UPPSALA
UNIVERSITET

Statistical Machine Learning

Lecture 1 – Introduction

UPPSALA
UNIVERSITET

Andreas Lindholm
Division of Systems and Control
Department of Information Technology
Uppsala University

andreas.lindholm@it.uu.se
www.it.uu.se/katalog/andsv164

UPPSALA UNIVERSITET

What is the course about?

Machine learning

"Machine learning is about learning, reasoning and acting based on data."

"It is one of today's most rapidly growing technical fields, lying at the intersection of computer science and statistics, and at the core of artificial intelligence and data science."

Ghahramani, Z. **Probabilistic machine learning and artificial intelligence**. *Nature* 521:452-459, 2015.

Jordan, M. I. and Mitchell, T. M. **Machine Learning: Trends, perspectives and prospects**. *Science*, 349(6245):255-260, 2015.

UPPSALA UNIVERSITET

This course

What is this course about? Supervised machine learning

In one sentence:

Methods for automatically learning (training, estimating, ...)
a model for the relationship between

- the **input** $\mathbf{x}$, and the
- the **output** y

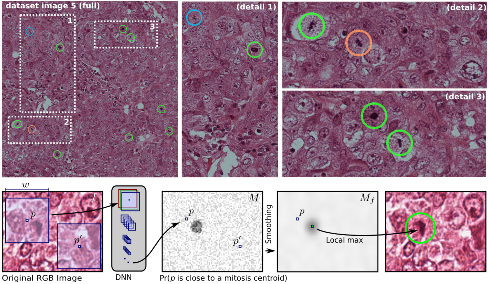
from observed **training data**

$$\mathcal{T} \stackrel{\text{def}}{=} \{(y_1, \mathbf{x}_1),\ (y_2, \mathbf{x}_2),\ \ldots,\ (y_n, \mathbf{x}_n)\}.$$

Seems dull...?! Can this be useful?

ex) Cancer diagnosis

Systems for detecting cell divisions (mitosis) in histology images can be used to improve (or automate) cancer diagnosis.

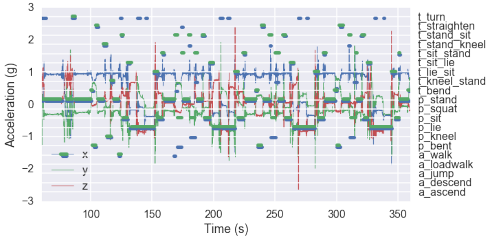


- Learn a model with
  input $\mathbf{x}$ = RBG histology image (pixel values)
  output y = number and locations (in the image) of mitosis detections
- Training data: Histology images labeled by experts.
- Uses Deep Learning to model f (Lectures 8–9)

D. C. Cireşan, A. Giusti, L. M. Gambardella and J. Schmidhuber. **Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks**. In *Medical Image Computing and Computer Assisted Intervention*, 411-418, 2013.

ex) Safe aging

SPHERE is a large UK-based research project for helping elderly people to live safely at home.



Machine learning problem: Non-intrusive activity recognition.

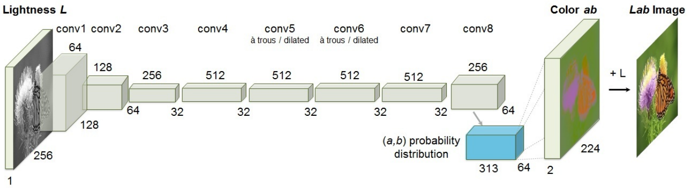
- Learn a model with

  input $\mathbf{x} =$ accelerometer, IR & positioning sensor data
  output $y =$ "`sitting`", "`walking`", "`ascending stairs`", etc.

- Boosting (Lecture 7) among the most successful methods.

N. Twomey, T. Diethe, M. Kull, H. Song, M. Camplani, S. Hannuna, X. Fafoutis, N. Zhu, P. Woznowski, P. Flach, and I. Craddock. **The SPHERE Challenge: Activity Recognition with Multimodal Sensor Data**. *arXiv:1603.00797*, 2016.

ex) Colorization (I/II)



Task: Colorize gray-scale photos.

