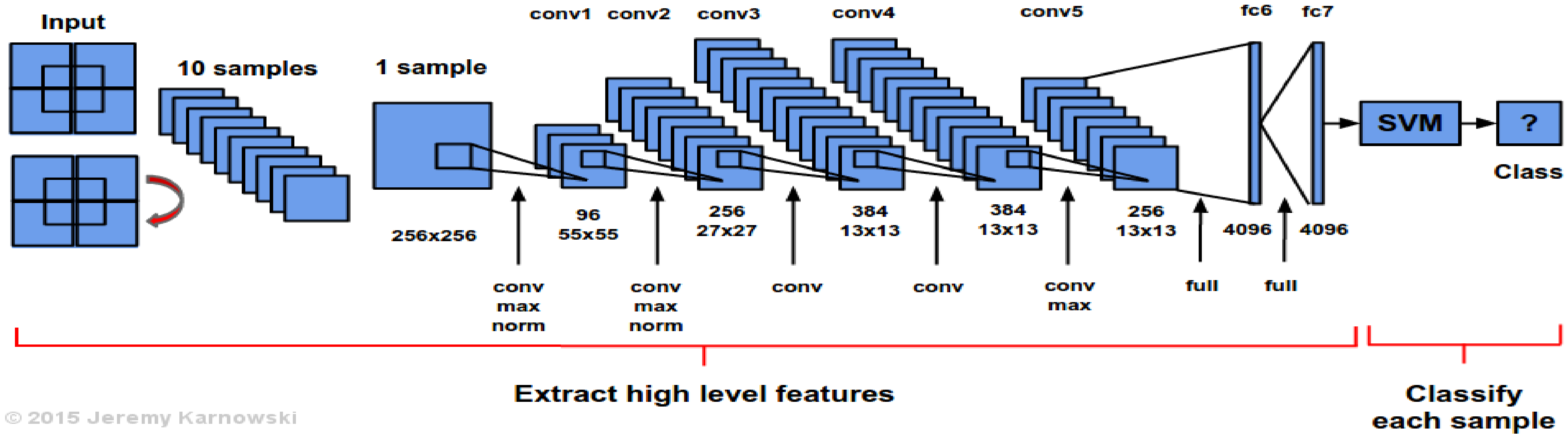
PR Stages - A Modern View

Pattern recognition after the deep learning revolution



Apparently, a single stage process with a data crushing-and-munching neural monster spitting out predictions

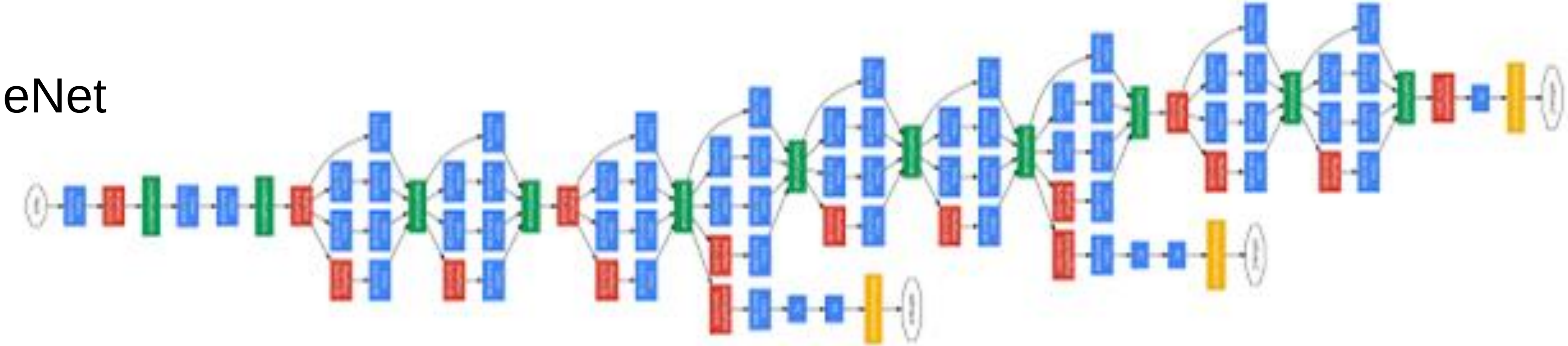
The Dawn of the Revolution



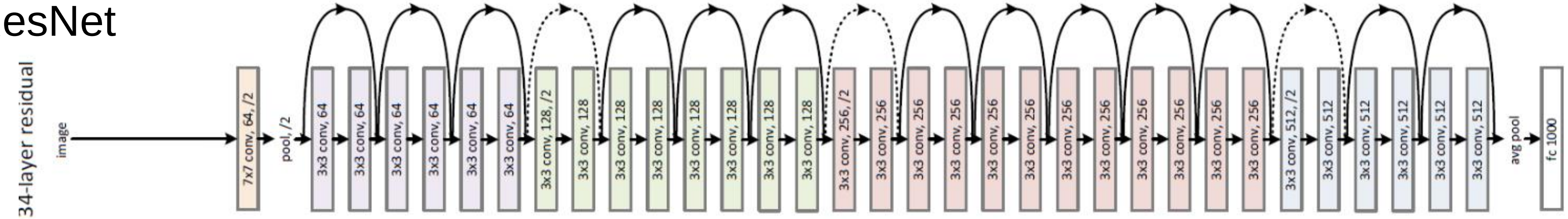
AlexNet kills the ImageNet 2012 competition outperforming runner-up by over 10%

