FA23 - INFO-I415 Introduction to Statistical Learning

Indiana University School of Informatics and Computing Indianapolis

Section Number: 32018
Credit Hours: 4 (Lecture 3, Lab 1)
Canvas Course Site: https://iu.instructure.com/courses/2133390
Instructor: Elie Salomon
Email: elisalom@iu.edu (use Canvas inbox for course matters)
Prerequisites: None
Instruction mode: This course is offered online only (asynchronous)

COURSE DESCRIPTION

Readings/Resources

All textbooks and readings are available via downloadable PDFs in the modules on the Canvas course site:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. An introduction to statistical learning with applications in R
- Additional readings as assigned

Assessments

Assessment Details

Each student should not only read the assigned material but also arrive at a competent understanding of it prior to assessment. These measures will be used to assess student learning outcomes:

1. Assignments (30%)
2. Quizzes (15%)
3. Labs (15%)
4. Midterm (20)
5. Final Exam (20%)

1. Assignments are selected among conceptual and applied problems at the end of Chapters 3 to 5 of James et al. and/or provided by the instructor on informatics and related applications.
2. Quizzes assess comprehension and skill acquisition. Structurally, the quizzes are administered via Canvas. Questions will include simple true/false, multiple-choice, matching, fill in the blank, and short essays. Students work under a time limit and will be notified of errors after the submission of the quiz. You will be given two chances for each quiz, and the highest score will be kept as your grade.
3. **Labs** provide students the opportunity to gain practical experience programming in R. Labs are assessed by the completion of lab exercises and/or by the creation of short programs that demonstrate skills employed in the lab exercises.

4. The **final and midterm** have a format resembling the assignments and quizzes.

### Learning Outcomes

* RBT: Revised Bloom’s

** PLUS: Profiles of Learning for Undergraduate Success

*** Applied Data and Information Science Program Learning Outcomes

#### Revised Bloom's Taxonomy (RBT)*

The revised Bloom’s taxonomy (RBT) presents a way to classify different types of learning experiences across two levels: 1) The revised Bloom’s taxonomy cognitive process (RBTCP) dimension and 2) the revised Bloom’s taxonomy knowledge (RBTK) dimension. The RBTCP dimension represents a continuum of increasing cognitive complexity—from remember to create—across six levels:

1. **Knowledge/Remembering:** The ability to recall or recognize specific information or data.
2. **Understanding:** Understanding the meaning of informational materials, translation, interpolation and interpretation of instructions and problems.
3. **Application:** The use of previously learned information in new and concrete situations to solve problems that have single or best answers.
4. **Analysis:** Breaks down information/concepts into smaller components. Each component is identified and understood as is the relationship of these components to the whole.
5. **Evaluation:** The ability to apply a criterion or set of standards to conclude a value judgment.
6. **Creation, Synthesis:** The ability to merge knowledge into creating a new meaning or structure including demonstrating how and why various diverse elements work together.

#### Profiles of Learning for Undergraduate Success (PLUS)**

- **1 Problem Solver – Thinks Critically**
- **3 Problem Solver – Analyzes, Synthesizes, And Evaluates**
- **4 Community Contributor – Anticipates consequences**
- **3 Community Contributor – Behave ethically**

#### ADIS Program Learning Outcomes***

**Applied Data and Information Science Learning Outcomes**

**A. Data Literacy**

1. Distinguish between data, information, and knowledge.

2. Analyze the value and key role data plays in society in providing opportunities to expand knowledge, to innovate, and to influence.

**Profiles of Learning of Undergraduate Success**

- **P2.1 Problem Solver** – Think critically
- **P2.3 Problem Solver** – Analyzes, synthesizes, and evaluates
- **P4.4 Community Contributor** – Anticipates consequences
3. Analyze datasets in context to determine data veracity including bias in data collection or representation.

4. Assess values with respect to the use of data technologies.

B. Data Science

1. Organize, visualize, and analyze large, complex datasets using descriptive statistics and graphs to make decisions.

2. Apply inferential statistics, predictive analytics, and data mining to informatics-related fields.

3. Assess the purpose, benefits, and limitations of visualizations as a human-centered data analysis methodology.

4. Conceptualize and design effective visualizations for a variety of data types and analytical tasks.

5. Identify, assess, and select appropriately among data analytics methods and models for solving real-world problems, weighing their advantages and disadvantages.

6. Understand data science concepts, techniques, and tools to support big data analytics.


8. Explore, transform, and visualize large, complex datasets with graphs in R.

9. Solve real-world problems by adapting and applying statistical learning methods to large, complex datasets.

10. Identify, assess, and select among statistical learning methods and models for solving a particular real-world problem, weighing their advantages and disadvantages.

