Inmas Machine Learning Workshop 2024

Internship Network in the Mathematical Sciences

Christian Kümmerle

January 13-14, 2024

Learning Goals

- Gain intuition about what are fundamental problems and concepts to **learn from data**
- Exposure to some **popular models and computational tools** to solve machine learning problems
- Learn about **different** data **types**
- Gain intuition about **challenges & peculiarities** of **high-dimensional** data
- Learn how to use **Python** popular packages to **apply** techniques

Schedule

Today, Saturday, Jan 13

- **10:00 AM - 1:00 PM ET (9:00 AM - 12:00 PM CT):**

  Framework of Statistical Learning, Regularization, High-Dimensional Data

- **90 minutes lunch break**

- **2:30 PM - 5:30 PM ET (1:30 PM - 4:30 PM CT):**

  Classification Problems, Natural Language Processing

Structure of Workshop:

~ <=1 h per Session: Presentation
~ >=2 h per Session: Work in Groups of 5-6 on Python Jupyter Notebooks

Schedule

Tomorrow, Sunday, Jan 14

- **10:00 AM - 1:00 PM ET (9:00 AM - 12:00 PM CT):**

  - Unsupervised Learning: Principal Component Analysis, Clustering

- **90 minutes lunch break**

- **2:30 PM - 5:30 PM ET (1:30 PM - 4:30 PM CT):**
  - Neural Networks and Deep Learning

A bit about myself

ckuemmerle.com

Current Position:

- Assistant Professor in Computer Science at University of North Carolina at Charlotte since 2022

Background:

- Ph.D. in Mathematics (Technical University of Munich, Germany)
- Postdoc at Johns Hopkins 2020-2022

Research Interests:

- **Make machine learning & AI more powerful, more resource-efficient, more data-efficient**

Optimization for machine learning, development of scalable algorithms, few-shot learning, recommender systems, high-dimensional probability

Our TA Team

- Ben Brindle (JHU)
- George Kevrekidis (JHU)

What is Data Science?

- [Tukey ’62 “The Future of Data Analysis]:
  “Data Analysis” as an **empirical science**:

  - Procedures for gathering data, for interpreting data
  - Uses mathematical statistics
  - **“reliance upon the test of experience as ultimate standard of validity"**

Focus in this workshop: Prediction instead of Inference

Why has Data Science/ ML become so big?

In last 15-20 Years: Massive technological advances in
Pattern & image recognition, machine translation,
targeted advertisement, semiautonomous cars

- [Liberman 2010; Donoho 2015]: One crucial ingredient:
  Common Task Framework

- Public “training” data set: List of observations with labels
- Competitive participants with **common task** to infer label **prediction rule** from training data, **submit to**
-> Referee mechanism which reports accuracy of prediction rule when applied to (hidden) test dataset.

Examples:
- $1M Netflix Prize: Recommender Systems
- ImageNet: Image Classification

What is Machine Learning?

Tom Mitchell (CMU), 1997:

*"A computer program is said to **learn from experience E** with respect to some **class of tasks T**, and **performance measure P**, if its performance at tasks in T, as measured by P, improves with experience E."*