Then.. Things Started Going Offhand

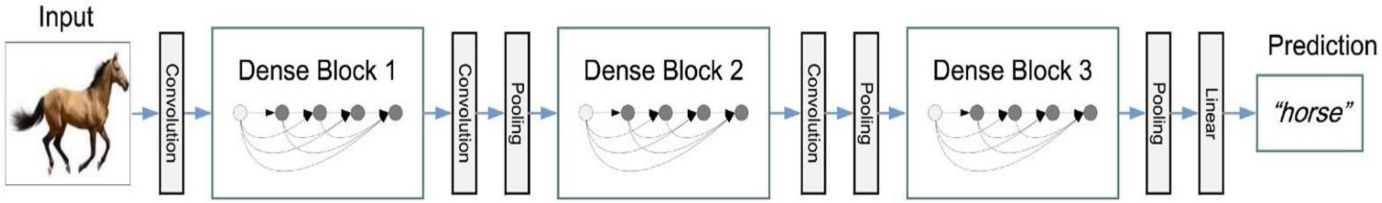
GoogleNet



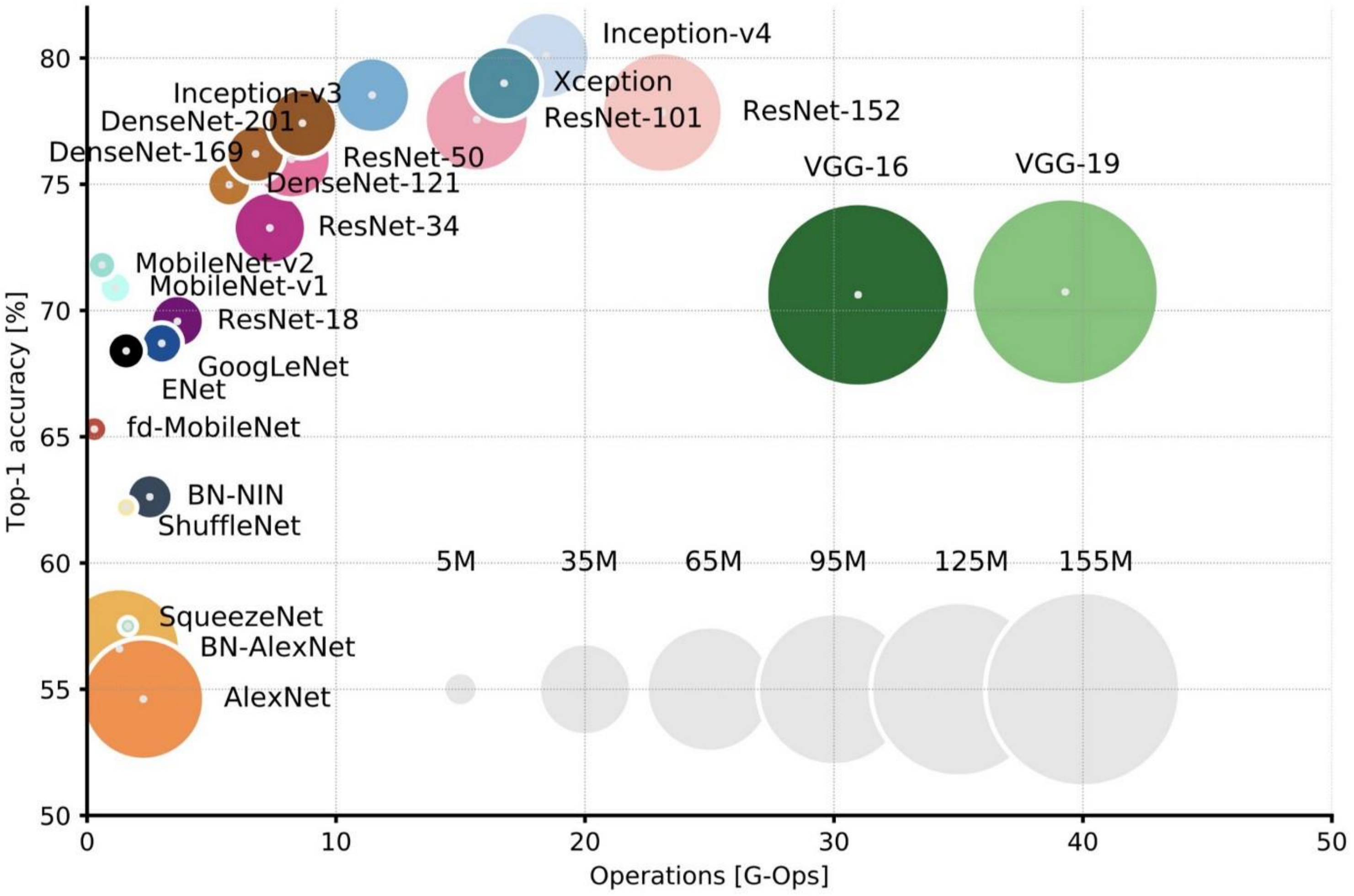
ResNet



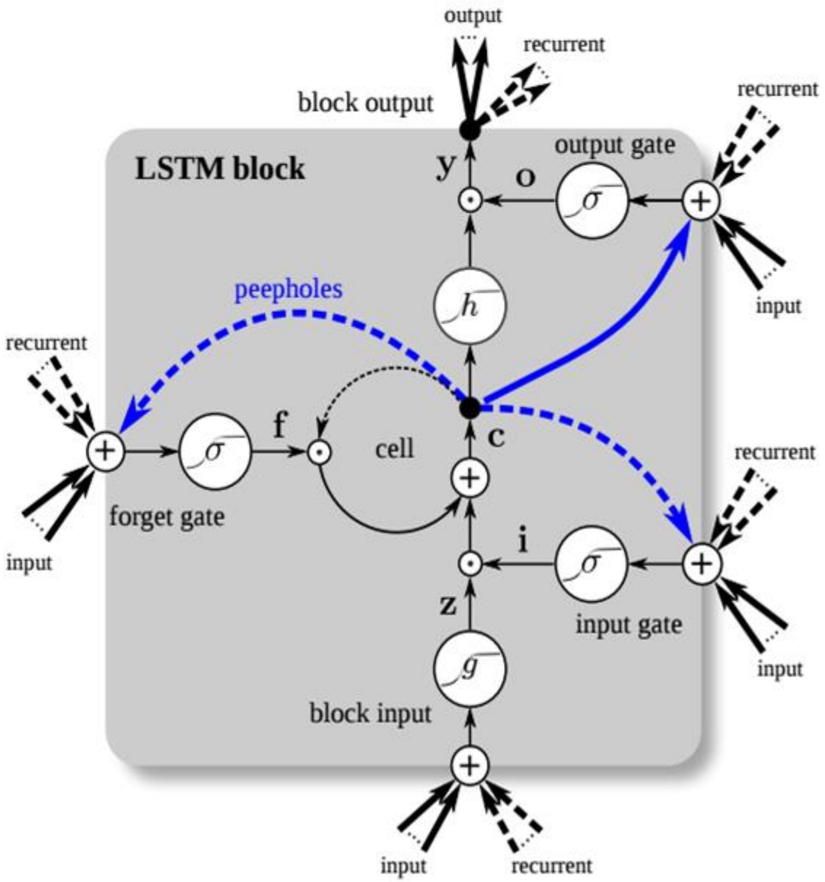
DenseNet



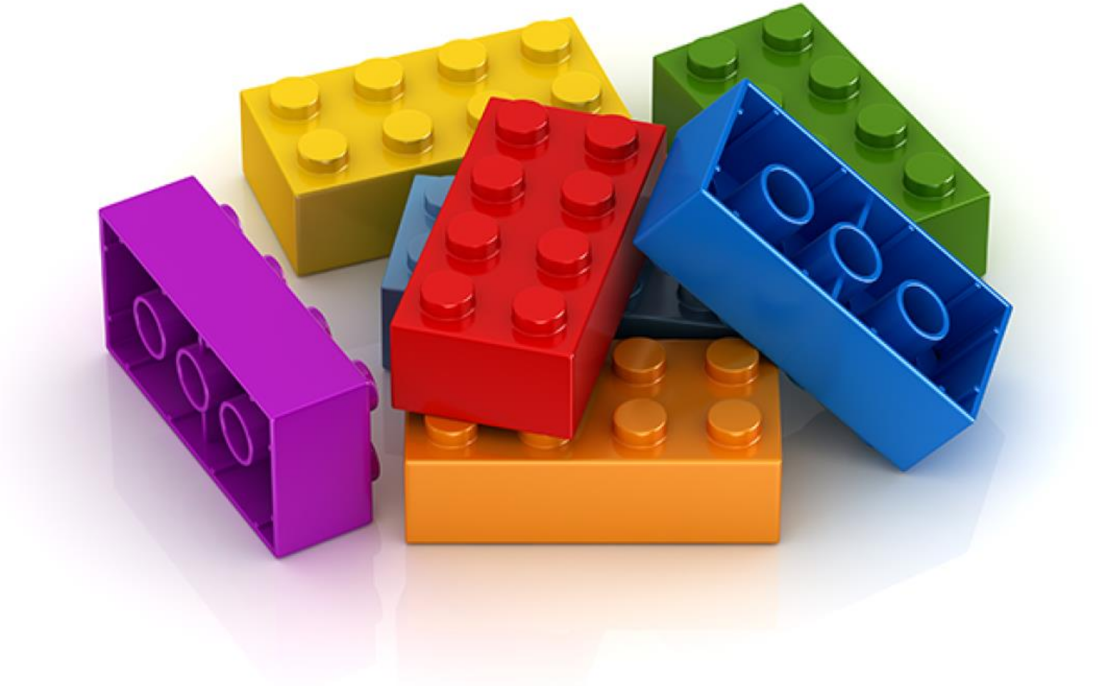
CNN Evolution



Long Short Term Memory



Processing sequences and rescuing gradients since 1996

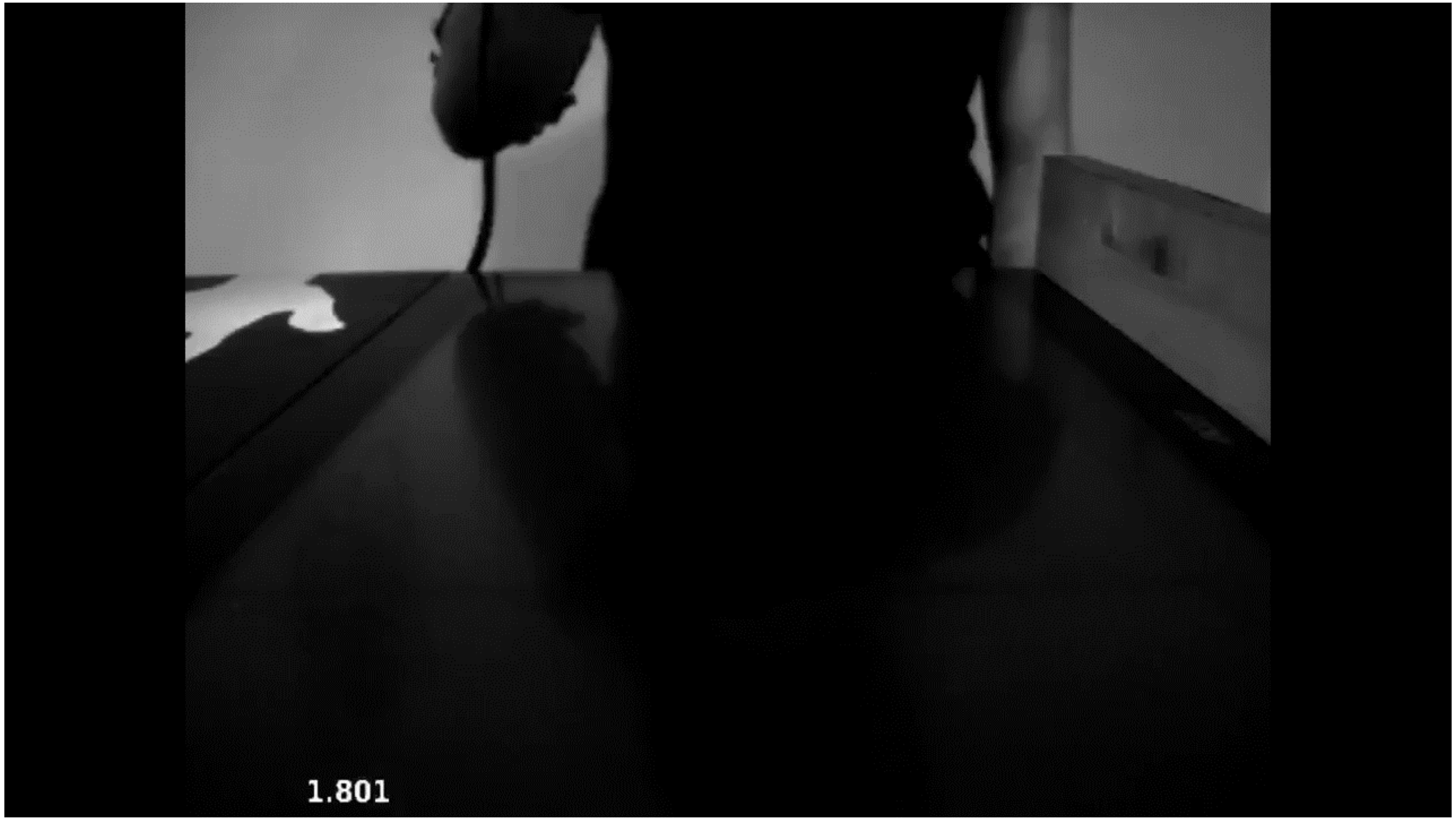


The Deep Learning Lego