11. Write programs to perform data analytics on large, complex datasets in R.
12. Analyze data from case studies in informatics related fields.

**C. Information Science**

1. Demonstrate an understanding of the data lifecycle, including data curation, stewardship, and long-term preservation.

2. Apply the principles of consistency and uniformity to recognize the need for authorized terms for describing various types of data.

3. Understand the principles of data organization including file name conventions, version control, and data documentation.

4. Understand the characteristics of various data types generated and used by a variety of disciplines, subdisciplines, research communities, and government organizations.

5. Understand critical issues associated with the storage, backup, and security of data.

6. Analyze data policies to compare possible outcomes.

**D. Data Ethics**

1. Understand the relation between data, ethics, and society.

2. Identify and understand the social, political, ethical, and legal aspects of data creation, access, ownership, service, and communication.

3. Develop substantive arguments using ethical reasoning to suggest improvements to data-driven systems and practices.

4. Differentiate between surveillance systems that promote and inhibit values.
E. Other Topics

1. Design, conduct, and write up results of research.
   - P3.1 Innovator – Creates/designs*
   - P1.4 Communicator – Conveys ideas effectively*
   - P2.3 Problem Solver – Analyzes, synthesizes, and evaluates*

2. Understand tools and techniques of project management.
   - P2.3 Problem Solver – Analyzes, synthesizes, and evaluates*
   - P2.1 Problem Solver – Thinks critically*

3. Understand legal and business aspects of technology and media.
   - P2.1 Problem Solver – Thinks critically*

Major emphasis** Minor emphasis* Some emphasis: no mark

Course Schedule

This is only a tentative schedule. The instructor may adjust it to better fit the class.

<table>
<thead>
<tr>
<th>Weeks</th>
<th>Topics</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 09, 2023</td>
<td>Introduction to Statistical Learning</td>
<td>Quiz 1</td>
</tr>
<tr>
<td>Jan 15, 2023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan 16, 2023</td>
<td>Martin Luther King Jr. Day - No Classes</td>
<td></td>
</tr>
<tr>
<td>Jan 22, 2023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan 23, 2023</td>
<td>Introduction to R Programming</td>
<td>Lab 1</td>
</tr>
<tr>
<td>Jan 29, 2023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan 30, 2023</td>
<td>R Programming Continue...</td>
<td>Lab 2</td>
</tr>
<tr>
<td>Feb 05, 2023</td>
<td></td>
<td>Quiz 2</td>
</tr>
<tr>
<td>Feb 06, 2023</td>
<td>Linear Regression (Part 1)</td>
<td>Assignment 1</td>
</tr>
<tr>
<td>Feb 12, 2023</td>
<td></td>
<td>Quiz 3</td>
</tr>
<tr>
<td>Feb 13, 2023</td>
<td>Linear Regression (Part 2)</td>
<td>Lab 3</td>
</tr>
<tr>
<td>Feb 19, 2023</td>
<td></td>
<td>Assignment 2</td>
</tr>
<tr>
<td>Feb 20, 2023</td>
<td>Classification (Part 1)</td>
<td>Quiz 4</td>
</tr>
<tr>
<td>Feb 26, 2023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb 27, 2023</td>
<td>Classification (Part 2)</td>
<td>Lab 4</td>
</tr>
<tr>
<td>Mar 05, 2023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Event</td>
<td>Quiz</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Mar 06, 2023</td>
<td>Midterm Exam Week</td>
<td>5</td>
</tr>
<tr>
<td>Mar 12, 2023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 13, 2023</td>
<td>Spring Break Week – No Class</td>
<td></td>
</tr>
<tr>
<td>Mar 19, 2023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 20, 2023</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Mar 26, 2023</td>
<td>Resampling Methods</td>
<td>6</td>
</tr>
<tr>
<td>Mar 27, 2023</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Apr 02, 2023</td>
<td>Linear Model Selection and Regularization</td>
<td>6</td>
</tr>
<tr>
<td>Apr 03, 2023</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Apr 09, 2023</td>
<td>Moving Beyond Linearity</td>
<td>7</td>
</tr>
<tr>
<td>Apr 10, 2023</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Apr 16, 2023</td>
<td>Tree-Based Methods</td>
<td>8</td>
</tr>
<tr>
<td>Apr 17, 2023</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Apr 23, 2023</td>
<td>Support Vector Machines</td>
<td>9</td>
</tr>
<tr>
<td>Apr 24, 2023</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Apr 30, 2023</td>
<td>Unsupervised Learning</td>
<td>10</td>
</tr>
<tr>
<td>May 01, 2023</td>
<td>Final Exam Week – No more classes</td>
<td></td>
</tr>
<tr>
<td>May 07, 2023</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
EXPECTATIONS, GUIDELINES, AND POLICIES

Deliverables

You are responsible for completing each deliverable (e.g., task, final project) by its deadline and submitting it by the specified method. Deadlines and submission instructions are outlined in the syllabus or in supplementary documents accessible through Canvas.

