Computational Intelligence Laboratory

Lecture 0

Overview & Organization

Thomas Hofmann

TAs: **Aurelien Lucchi**, Gary Becigneul, Andrew Bian, Hadi Daneshmand, Octavian Ganea, Paulina Grnarova, Matthias Hüser, Yannic Kilcher, Francesco Locatello, Xinrui Lyu

ETH Zurich – `cil.inf.ethz.ch`

24 February 2017

Section 1

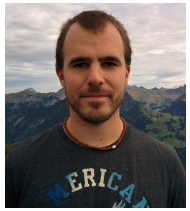
Organization

CIL Team

Week 8-11

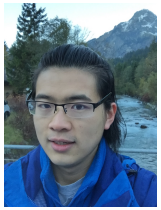
Octavian Ganea

Yannic Kilcher



Week 12-14

Andrew Bian



Francesco Locatello

Matthias Hüser

Week 15-18

Gary Becigneul

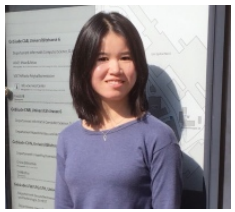
Paulina Grnarova

Aurelien Lucchi

Week 19-21

Hadi Daneshmand

Xinrui Lyu



Class Credit

- CIL developed and taught until 2014 by Joachim Buhmann and various research associates/assistants
- Ongoing changes, amendments, new lectures by Thomas Hofmann & Co. since 2015
- Projects by Aurelien Lucchi, Thomas Hofmann & Co.

Weekly Schedule: 2 + 2 + 1

- **Lecture**: Fri 10-12, ML D 28
- **Exercises**: Thu 15-17, CAB G51 **or** Thu 16-18, CAB G61 **or** Fri 15-17, CAB G61
  - all three exercise sessions are "identical"
  - first hour: pen-and-paper exercise, immediate discussions
  - second hour: group work on programming assignment
  - may need to "load balance", if distribution is skewed
- Voluntary presence time: Mo 11-12, CAB H53
  - TAs help completing programming assignments
- **Webpage:** `http://cil.inf.ethz.ch`
  - only accessible from the ETH network or VPN

Recording & Slides

- **Lecture slides** are (generally) posted before class
- Lectures are automatically **recorded** (voice + screen)
- Heavy use of **blackboard** (!) – not recorded (ETH limitation)
- Photos of board content posted after class – on a best effort basis

- *it is recommended that you come to class*

Digital Interactions

- Q&A platform – `piazza.com`
  - please sign up for an account
- ETH Edu App – used for classroom interaction
  - please download in app store
  - see ETH documentation
  - use nethz login – questions automatically pushed to you

Programming Project

- Joint work in groups of **two or three** students
- Solving problems by applying techniques learned in class
- Submitting solution to the CIL online evaluation & ranking system

Been here before? :-/

- You can re-submit previous year's project.
- You have to let us know that you are a "one person group"
- You have the choice to redo the project and join a group (as anyone else)

First Week

Coming week: exercise session on **Python** programming.

- Bring laptops with Python installed
- "Hands on" exercises in groups of 2-3 students

Reading material

- Linear algebra background – on the course website

Grading Criteria

Written Exam

- 120 minutes written exam, closed book (2x A4-pages of notes)
- Problems in the spirit of the pen-and-paper exercises

Semester Work

- Develop a novel solution for one of the application problems
- Compare with two baseline techniques already implemented in the weekly programming assignments
- Write up in the form of a short paper
- Competitive criteria: run-time, accuracy, ..
- Non-competitive criteria: paper review, creativity of solution, quality of implementation

Grading Formula

- **Final examination** – during examination period
- Project work in small teams – highly recommended
- Grading
  - exam < 3.5: $\Longrightarrow$ fail
  - exam ≥ 3.5:

$$\text{grade} = \frac{2}{3}\,\text{exam grade} + \frac{1}{3}\,\max\left\{\text{exam grade}, \underbrace{\text{project grade}}_{\text{potential uplift}}\right\}$$

Section 2

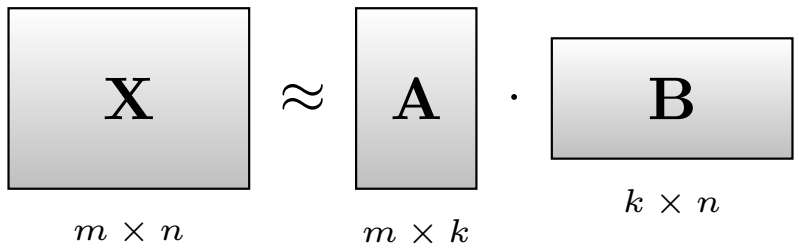
Goals & Content

Learning Goals

- Acquire and deepen fundamental concepts of machine learning
- Focus on: matrix factorization and representation learning
- Implement and compare techniques and models
- Goal:
  - understand and analyse methods
  - find best solutions to practical problems
- Expected background:
  - basic class on machine learning (e.g. LIS@ETHZ)
  - mathematics: linear algebra, vector analysis, prob & stats
  - programming: Python (easy to learn)

Matrix Factorization

Core problem: **matrix factorization** (MF) of a data matrix into factors, e.g.



