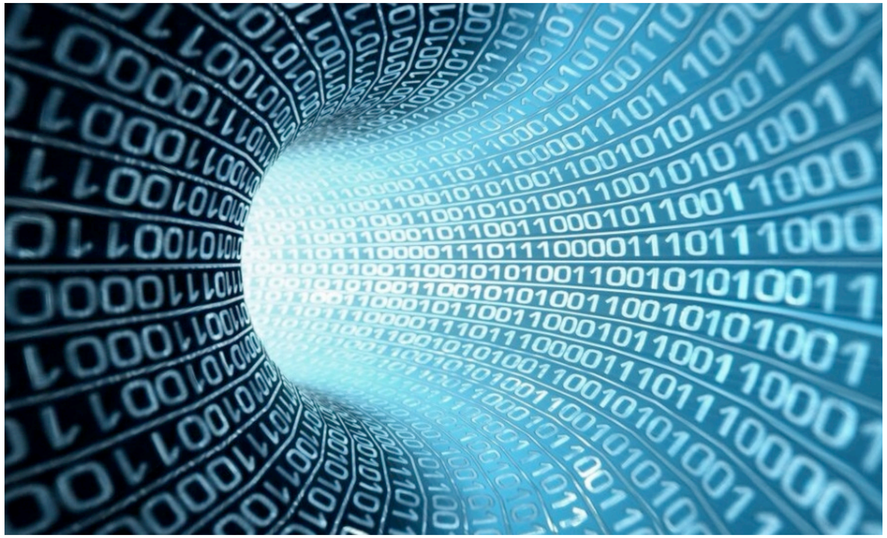
DD2434 Machine Learning, Advanced Course

Lecture 1: Introduction

Hedvig Kjellström
hedvig@kth.se
https://www.kth.se/social/course/DD2434/

Big Data

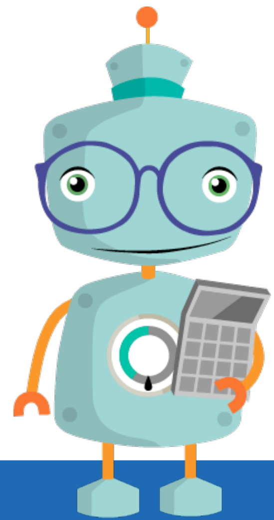


Making sense of signals (RGB-D video): Hand Tracking from MSR Cambridge

https://www.youtube.com/watch?v=A-xXrMpOHyc

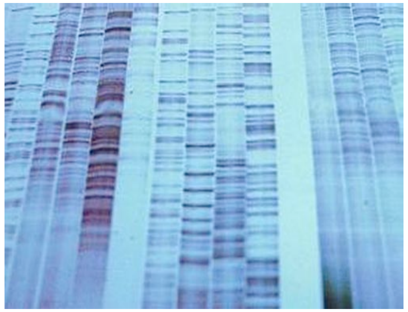
Predicting future events knowing the history: Botten Ada from Linköping U

http://bottenada.se



Learning to see suddle patterns in huge amounts of data: Cancer Therapy based on DNA Sequencing from IBM

https://www.youtube.com/watch?v=0M1DMdc1mQ0

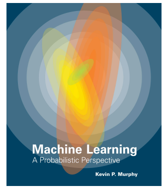


Today

Check the homepage at least 2 times / week!
Or set it to send you emails!

Course preliminaries

- All info at https://www.kth.se/social/course/DD2434/
- **Ask questions through the News forum!**
- Buy the book by Kevin Murphy:



The three teachers

- Jens Lagergren
- Carl Henrik Ek
- Hedvig Kjellström

Introduction to Machine Learning

- Murphy Chapter 1

Course Preliminaries

Learning outcomes

Upon completion of the course, the student should be able to

1. explain, derive, and implement a number of models for supervised, unsupervised learning,
2. explain how various models and algorithms relate to one another,
3. describe the strengths and weaknesses of various models and algorithms,
4. select an appropriate model or approach for a new machine learning task.

