ΠΙΠ

Seminar: Robust Data Mining Techniques
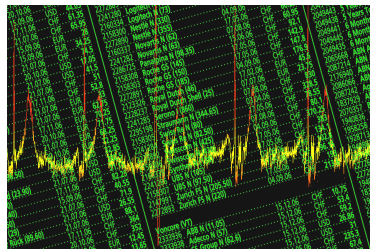
Pre-course meeting

Technische Universität München
Department of Informatics
Data Mining and Analytics
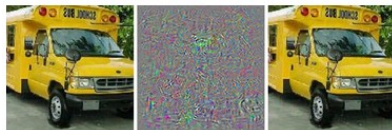kdd.in.tum.de

January 27, 2017

Why robustness matters?

- Data is inherently noisy
- Guarantees in performance-critical areas
- Adversarial scenarios
- Crowdsourcing
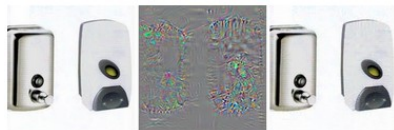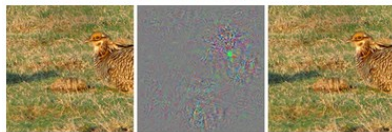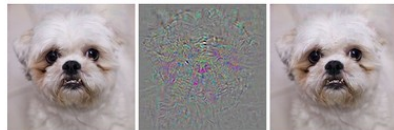
Example: Linear Regression

| correct | +distort | ostrich |
| correct | +distort | ostrich |

Data Mining
and Analytics
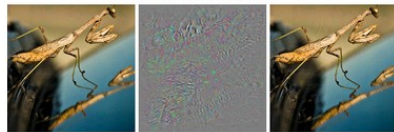
Topics

Robust extensions to data mining algorithms

- Matrix factorization
- Regression
- Classification
- Clustering
- Time series analysis

Further topics in robust machine learning

- Robustness of complex networks
- Differential privacy
- Fooling learning algorithms with adversarial examples
- Designing algorithms for the adversarial setting

Data Mining
and Analytics

Learning Outcome

You are going to learn

- about the design of robust data mining algorithms
- to read and understand scientific publications
- to write a scientific report
- how to prepare and give a technical talk

Requirements

Paper

- 5 - 8 pages
- Latex template on the course webpage

Presentation

- 30 minutes talk
- 15 minutes discussion

Reviews

- Everyone has to review 2 papers by other students

Data Mining
and Analytics

Deadlines

- **1 week before the talk** - submission of extended abstract and slides
- **Day of the talk** - submission of preliminary paper for review
- **1 week after the talk** - receiving comments from reviewers
- **2 weeks after the talk** - submission of the final paper

Data Mining
and Analytics

Grading

The grade is determined based on

- Report
- Presentation (slides and speech)
- Reviews written by **you**
- Involvement in the class
- Interactions with the supervisor
- Extra bonuses for own contributions (e.g. visualizations, demos, experiments)

Data Mining
and Analytics

Schedule

- **Before 08.02.** - fill out the pre-course survey
  <https://goo.gl/forms/rWxXhsMpZPW6kpcE2>
- **03.02.** - **08.02.** - registration via the matching system
- **After 15.02.** - notification of the participants and selection of topics
- **April** - **June** - weekly sessions every Monday 14:30 - 16:00

Questions?