Creating application by putting together various combinations of CNN and LSTM modules

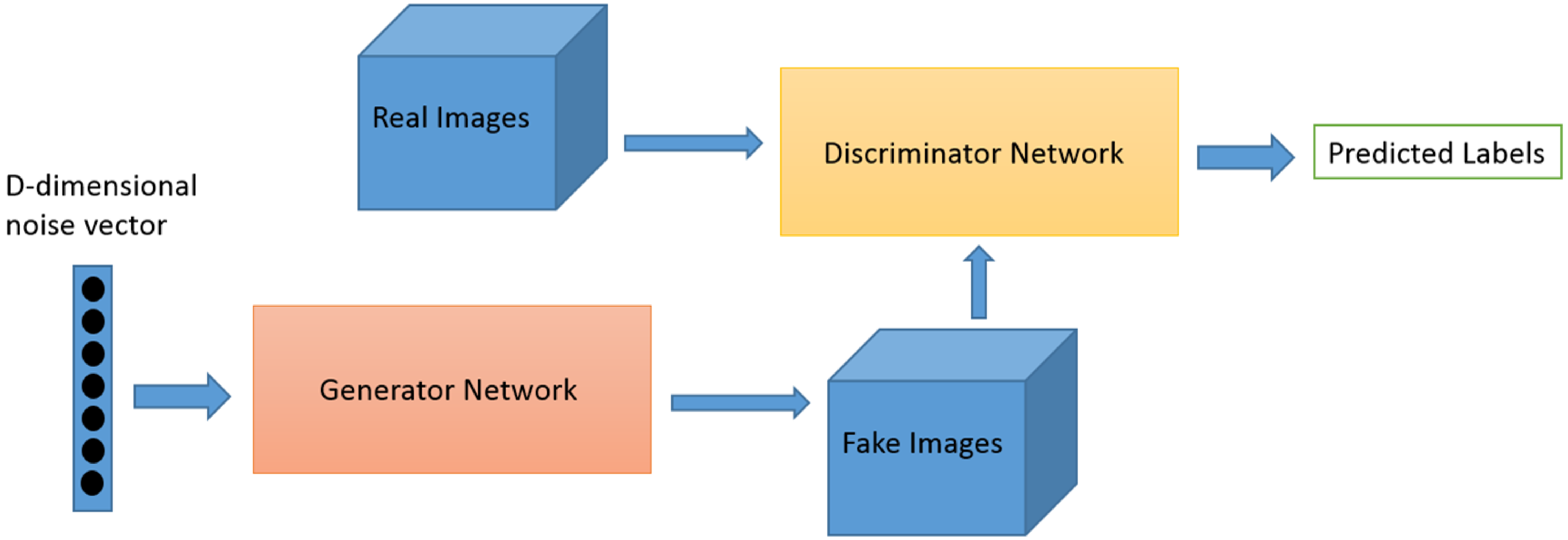
Autonomous Driving

Teaching Robots to Manipulate



Teaching Robots to Manipulate

Top primitive:
the object is approached from the top with palm down parallel to the table. Object center is approximatively at the level of middle phalanx. When contact is established all fingers are simultaneously closed, achieving a firm power-like grasp.