- Learn a model with

  input $\mathbf{x} =$ gray-scale pixel values.
  output $y =$ color pixel values (Lab).

- Typical task for deep learning (Lectures 8–9).

R. Zhang, P. Isola, and A. A. Efros. **Colorful Image Colorization**. *ECCV*, 2016.

ex) Colorization (II/II)

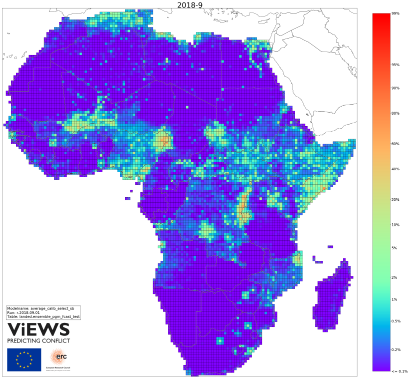
Model applied to legacy grayscale photos.

ex) Predicting conflicts

- Predicting the risk of violent conflicts across Africa
- Learn a model with

  input $\mathbf{x}$ = conflict history, protests, population, economic indicators, geography, ...
  output y = risk of violent conflict

- Random forests (Lecture 6) used



https://www.pcr.uu.se/research/views

ex) Autonomous driving

- Recognizing objects in a street view
- Learn a model with

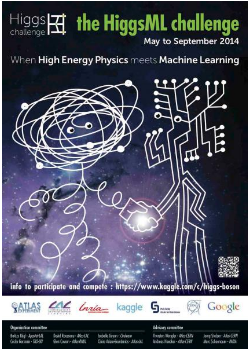
  input $\mathbf{x}$ = the RGB values of the image
  output y = the location of pedestrians in the image

- Deep learning (Lectures 8–9)

ex) Higgs Machine Learning Challenge

HiggsML: Crowd-sourcing initiative by CERN (hosted at Kaggle)

- Separate $H \to \tau\tau$ from background noise.
- Lean a model with
  - input $\mathbf{x} =$ 30-dimensional vector of "features" recorded during the experiment.
  - output $y =$ "`signal`" or "`background`"
- Bagging (Lecture 6), Boosting (Lecture 7) and Neural Networks (Lectures 8–9) among the winning methods.



C. Adam-Bourdarios, G. Cowan, C. Germain, I. Guyon, B. Kégl and D. Rousseau. **The Higgs boson machine learning challenge**. *NIPS 2014 Workshop on High-energy Physics and Machine Learning*, 2014.

Statistical machine learning

Why the word "statistical" in the course title?

- Probability theory is used to define the models.
- Statistical tools are used to learn the models from training data.

Allows us to reason about the ***uncertainties*** in the data, models, predictions, etc.!

Quantitative and qualitative variables

Both input variables ($\mathbf{x}$) and output variables (y)
can be either **quantitative** or **qualitative**.

- **Quantitative** variables take on numerical values (real numbers, integer values, . . .).
- **Qualitative** variables take on values in one of K distinct classes, e.g. “true or false”, “disease type A, B or C”.

We will mostly use integer coding (like $\{1, 2, 3, 4\}$ or $\{0, 1\}$), but the coding (or labeling) of qualitative variables is arbitrary and unordered.

Regression vs. classification

We will distinguish between two types of problems:
regression and **classification**

Regression is when the output y is quantitative, e.g.

- Climate models ($y =$ "increase in global temperature")
- Economic models ($y =$ "change in GDP")

Classification is when the output y is qualitative, e.g.

- Spam filters ($y \in \{\texttt{spam}, \texttt{good email}\}$)
- Diagnosis systems ($y \in \{\texttt{ALL}, \texttt{AML}, \texttt{CLL}, \texttt{CML}, \texttt{no leukemia}\}$)
- Fingerprint verification ($y \in \{\texttt{match}, \texttt{no match}\}$)

Aim of the course

What will we learn in this course?

