

STA 7934, Fall 2013

Dimension Reduction

Course Instructor

Z. Su, 205 Griffin-Floyd Hall (352-273-2993)

email: zhihuasu@stat.ufl.edu

Office Hours: 2:00-3:00 TR, 205 Griffin-Floyd Hall, or by appointment.

Lectures

T: 10:40 am -11:30 am, R: 10:40 am - 12:35 pm.

Text

None, but the following book may be useful for reference: “Regression Graphics: Ideas for Studying Regressions through Graphics” by R. D. Cook. The web page for the text is at <http://www.stat.umn.edu/RegGraph/>.

Homework

Homework is a required part of the course. There will be homework assignments throughout the semester, portions of which will be graded.

Grading

A grade of “B” requires satisfactory completion of the homework problems and reading assignments, along with regular attendance and participation in classroom discussion. A grade of “A” requires completion of a class project involving detailed study of some aspect of the course material. Project, which must be approved in advance, should be underway by early November. Project suggestions will be given in class from time to time. You should expect to spend about 1/4 of your time on the project.

Exam

None planned at present. Some project presentations might be scheduled during finals week.

Incompletes

Grades of “I” will be given only in extraordinary circumstances, and then only by written agreement between the instructor and the student.

Computing

Matlab will be the primary computing platform for this course. Some methods are available in R via Weisberg’s `dr` package, but many of the novel methods have been written only in Matlab. The Matlab code and documentation are available at <https://sites.google.com/site/lilianaforzani/ldr-package> and <http://code.google.com/p/envlp/>.

Coverage

This course will consider both traditional and modern methods of dimension reduction, and attempt to construct a common framework that may suggest new theory and methods. Traditional methods to be discussed include principal components and partial least squares. More modern methods include several methods that fall under the heading of “sufficient dimension reduction”. Emphasis will be placed on contrasting historical and modern conditions. There will likely be more questions than answers.

Reduction of the dimensionality of the predictor vector is the primary goal in regressions with a univariate response. There are several reasons why dimension reduction may be useful in this context, including the possibilities of mitigating the effects of collinearity, facilitating model specification by allowing visualization of the data in low dimensions, providing a relatively small set of predictors on which to base prediction or interpretation, and dealing usefully with large-p-small-n problem. When the response is multivariate, reduction of the response vector and the predictor vector may be considered separately or simultaneously.

Most of classical Twentieth-century Fisherian statistics focused on problems where the number of unknown “p” was small and, in particular, much smaller than the number of observations or experimental units. However, with advances in computing and the emergence of applications with relatively large p, the practical environment has changed dramatically over the past 20 years. The statistical community has not yet decided how to deal effectively with related issues. This course is intended in part to be a contribution to the discussion.

Assignment 1

Reading: C.J.C. Burges, “Dimension reduction: a guided tour”, Foundations and Trends in Machine Learning, Vol 2, No. 4, 275-365, 2010.

Writing: As discussed in class, search the Internet for types of dimension reduction and write a brief introduction to one type. To avoid overlap, your choice must be approved by the instructor prior to writing (an email will be sufficient). Due: Thursday, Aug. 29.

Disability access statement

This material is available in alternative formats upon request.