Many important data analysis techniques can be written as MF problems by specifying ...

- **Type of data:** e.g. Boolean: $x_{ij} \in \{0, 1\}$ or non-negative: $x_{ij} \in \mathbb{R}_{\geq 0}$
- **Approximation:** e.g. exact: $\mathbf{X} = \mathbf{A} \cdot \mathbf{B}$ or minimal Frobenius norm: $\min_{\mathbf{A},\mathbf{B}} \|\mathbf{X} - \mathbf{A} \cdot \mathbf{B}\|_F$
- **Constraints on factors:** e.g. $\mathbf{A}$ has to be a basis or $\mathbf{B}$ must be sparse.

Learning Representations

More broadly: study methods of **learning data representations**

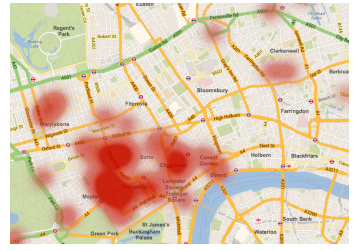
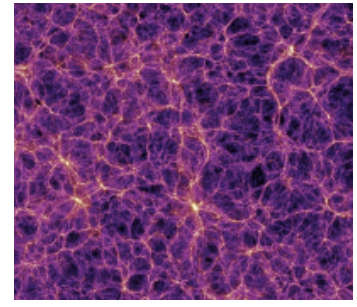
- Dimension reduction
- Latent variable models
- Data clustering
- Sparse coding
- Deep neural networks
- Generative models

Section 3

Projects

Project Overview

- Real world data sets and challenges – you pick one!

1	2	3	4
			
preferences recommender	texts & tweets sentiment	aerial imagery segmentation	cosmology galaxies

- Combine and extend techniques ⇒ novel solution
  - compare to baselines developed during the course
- Produce a write up of your findings ⇒ scientific short paper
  - emphasize experimental protocol, metrics, and reproducability

Project 1: Collaborative Filtering

Recommender system: present items of likely interest to a user

- Products: Amazon
- Movies: Netflix, IMDB
- Music: LastFM, Spotify

Collaborative filtering (CF) makes recommendations based on:

- (known) preference of a user towards other items
- collective preferences of other users

Project 1: Collaborative Filtering

Viewers were asked to rate some movies (items):

	Ben	Tom	John	Fred	Jack
Star Wars	?	?	1	?	4
WallE	5	?	3	4	?
Avatar	3	4	?	4	4
Trainspotting	?	1	5	?	?
Shrek	5	?	?	5	?
Ice Age	5	?	4	?	1

- Not all viewers rated all movies.
- We want to predict unrated user-movie pairs (**matrix completion**)
- Should we recommend Fred to watch "Ice Age"?

Project 2: Sentiment Classification

Automatic sentiment analysis to give a machine the ability to understand text and its polarity.

- **Data**: we provide a large set of training tweets.
- **Ground-truth**: each tweet is labeled as {negative, positive}.

Positive: "i have the worlds best dad"

Negative: "pouring rain outside . wish i could go out"

- **Goal**: train classifier using word vectors to predict polarity

Project 3: Semantic Segmentation

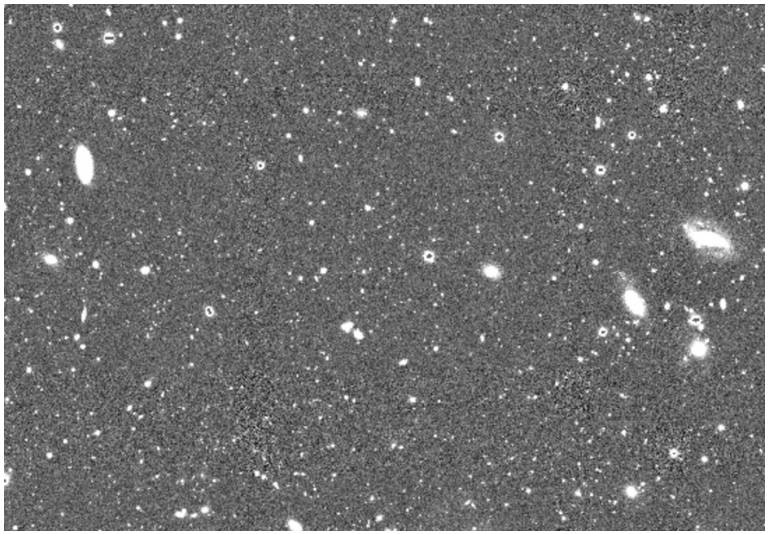
Extract roads from satellite images

- **Data**: set of satellite/aerial images acquired from GoogleMaps
- **Ground-truth**: images with pixels labeled as {road, background}.

- **Goal**: train a classifier to segment roads in these images, i.e. assign a label {road=1, background=0} to each pixel.

Project 4: Galaxy Image Generation

- **Data**: astronomical images acquired from wide field imaging surveys.
- **Ground-truth**: pixels labeled as {background, star, galaxy}



- **Goal**: train a generative model that can generate galaxy images.