- Use various methods for solving regression and classification problems, ranging from fundamental (linear regression) to state-of-the-art (deep learning, boosting, . . .)
- Identify and apply suitable methods to a given problem
- Evaluate the performance of a method and rationally choose between different competing models and methods
- Work with real data, reason about data representations and how data is used in realistic machine learning applications

Course information

UPPSALA UNIVERSITET

Lecture outline

1. Introduction
2. Linear regression, regularization
- Introduction to Python & scikit-learn
3. Classification, logistic regression
4. Classification, LDA, QDA, k-NN
5. Bias-variance trade-off, cross validation
6. Tree-based methods, bagging
7. Boosting
8. Deep learning I
9. Deep learning II
10. Summary and guest lecture

”Warm-up videos” for each lecture linked from the home page.
Watch before coming to the lecture (the night before, or so).

UPPSALA UNIVERSITET

Course elements

- 10 lectures + 1 introductory lecture to Python
- 10 problem solving sessions
- 1 mini project (3-4 students, written report)
- 1 computer lab (4h, no report)
- Complete course information (including lecture slides) is available from the course home page:

  `www.it.uu.se/edu/course/homepage/sml`

Problem solving sessions

10 problem solving sessions:

- Solve problems, **discuss and ask questions**! ("räknestuga")
- 5 pen-and-paper sessions
- 5 computer-bases sessions (using Python or R)
- Feel free to use your own laptops – Python and freely available
- Exercises available via homepage or the student portal

The computer-based sessions are scheduled in 1 computer room + 1 normal class room. The latter is intended for students who choose to work on their own laptops.

A great opportunity to discuss and ask questions!

Examination

Mini project:

- Solved in groups of 3 or 4 students (sign up before January 30)
- Written report (deadline: February 22)
- Peer-review: read and review another group's report (anonymously)
- Material most relevant for the mini project presented at lectures 3–7, **but you can start working on the solution after lecture 3**
- Graded U/G, however we will hand out "gold stars" for projects of notable quality. **A gold star gives bonus points on the exam!**

Laboration:

- 4 h computer laboration, solved in groups of 2 students, graded U/G
- 4 sessions available – sign up for one of these
- **Solve the preparatory exercises before the lab session!**

Written exam:

- Written pen-and-paper exam, graded as U, 3, 4, or 5.
- You can bring one (1) page with your own notes
- Old exams are on the course home page

Course literature

Lecture notes ($\sim$ 130 pages), available via the course home page (chapter 6 not yet in place)

Additional reading suggestions: extensive list on the home page

The lecture notes will eventually be turned into a book. If you provide us with many useful comments, you will get a free copy of the book!

Teachers

Teachers involved in the course (in approximate order of appearance):

						
Andreas Lindholm	Thomas Schön	Fredrik Lindsten	Niklas Wahlström	Carl Andersson	David Widmann	Carmen Lee
Room: 2340	Room: 2209	Room: 2335	Room: 2319	Room: 2353	Room: 2303	

All room numbers are at ITC Polacksbacken.

You can reach us by email: `<firstname.lastname>@it.uu.se`.

Statistical Machine Learning

Probability theory refresher

- If z is a random variable with PDF $p(z)$, then the **expected value** or **mean** of z is given by

$$\mathbb{E}[z] = \int z p(z) \mathrm{d}z.$$

More generally, $\mathbb{E}[g(z)] = \int g(z)p(z)\mathrm{d}z$.