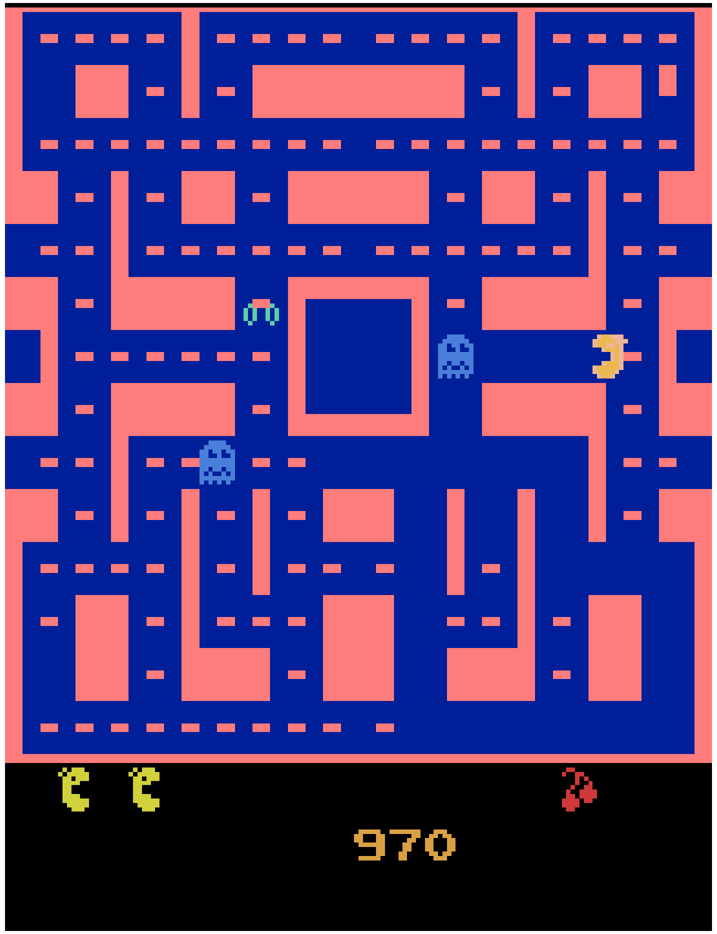


Generative Adversarial Networks

At the roots of the generative deep learning wave

Early bedroom uses…

…and Psychedelic Pacman



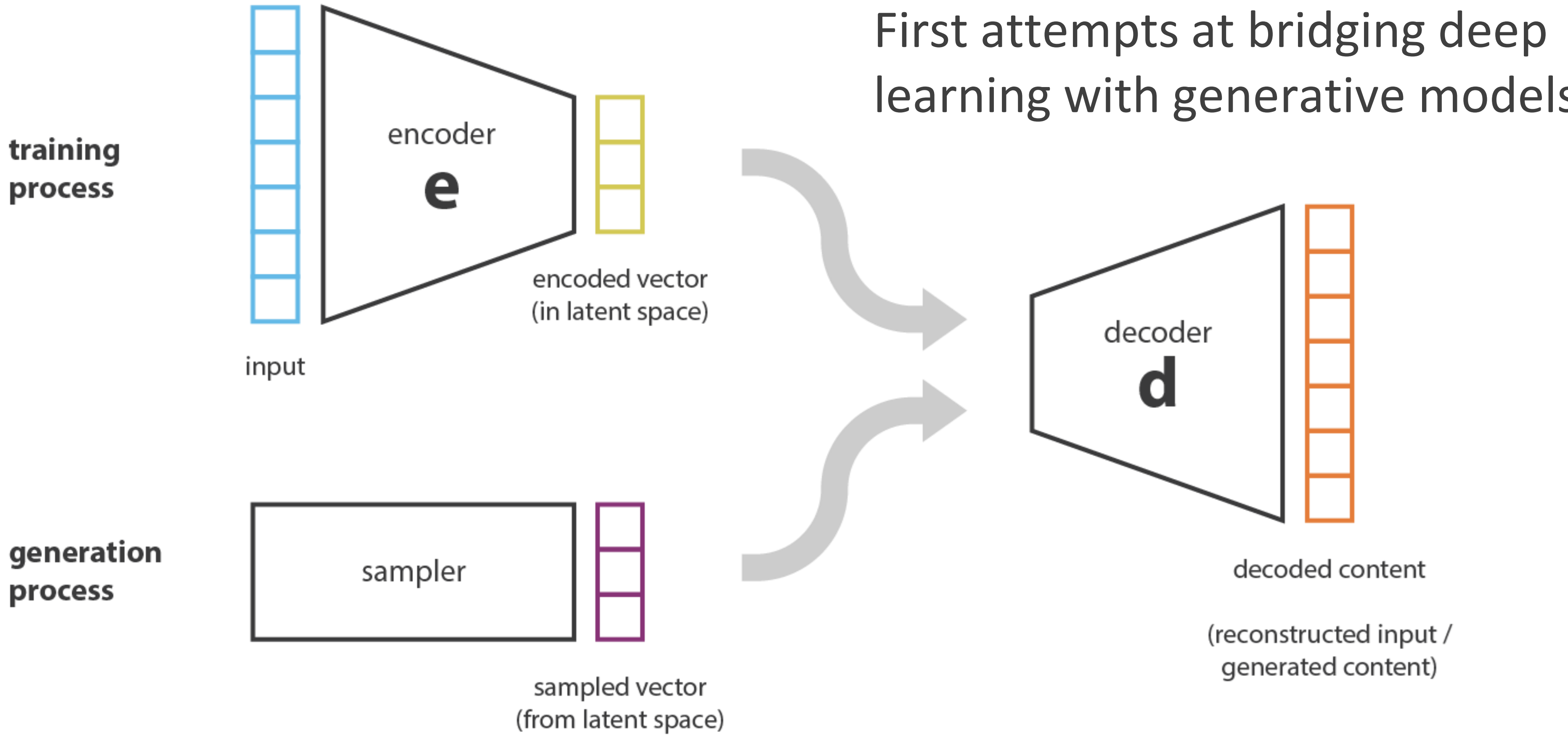
Starting to get better at face generation

