

IE 8930 – Special Topics on High-Dimensional Data Analytics

Instructor: Dan Li, 269 Freeman Hall, dli4@clemson.edu

Textbook

There is no required textbook for this course, but a reading list will be provided for complementary information.

Software: R, MATLAB, and Python

Catalog Course Description

This course focuses on analysis of high-dimensional structured data including profiles, images, and other types of functional data using statistical machine learning. A variety of topics such as functional data analysis, image processing, multilinear algebra and tensor analysis, and regularization in high-dimensional regression and its applications including low-rank and sparse learning is covered. Optimization algorithms commonly used in statistical modeling and machine learning and deep learning will also be discussed.

Course Outcomes

Students will

- Learn machine learning and statistical methods for image processing and analysis of functional data;
- Learn a variety of regularization techniques and their applications;
- Be able to use multilinear algebra and tensor analysis techniques for performing dimension-reduction on a broad range of high-dimensional data.
- Understand how to use well-known optimization methods to create efficient learning algorithms.

Grading Policy

- Homework 40%
- Exam I 30%
- Final Project 30%

Grading Scale

Your final grade will be assigned as a letter grade according to the following scale:

- A 90-100%
- B 80-89%
- C 70-79%
- D 60-69%
- F 0-59%

Topics Covered:

The topics covered to be covered are subject to change.

1. Functional data analysis
 - 1.1 Introduction to HD & Functional Data
 - 1.2 Review of Regression
 - 1.3 Splines
 - 1.4 BSplines
 - 1.5 Smoothing Splines
 - 1.6 Kernel Smoothers

- 1.7 FPCA
- 1.8 Gaussian Process
- 2. Image processing (5 lessons)
 - 2.1 Introduction to Image Processing
 - 2.2 Image Transformation
 - 2.3 Convolution & Image Filtering
 - 2.4 Image Segmentation
 - 2.5 Edge Detection
- 3. Tensor data analysis
 - 3.1 Multilinear Algebra Preliminaries
 - 3.2 Tensor Decomposition Methods
 - a. Tucker
 - b. CANDECOMP/PARAFAC (CP)
 - c. DEDICOM
 - 3.3 Tensor Regression
 - 3.4 Tensor Analysis Applications
- 4. Optimization applications in HD analysis
 - 4.1 Introduction
 - 4.2 First Order Methods
 - 4.3.1 Gradient Descent
 - 4.3.2 Accelerated algorithms
 - 4.3.3 Stochastic Gradient Descent
 - 4.3 Second Order Methods
 - Newton method
 - Quasi-Newton method
 - Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm
 - 4.4 Decomposable problems
 - 4.5 Coordinate descent
 - 4.6 Proximal gradient descent
 - 4.7 Augmented Lagrangian Methods
 - 4.8 Alternating Direction Method of Multipliers
- 5. Regularization
 - 5.1 Introduction
 - 5.2 Ridge
 - 5.3 Lasso
 - 5.4 Non-negative Garrote and Adaptive Lasso
 - 5.5 Group Lasso
 - 5.6 Elastic Net
- 6. Applications of regularization
 - 6.1 Compressive sensing
 - 6.2 Matrix completion
 - 6.3 Robust PCA
 - 6.4 Smooth-Sparse Decomposition (SSD)
 - 6.5 Kernel Ridge Regression and Reproducing Kernel Hilbert Space (RKHS)
 - 6.6 Anomaly detection for streaming data
 - 6.7 Spatiotemporal SSD