- Let $\mu = \mathbb{E}[z]$. The **variance** of z is defined as

$$\mathrm{Var}[z] = \mathbb{E}[(z-\mu)^2] = \mathbb{E}[z^2] - \mu^2.$$

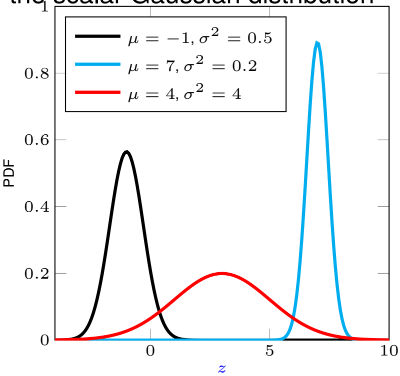
- If z_1 and z_2 are **independent** random variables, then $p(z_1, z_2) = p(z_1)p(z_2)$ and $\mathbb{E}[z_1 z_2] = \mathbb{E}[z_1]\mathbb{E}[z_2]$.

Gaussian (Normal) distribution

Probability density function (PDF) for the scalar Gaussian distribution

$$\mathcal{N}(z \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

- μ is the mean (expected value of the distribution)
- σ is the standard deviation
- σ^2 is the variance



$z \sim \mathcal{N}(\mu, \sigma^2)$ means that z is a Gaussian random variable with mean μ and variance σ^2. $\sim$ reads "distributed according to".

Workflow for machine learning

1. Formulate the problem as a machine learning problem
2. Collect training data
3. Pre-process the data
4. Choose which model to use
5. **Learn**/**train**/estimate/fit/…the model from training data
6. Feed $x_\star$ into the model to make a **prediction** $\widehat{y}_\star$
7. Evaluate the prediction, and thereby the usefulness of the model

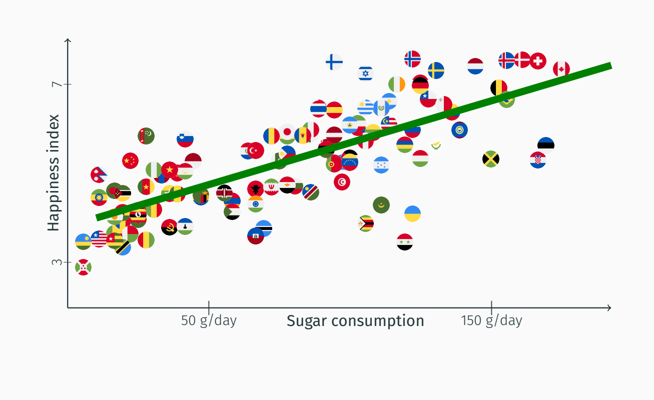
Training a model

When training a model on training data

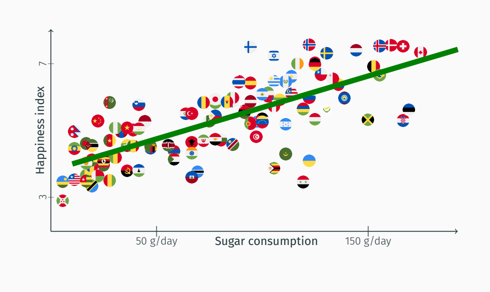
$$\mathcal{T} \stackrel{\text{def}}{=} \{(y_i, \mathbf{x}_i)\}_{i=1}^n,$$

the model is somehow adopted to the training data.

The most basic learning is fitting a straight line to data:



Training a model



This is **linear regression** (next lecture). A good entrance into the world of supervised machine learning.

We will often understand training from a statistical **maximum likelihood** perspective – adapting the model such that the data is as likely as possible to have been observed.

Randomness of the learned model

We use training data $\mathcal{T} = \{(x_i, y_i)\}_{i=1}^n$ to **learn a model**

If the training data $\mathcal{T} = \{(x_i, y_i)\}_{i=1}^n$ used to learn the model is random, then so is the learned model!

We will use statistical tools and probability theory to reason about the properties of the learned model: Bias-variance trade off in lecture 5.

UPPSALA UNIVERSITET

A few concepts to summarize lecture 1

Machine Learning: Deals with learning, reasoning and acting based on data.

Regression: Learning problem where the *output* is quantitative.

Classification: Learning problem where the *output* is qualitative.

Training data: The dataset that is used to learn a model from data. (The model should not be evaluated on this dataset.)

Maximum likelihood: Learning objective based on probability theory.