Face Swapping in Back to the Future

Or in any other clip you like

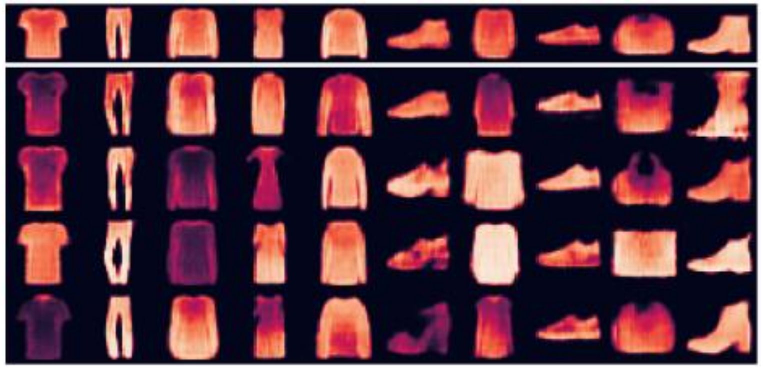
Variational Deep Learning

First attempts at bridging deep learning with generative models



Learning Entities and Relations from Images

Numbers



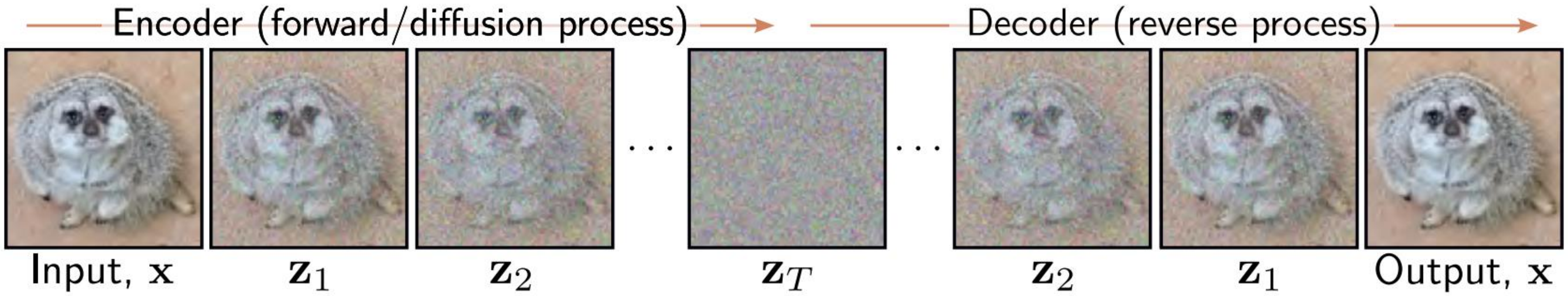
Clothing

Addition

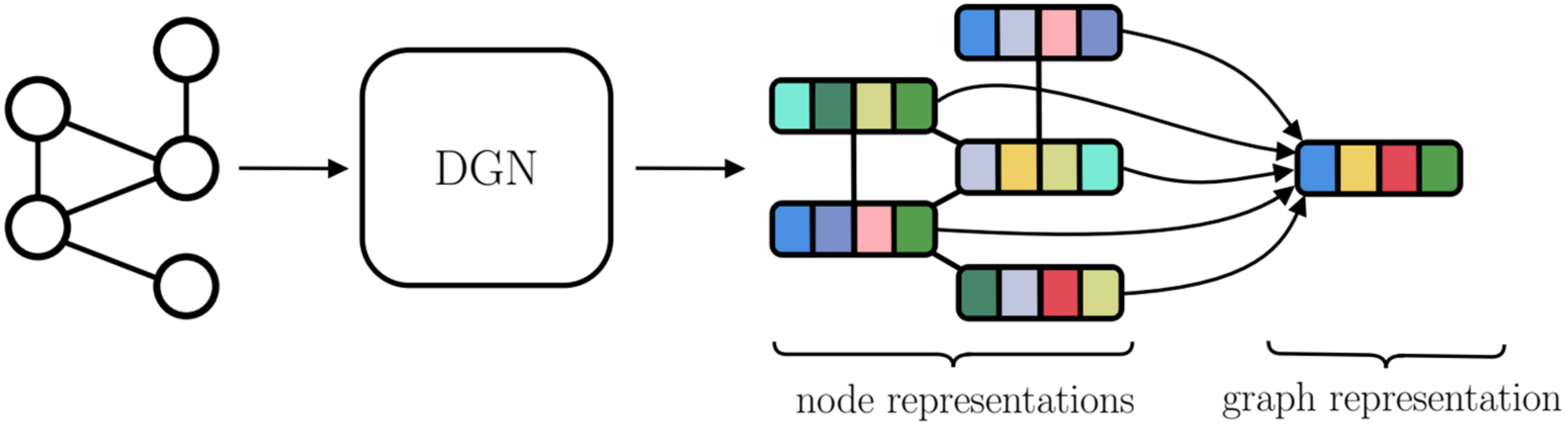
Multiplication

But nowadays nobody cares because we have…

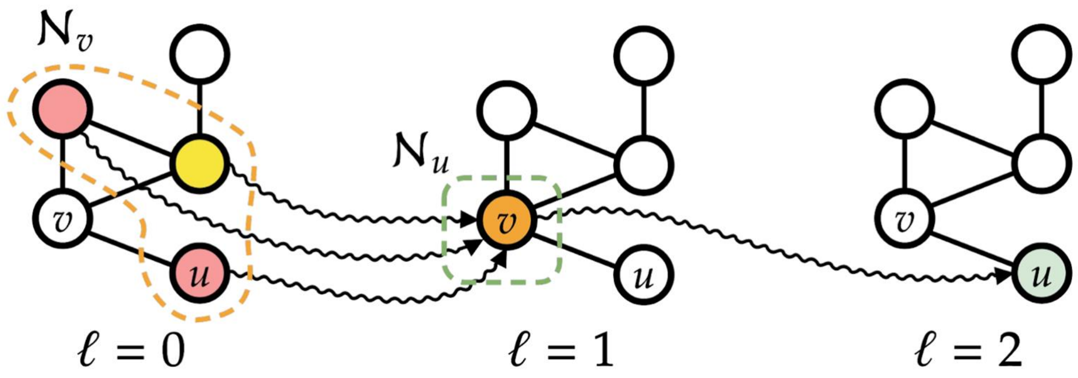
…diffusion models!



