

New Course Request

Indiana University

Indianapolis Campus

Check Appropriate Boxes: Undergraduate credit Graduate credit Professional credit

1. School/Division Science, Biostatistics 2. Academic Subject Code BIOS
3. Course Number S612 (must be cleared with University Enrollment Services) 4. Instructor Harezlak
5. Course Title Modern Statistical Learning Methods

Recommended Abbreviation (Optional) _____
(Limited to 32 Characters including spaces)

6. First time this course is to be offered (Semester/Year): Fall 2011

7. Credit Hours: Fixed at 3 or Variable from _____ to _____

8. Is this course to be graded S-F (only)? Yes _____ No X

9. Is variable title approval being requested? Yes _____ No X

10. Course description (not to exceed 50 words) for Bulletin publication: P: STAT 52500

This course will cover the topics pertaining to the modern methods of high-dimensional data analysis.

11. Lecture Contact Hours: Fixed at 3 or Variable from _____ to _____

12. Non-Lecture Contact Hours: Fixed at N/A or Variable from _____ to _____

13. Estimated enrollment: 15 of which 100 percent are expected to be graduate students.

14. Frequency of scheduling: annually Will this course be required for majors? yes

15. Justification for new course: Course for new Biostatistics ^{Ph.D.} program

16. Are the necessary reading materials currently available in the appropriate library? yes

17. Please append a complete outline of the proposed course, and indicate instructor (if known), textbooks, and other materials.

18. If this course overlaps with existing courses, please explain with which courses it overlaps and whether this overlap is necessary, desirable, or unimportant.

19. A copy of every new course proposal must be submitted to departments, schools, or divisions in which there may be overlap of the new course with existing courses or areas of strong concern, with instructions that they send comments directly to the originating Curriculum Committee. Please append a list of departments, schools, or divisions thus consulted.

Submitted by:

[Signature] Date 10-5-2009
Department Chairman/Division Director

Approved by:

[Signature] Date 10/30/2009
Dean

Date _____
Dean of Graduate School (when required)

Date _____
Chancellor/Vice-President

Date _____
University Enrollment Services

After School/Division approval, forward the last copy (without attachments) to University Enrollment Services for initial processing, and the remaining four copies and attachments to the Campus Chancellor or Vice-President.

BIOS 598: Modern Statistical Learning Methods **SPRING 2009**

Instructor:

Jaroslav Harezlak, PhD, Assistant Professor

Contact information:

Jaroslav Harezlak, PhD
Division of Biostatistics
Indiana University School of Medicine
410 West 10th Street, Suite 3000
Indianapolis, IN 46202

Office: (317) 274-2682

Email: harezlak@iupui.edu

Prerequisites: STAT 525 and familiarity with regression analysis.

Course Description:

Recent years have brought rapid growth in the amount and complexity of data in biostatistical applications. Among others, data collected in the genomic, proteomic and other “-omic” areas call for new statistical techniques in both predictive and descriptive learning. This course will cover the topics pertaining to the modern methods of high-dimensional data analysis. The specific topics will include penalized regression methods, linear classification methods, basis expansions and regularization, kernel methods, model assessment and selection, model inference and averaging, boosting and additive trees, neural networks, support vector machines, nearest-neighbor methods, and unsupervised clustering. Class presentations and homework assignments will cover both the methodological issues and the applications of the methods to data analysis.

Educational objectives:

At the end of the course students will be able to:

- use modern statistical methods for the analysis of high dimensional data
- utilize statistical software R to conduct data analysis of such data

Meeting Times: Lectures twice per week

COURSE OUTLINE

Lectures 1-3.	Introduction; Supervised Learning
Lectures 4-5	Linear Methods for Regression
Lecture 6.	Linear Methods for Classification
Lectures 7-8.	Nonparametric Methods
Lectures 9-10	Basis Expansions and Regularization
Lectures 11-12.	Kernel Methods
Lectures 13-14	Model Assessment and Selection
Lectures 15-16	Model Inference and Averaging
Lectures 17-19	Additive Models, Trees, and Related Methods
Lectures 20-21	Boosting and Additive Trees
Lectures 22-23	Neural Networks
Lectures 24-26	Support Vector Machines & Discriminant Analysis
Lecture 27	Nearest-Neighbor methods
Lectures 28-29	Unsupervised Learning
Lectures 30-31	Student presentations

Required and Recommended Texts:

The course will use the following textbook:

Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning, Springer, New York.

The lectures will follow the textbook closely, but we will supplement the textbook with instructors' own lecture notes,

Evaluation and Grading: Students will be evaluated based on their performance on the homework assignments (50%), and the final project (50%). Letter grades for the course are assigned using the following scale: A: 90-100; B: 80-89; C: 70-79; D: 60-69; F: less than 60. Within each letter grade, "+" and "-" will be assigned if the numeric score is in the top and bottom quintiles, respectively.

Cheating and Plagiarism:

Academic misconduct will *not* be tolerated and all cases will be reported. Examine the IU Code of Student Rights, Responsibilities, and Conduct at <http://life.iupui.edu/help/code.asp> and in particular examine the rules regarding academic misconduct at http://life.iupui.edu/help/docs/Part_3all.html. Violations of these rules will result in a grade of "F" (or 0%) for the assignment in question, and may result in an "F" for the course or even expulsion from the university (see http://life.iupui.edu/help/docs/Part_4all.html#sanction).

Americans with Disabilities Act:

If you need any special accommodations due to a disability, please contact Adaptive Educational Services at (317)-274-3241. The office is located in CA 001E.