Course organization

Assignments, detailed schedule with reading, etc, on the homepage

Nov 25
1: Graphical models
Jens Lagergren

Dec 3
2: Representations
Carl Henrik Ek

Dec 15
3: Applied ML
Hedvig Kjellström

Dec 18
Project

The Three Teachers

Jens Lagergren

Professor of Computer Science
at KTH / Science for Life Laboratory
Research area: Bio-informatics

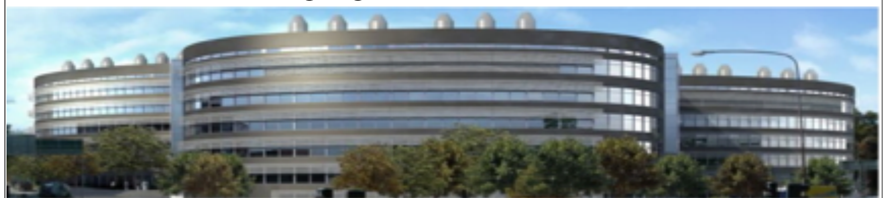
Responsible for
- Lectures 2-5
- Practicals 1-2
- Assignment 1

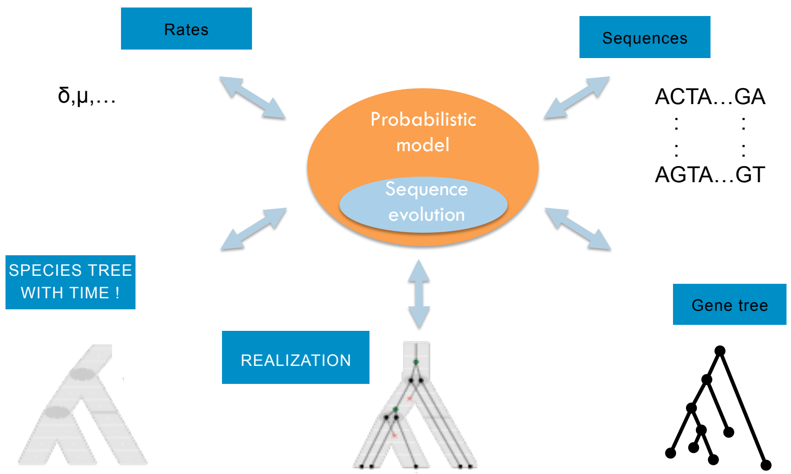
Royal Institute of Technology

Computational Biology

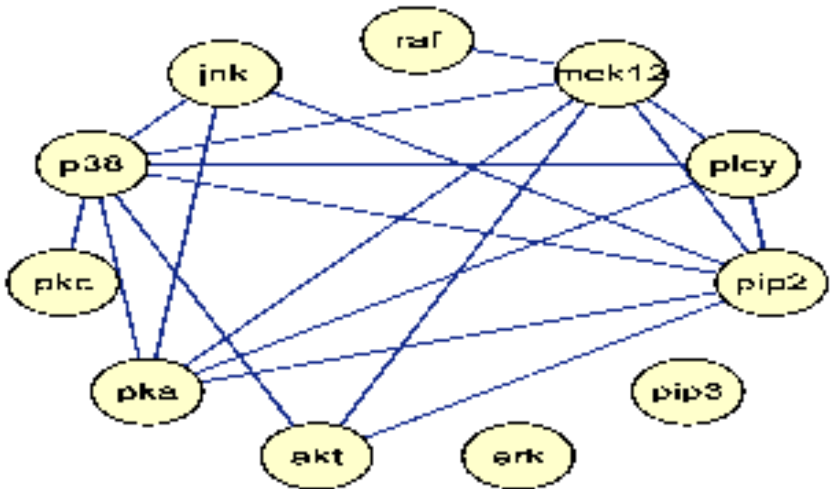
Machine Learning – a main tool

Jens Lagergren

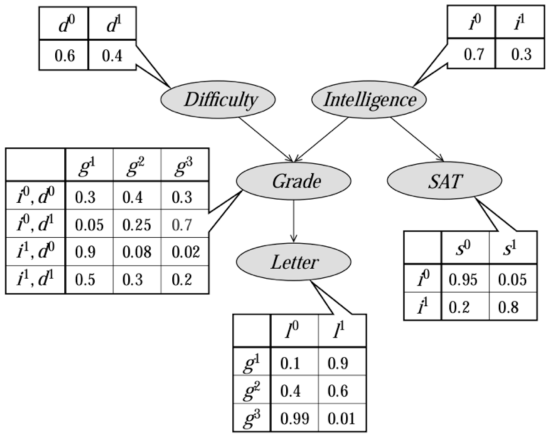




DLRS (Delirious)



DISCOVERING GRAPH STRUCTURE



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

DIRECTED GRAPHICAL MODELS

DD2434 - Advanced Machine Learning

Lecture 6-8, Assignment 2, Practical 3-4

Carl Henrik Ek
{chek}@csc.kth.se

Royal Institute of Technology

November 3, 2014

My Research

- Representation Learning
  - multi-view representations
  - correspondence/alignment learning
- Non-parametric methods
  - Gaussian Processes
- Structural representations
- Applications
  - Animal welfare
  - Motion modelling
  - Computational Biology

Lectures

Theme

"How can I incorporate my knowledge/belief with observations such that data reduces my uncertainty?"