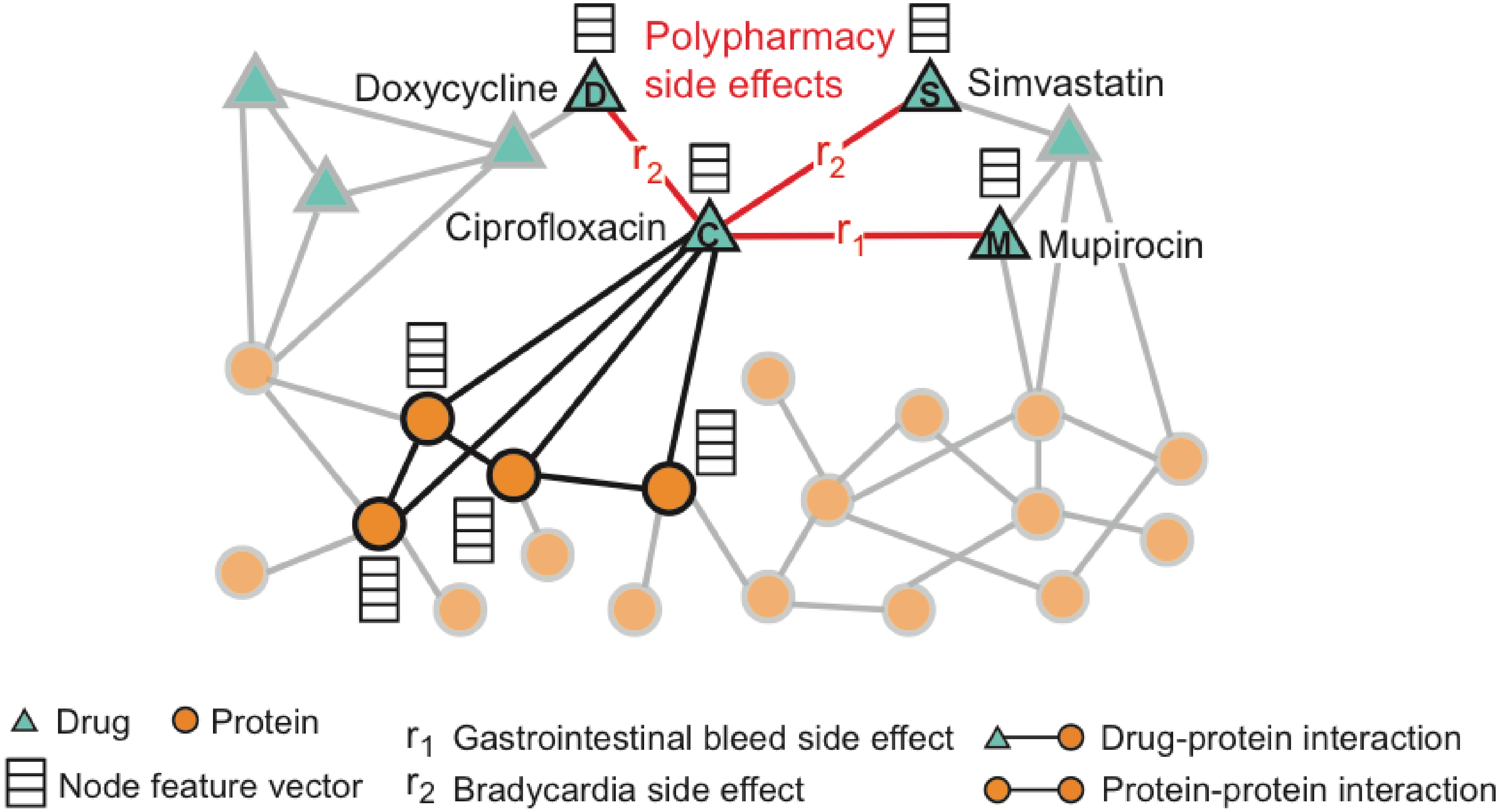
Graph Neural Networks



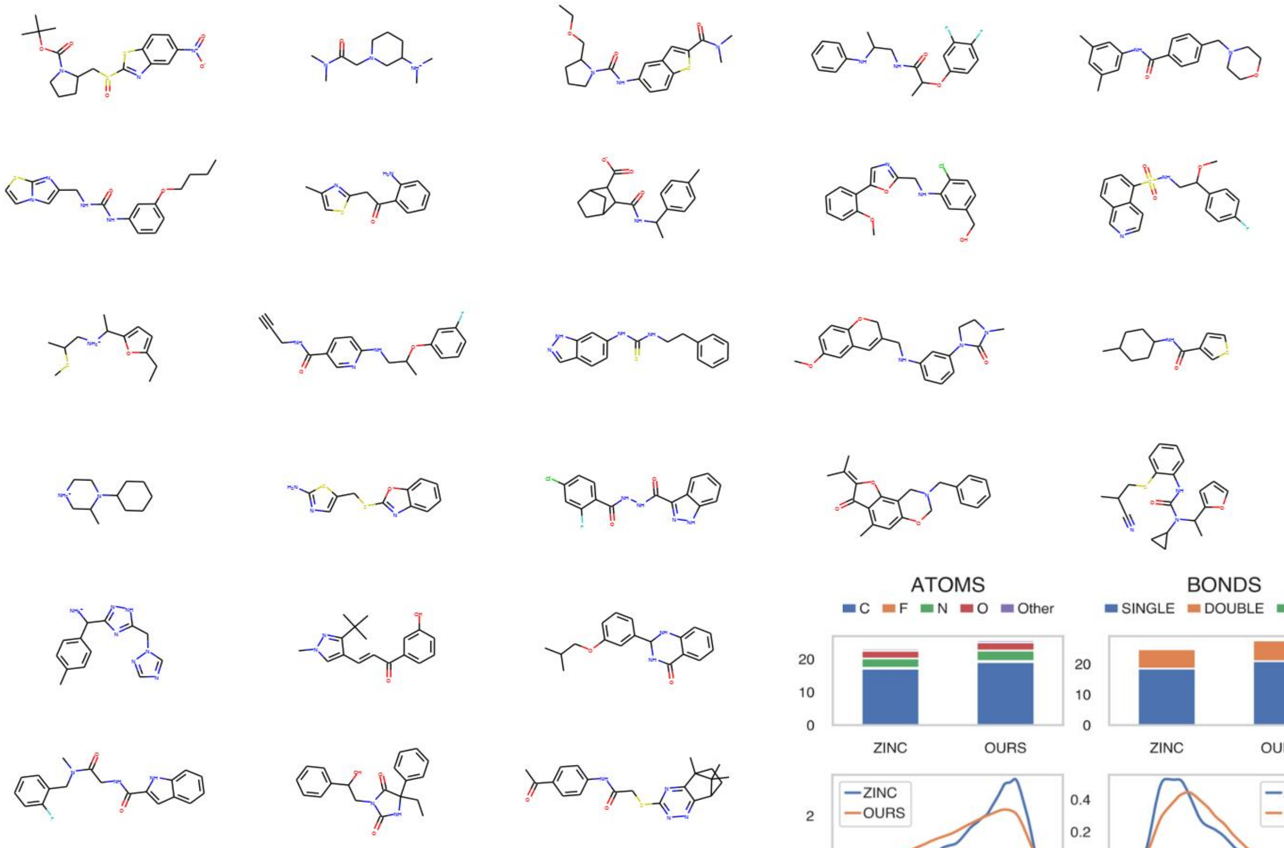
An exploding field
in Deep Learning

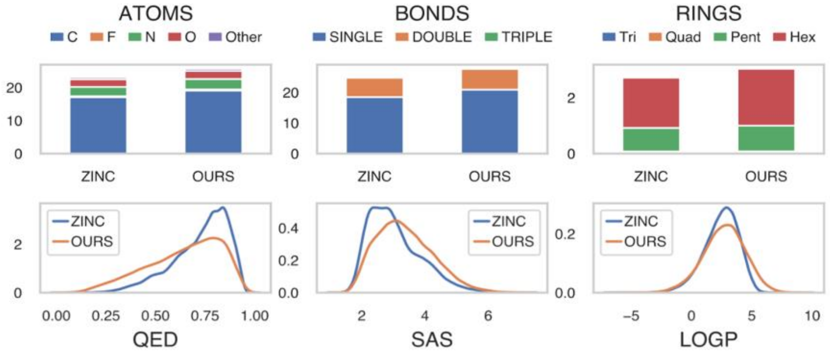


Drug Repurposing

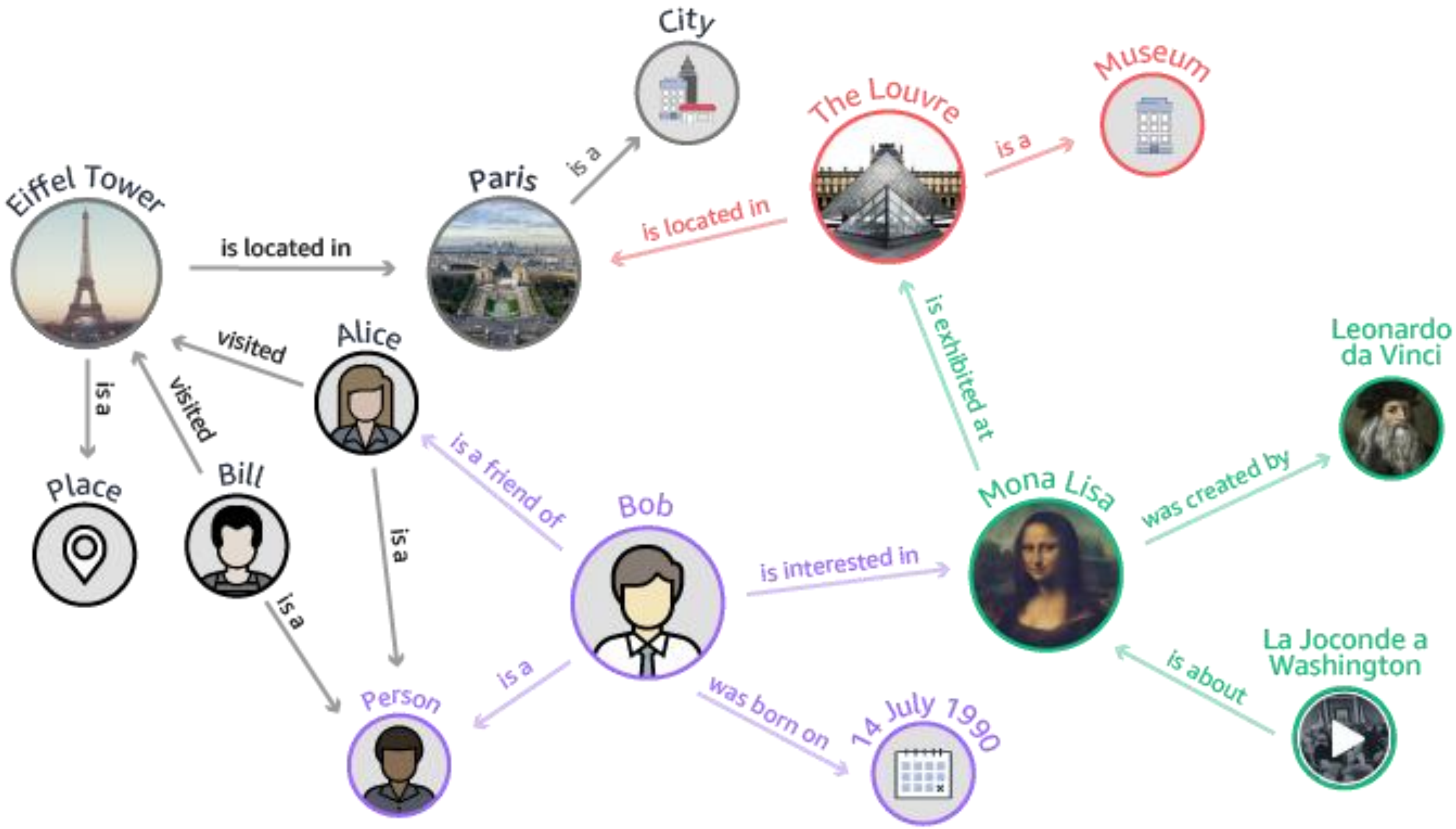


Generating Molecules

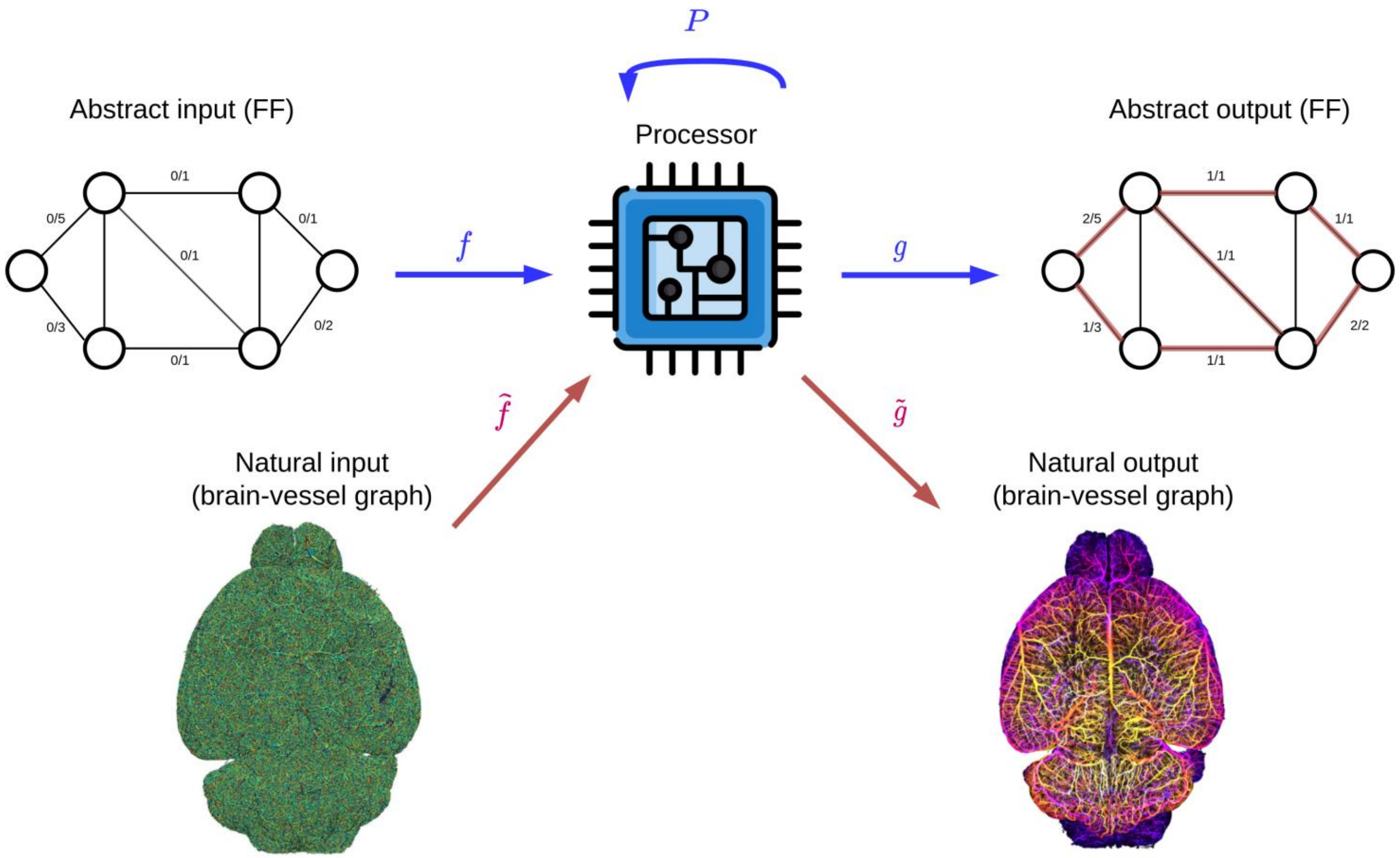




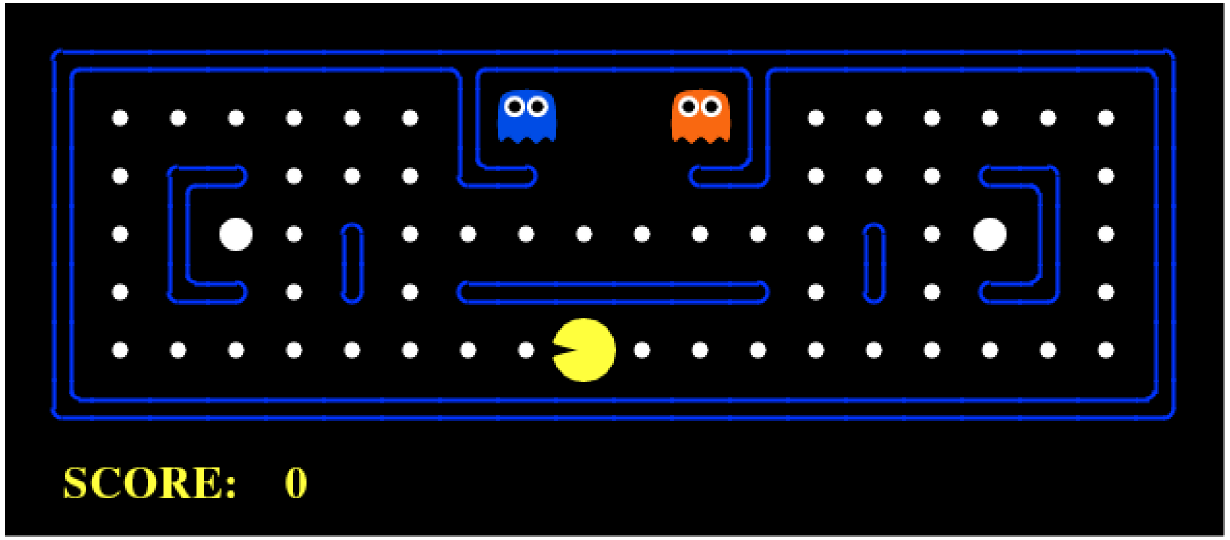
Incorporating Knowledge Graphs



Neural Algorithmic Reasoning



Learning intelligent agents



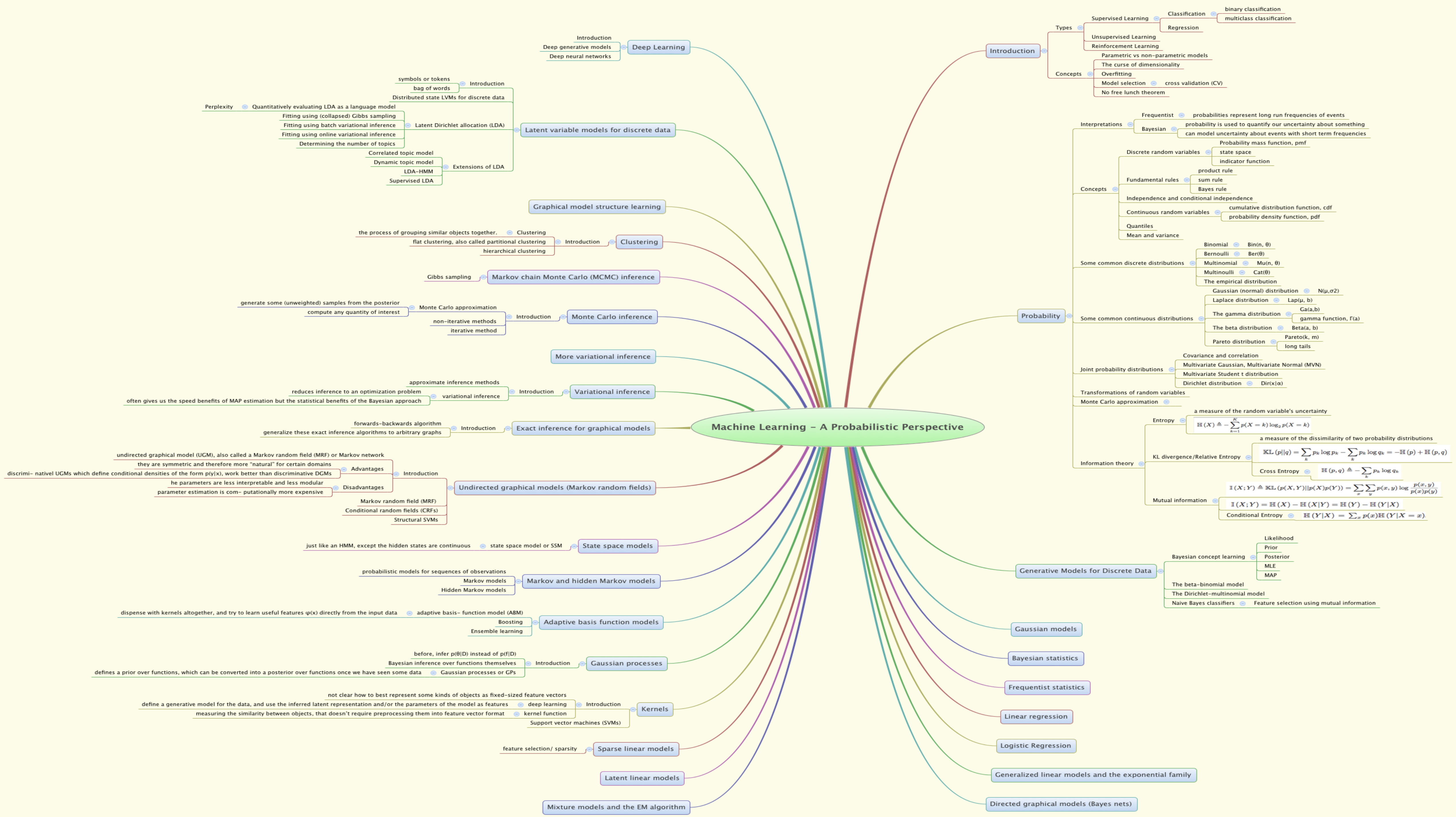
The Course Philosophy

- Start from traditional PR approaches
  - Introduce problems and tasks
  - Learn some useful techniques
- Learn how old-school stuff has been reused in a modern way
- Understand how probability is fundamental to machine-deep-reinforcement learning
- Connect the dots between traditional PR, generative and deep learning

A practical approach with code complementing theory when possible

On probabilities and pains of the sort

- From student anonymous advices (some 2 years ago)
  - Too much attention on probabilistic models, which are not state of the art….
  - The course should only briefly mention probabilistic models and focus on state-of-the-art models…



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 Languages

Reference language for the course is Python (but some Matlab might pop-up)

- 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 signal processing, image processing and machine 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 lecture 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 alternative way (for working students, those who fail or don't like the other way)

