Machine Learning Lab Course

Organizational Meeting

lecturer: Prof. Dr. Stephan Günnemann
This is a practical course (Praktikum) for **Master** students!

*Name of module: Large-Scale Machine Learning (IN2106, IN4192)*

website: [ml-lab.in.tum.de](http://ml-lab.in.tum.de)
Why attend our Machine Learning lab course?

1. Get the chance to implement and apply state-of-the-art ML algorithms

2. Gain hands-on experience working on real-world data, solving real-world tasks by working on projects offered by our industry partners.
   – Successful projects might even qualify for a subsequent master thesis.

3. Work on large-scale problems with the support of state-of-the-art GPU computing resources.
Requirements

- Requirements for the lab course
  - strong programming skills (Java, Python, C++, Java, etc.)
  - strong knowledge in data mining/machine learning
  - you should have passed relevant courses (the more, the better)
    - Mining Massive Datasets
    - Machine Learning
    - Our seminars
  - self-motivation

- Additional selection criteria
  - other relevant experience (projects in companies, experience as a HiWi)
    - you can send an overview of your experience to us (see end of slides)
Organization

- Groups of 3 students
- Each team will work on a different project in cooperation with one of our industry partners.

- Groups are allowed to (should) collaborate!
  - exchange your experience with the other groups
  - how do the other groups tackle certain problems?

- Technical aspects:
  - each group will get exclusive access to one high-end GPU server with
    - 4x NVIDIA GPU w/ 11GB RAM
    - 10-core CPU
    - 256 GB RAM
  - scale up your models and data!
Organization

- Weekly meetings (around 90-120 minutes)
  - each group should briefly report their progress, open problems, and next steps
- Regular documentation of your work
  - status reports and documentation (we have set up a wiki)
  - use of a central code repository
This semester’s industry partners

BMW

ilass AG

SIEMENS
Industry project: Situation recognition

- **Autonomous driving** is expected to be one of the most disruptive technologies of the 21st century and an active field of research. One crucial component is the robust and reliable recognition of situations.

- In this project, the goal is to use various (sensor) data collected in cars to recognize certain **driving situations** (e.g. turns, signals, parking, ...).

- There are many different **data streams** available for the students (e.g. velocity, acceleration, steering wheel angle, turn signal status) for this task. GPS data can be used for verification of the results.
Industry project: Oktoberfest food classification

- Industry partner: **ilass AG**, maker of software for gastronomy and party tents (e.g. Oktoberfest).

- The project will be about detecting and classifying **food items on images** to be extracted from a **video** stream.

- This semester’s focus will be to incorporate the temporal dimension of the video stream.
Industry project: Virtual sensors

- 3D simulations are an important tool for testing products such as gas turbines or motors.

- During a simulation, we can find out the condition (temperature, rotation speed, etc.) of the product at any position we are interested in. In a physical product during operation, the positions at which we can place hardware sensors are very limited.

- The goal of this project is to create a machine learning model that, based on sensor data at certain positions, can accurately infer the conditions at virtual sensor positions where we cannot place hardware sensors (e.g. a rotor) over time.
Industry project: Health indexing

- **Predicting failures** in a device during operation **before** they take place can increase product safety, reduce downtime, and reduce repair costs.

- However, it is very difficult to detect **anomalies** or possible failures in the vast stream of sensor data, and the **rarity** of such events means that we have very little ‘labeled’ data.

- In this project, the goal is to simulate **device failures** in products such as gas turbines to record **labeled data** and use this data to **predict** these **failures** based on new data.
Registration via the matching system!

*Module name: Large-Scale Machine Learning (IN2106, IN4192)*

+ fill out the application form (see next slide)
Your Experience

- Fill out our brief online form about your experience **by July 4, 2018**
  - you can provide us with a list of your experience in data mining/machine learning (courses, projects, ...)
  - please send a short overview only (bullet list); not a complete CV

- A link to the registration form can be found at [ml-lab.in.tum.de](http://ml-lab.in.tum.de).