Lectures

6 Basic modelling
- Likelihood, Prior & Posterior
- Kernels

7 Non-parametric modelling
- function uncertainty

8 Representation Learning
- Pattern discovery

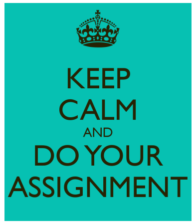
9 Hierarchical modelling
- layered structures

Assignment

- Regression

$$f : \mathbf{Y} \rightarrow \mathbf{X}$$

- Three parts
  1. Building models
  2. Learning models
  3. Evaluating models

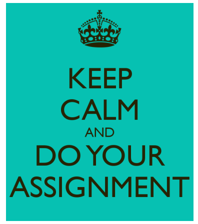


Assignment

- Regression

$$f : \mathbf{Y} \rightarrow \mathbf{X}$$

- Three parts
  1. Building models
  2. Learning models
  3. Evaluating models



MSc Thesis Work

- Start January/February
- Related to my research
- Associated with CVAP
- More on lecture 9

Practicals

My best friend the Gaussian

- Example of Conjugate priors
- Multiplication
- Marginalisation
- Derivatives of Matrices

Practicals

Learning: Tools of the trade
- How to fit models to data
- Beyond ML & MAP
- Variational Approximation

Hedvig Kjellström

Associate Professor of Computer Science at CSC / CVAP

Research area: Robotics and Computer Vision

Responsible for
- Entire course
- Lectures 1, 10-12
- Practical 5
- Assignment 3

Hedvig Kjellström: My research

Machine learning applied to Robotics and Computer Vision: Automatic perception of human activity in video

Object affordances, object-action complexes
"automatic understanding of how objects are used in human activities what happens to them during the activity"

Human non-verbal communication
"automatic understanding and modeling of non-verbal signals – face expressions, body motion - both conscious and unconscious"

Multi-modality and context in activity recognition
"using several modalities – vision, sound, touch etc – to better understand human activity"

See my webpage for master project proposals!

Hedvig Kjellström: My part of the course

Course block 3: Applied Machine Learning

Topic models
- Just one out of many methods, but important
- Chosen to complement the methods covered in DD2427 Image Based Recognition and Classification, DD2431 Machine Learning, EN2202 Pattern Recognition

Practical Machine Learning
- What happens to the performance when data is noisy and incomplete?

Introduction to Machine Learning

Uncertainty

Basic philosophy:

Data (observations) noisy and incomplete i.e. **uncertain**
Decision making (prediction, classification, detection, estimation) under **uncertainty**

Uncertainty is best modeled with **probability theory**

Common division:
Supervised / **Unsupervised**

Supervised/Predictive Learning

Data (**training set**): $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$

features/attributes   response variable

Task: Learn mapping $\mathbf{x} \rightarrow y$

5 min: Discuss with your neighbor

Give at least three examples of supervised learning problems

What is $\mathbf{x}$ and y in each problem?

What does the mapping look like (linear/non-linear, one-to-one/many-to-many, smooth/noisy)?

Supervised/Predictive Learning

unknown true function

Functional approximation: $y = f(\mathbf{x})$

Use $\mathcal{D}$ to learn an approximative function $\hat{y} = \hat{f}(\mathbf{x})$

Classification: $y \in \{1, \ldots, C\}$ is discrete and finite

Probabilistic formulation: Model

$p(y = 1|\mathbf{x}, \mathcal{D})$, $p(y = 2|\mathbf{x}, \mathcal{D})$, etc…

Best $y \equiv$ most probable y:

$$\hat{y} = \hat{f}(\mathbf{x}) = \arg\max_{c=1}^{C} p(y = c|\mathbf{x}, \mathcal{D})$$

Unsupervised/Descriptive Learning

Data (**training set**): $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^{N}$

Task: discover patterns in $\mathcal{D}$

Under-specified problem – what patterns? How measure error?