- **Final Project** - A written report on a topic of interest for the course, a software implementing a PR application, ….
- **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

- Delivery of a notebook/colab or a very short slide deck (e.g. 10 slides) on
  - A quick and dirty (but working) implementation of a simple pattern recognition algorithm
  - A report concerning the experience of installing and running a demo application realized using available libraries
  - A summary of a recent research paper on topics/models related to the course content

- 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

Final Project (I)

- 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
  - Report (10 pages, for survey type) or code (for SW type) and presentation (for all) delivery: by the standard exam date (appello) (strict)

Final Project (II)

Possible project types

- **Survey -** Read at least five relevant and distinct papers on a topic, prepare a presentation and write a report: not a simple summary, rather try to find connections between the works and highlight interesting open problems
- **Software -** Develop a tested and commented software implementing a non-trivial learning model and/or a pattern recognition application relevant for the course. Prepare a presentation describing code and its validation.

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

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 vote is given by the oral examination grade

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

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

UNIVERSITÀ DI PISA

Upcoming…

Introduction to Pattern Recognition

An introduction to the fundamental PR problems in signal and image processing and a summary of the old-school techniques to address them.

Topics

- Pattern recognition in time/spatial and spectral domain
- Timeseries and image analysis
- Convolution and correlation operators
- Visual feature descriptors
- Visual feature detectors
- Image segmentation

Next Lecture

Introduction to Signal Processing

- Timeseries
- Convolution and correlation
- Spectral analysis

Changes to next week schedule!

- ~~Tuesday 27 February~~ – **Cancelled**
- ~~Wednesday 28 February~~ - **Cancelled**
- ~~Thursday 29 February~~ - **Cancelled**

Extra (recovery) lectures

- Friday 01 March – Room L1 h14-16 (confirmation pending)
- Friday 08 March – Room L1 h14-16
- Friday 15 March – Room L1 h14-16

Onboarding

Remember to register on the course Moodle

https://elearning.di.unipi.it/course/view.php?id=278

Within the end of this week please signup on this shared spreadsheet

- Your email address for the course mailing list
- Name and curriculum/course
- Note the different sheet for M.Sc./Ph.D.

When you send me an email include tag [ISPR] (or may end up in thrash)

Questions?

UNIVERSITÀ DI PISA