In fairness to the instructor and students who completed their work on time:

- **Late assignments** will have a 5% deduction per day for up to six days late.
- After six days, an assignment may still be accepted but will suffer a 50% deduction, and any additional deduction based on the assignment criteria.
- Instructors have the discretion to provide exceptions based on special circumstances, such as illness or emergency.
- If you have difficulties completing assignments, please communicate with your instructor (at least 8 hours before the assignment is due, or, more preferably, as soon as you know you will be late completing the assignment) to create a timeline and plan for the assignment's completion.

Your Questions, Concerns, and Comments

Please do not hesitate to contact the instructor directly via Canvas mail with any questions. If needed, the instructor will also use Canvas Announcements to notify the entire group (e.g., syllabus change, instructor availability, etc.).

If you have problems accessing Canvas, please contact the University Information Technology Services (UITS) Support Center at uits.iupui.edu or 317-274-HELP. All course announcements will be found in Canvas along with the course schedule, assignments, and other course documents.

Attendance

The course will be taught entirely online including web-based readings and resources, threaded discussions, plus online presentations and activities.

This course assumes that students can work independently. There are no required face-to-face meetings. There are no required synchronous online meetings. However, students are encouraged to e-mail or arrange an online chat with the instructor at any time.

A basic requirement of this course is that you will participate in all class activities and conscientiously complete all required course assignments. Students are expected to complete the assignments, quizzes, and projects on time, which is your attendance.

Incomplete

Incompletes are not automatically granted. You may arrange a grade of “I” or incomplete for a course with an instructor for special circumstances. Students need to have completed the majority of course work (75%+) at an acceptable level of achievement. You and the instructor must agree upon the terms for completing the course. Students who have multiple incompletes (2 or more) will be blocked from registering for additional LIS courses.
until there is only one (or zero) outstanding incomplete, or the student presents the department chair with a plan of action for completing all incompletes in a timely way.

Deadlines for the work for an incomplete to be finished are at the instructor’s discretion. The deadline can be no longer than 1 year from the end of the semester but can be earlier if the instructor specifies that. Left unchanged, an Incomplete automatically becomes an F after one year. See: Student Central: Incompletes (Links to an external site.) (https://studentcentral.iupui.edu/grades-progress/incompletes.html (Links to an external site.))

**GRADING SCALE**

Grades will be assigned based on the IUPUI grading scale (Links to an external site.).

- A+ 97–100% Professional level work, showing highest level of achievement
- A 93–96.99% Extraordinarily high achievement, quality of work; shows command of the subject matter
- A– 90–92.99% Excellent and thorough knowledge of the subject matter
- B+ 87–89.99% Above average understanding of material and quality of work
- B 83–86.99% Mastery and fulfillment of all course requirements; good, acceptable work
- B– 80–82.99% Satisfactory quality of work
- C+ 77–79.99% Modestly acceptable performance and quality of work
- C 73–76.99% Minimally acceptable performance and quality of work
- C– 70–72.99% Unacceptable work (Core course must be repeated for credit)
- D+ 67–69.99% Unacceptable work (Course must be repeated for credit)
- D 63–66.99% Unacceptable work
- D– 60–62.99% Unacceptable work
- F Below 60 Unacceptable work

No credits are granted for a grade below C.

**APPLIED DATA AND INFORMATION SCIENCE PROGRAM OUTCOMES**

**CODE OF CONDUCT**

**OTHER POLICIES**

**MISSION STATEMENT**

**STATEMENT OF VALUES**

IUPUI values the commitment of students to learning; of faculty to the highest standards of teaching, scholarship, and service; and of staff to the highest standards of service. IUPUI recognizes students as partners in learning. IUPUI values the opportunities afforded by its location in Indiana’s capital city and is committed to serving the needs of its community. Thus, IUPUI students, faculty, and staff are involved in the community, both to provide educational programs and patient care and to apply learning to community needs through service. As a leader in fostering collaborative relationships, IUPUI values collegiality, cooperation, creativity, innovation, and entrepreneurship as well as honesty, integrity, and support for open inquiry and dissemination
of findings. IUPUI is committed to the personal and professional development of its students, faculty, and staff and to the continuous improvement of its programs and services.