5 min: Discuss with your neighbor

Give at least three examples of unsupervised learning problems

What is $\mathbf{x}$ in each problem?

What kind of patterns are found?

What is the purpose?

Unsupervised/Descriptive Learning

Probabilistic formulation: Density estimation

Models of the form $p(\mathbf{x}_i|\theta)$

Use $\mathcal{D}$ to maximize the probability $p(\mathbf{x}_i|\theta)$ of seeing each $\mathbf{x}_i$ given the model θ

New obstacles: Multivariate distributions

Unsupervised learning is more similar to how humans and animals learn!

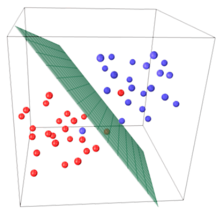
Practical advantage: No labeling of data required!

Basic concept: Parametric vs Non-Parametric

Models $p(\mathbf{x})$ and $p(y|\mathbf{x})$

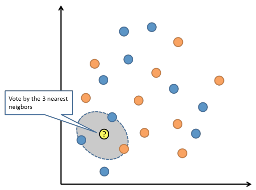
Parametric: Number of parameters constant with more data

E.g., linear classifier

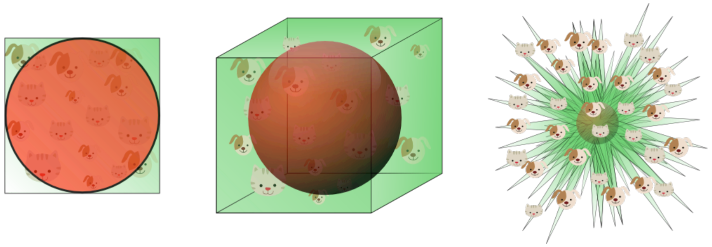


Non-parametric: Number of parameters grows with more data

E.g., kNN classifier



Basic concept: Curse of Dimensionality



2D cube/sphere $= \frac{\pi}{2^2}$

3D cube/sphere $= \frac{4\pi}{2^3 * 3}$

8D cube/sphere $= \frac{\pi^4}{2^8 * 24}$

Adressed by using parametric models (fewer parameters – more robust)

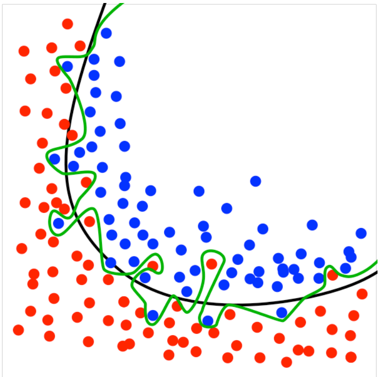
Basic concept: Overfitting

Model fits training data perfectly but not novel data

Reasons: Too little data, to high dimension, too flexible model

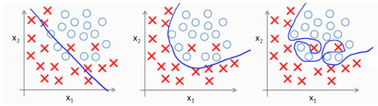
5 min: Discuss with your neighbor

How can you test if your classifier is overfitting the training data?



Basic concept: Model Selection

Overfitting and **underfitting**

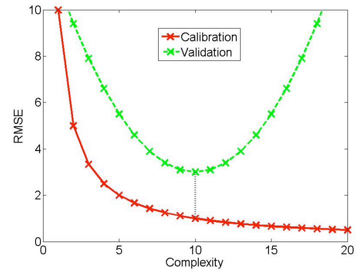


More complex model always have lower training data error

Solution from last slide:

Divide data into training set and **validation set**

Evaluate each model, each parameter setting with the validation set



Basic concept: No Free Lunch Theorem

Do not believe the preachers…

There is no universally best model! All models contain assumptions that work well in one domain but not in another.

What is next?

Check the homepage at least 2 times / week!
Or set it to send you emails!

We use the homepage a lot: links to video lectures, readings for lectures, lecture slides, questions answered through the News forum

https://www.kth.se/social/course/DD2434/

Next on the schedule

Wed 5 Nov 13:15-15:00 V32

Lecture 2: Graphical Models

Jens Lagergren

Readings: Murphy Chapter 10 (except 10.2.4, 10.2.5, 10.4)

Assignment 1 published today, deadline November 25