ILLINOIS
School of Information Sciences

# Concepts of Machine Learning

**IS 327**
*Spring 2025*
**Mondays, Wednesdays, and Fridays, 01:00 PM - 01:50 PM**
**331 Armory**

**Instructor: Yaoyao Liu**
**Office hours: Wednesdays, 4:00 PM - 5:00 PM**
**Office hour location: Office 5125, 614 East Daniel St**
**Email: lyy@illinois.edu**

**TA: Naren Sudhir Shetty**
**Office hours: Tuesdays, 2:00 PM – 3:00 PM**
**Office hour location: Office 4170, 614 East Daniel St**
**Email: narenss2@illinois.edu**

***Note for contact.*** *The course has a Discussions tab set up on Canvas. All questions regarding course content or that might be of interest to other students, in particular, programming or course logistics questions, should be posted to the Discussions tab. If your question requires confidentiality or is only pertinent to you, please use email.*

**Course Description**
A dramatic increase in computing power has enabled new areas of data science to develop in statistical modeling and artificial intelligence, often called “machine learning”. Machine learning covers predictive and descriptive learning, and bridges theoretical and empirical ideas across disciplines. We will focus on concepts and methods for predictive learning: estimating models from data to predict unknown outcomes. Model types will include decision trees, linear models, nearest neighbor methods, and others as time permits. We will cover classification and regression using these models, as well as methods needed to handle large datasets. Lastly, we will discuss deep neural networks and other methods at the forefront of machine learning. We situate the course components in the “data science life cycle” as part of the larger set of practices in the discovery and communication of scientific findings.

The course will include lectures, readings, homework assignments, exams, and a class project. Most of the course activities will use Python with the *Pandas* library, which students should

already be proficient using. Students will learn how to use the *scikit-learn* Python library for machine learning during this course.

**Pre- and Co-requisites**
Students should be familiar with the concepts of tabular data (tables) and data types (categorical, ordinal, continuous, etc.) and be able to implement these concepts in Python using Pandas. Either STAT/CS/IS 107, IS 205, INFO 407, or at least 1 semester of programming experience using Python and Pandas is recommended as a prerequisite. Students should also be comfortable with basic geometry concepts such as points, lines, and distances. Sophomore, Junior, or Senior standing.

**Credit hours**
3

**Student Learning Objectives or Outcomes**
Upon successful completion of the course, students will be able to:

- Identify and abstract relevant information to formulate and clearly state a machine learning problem
- Construct and execute a logically appropriate process for solving the problem, including identification of assumptions and expected outcomes
- Devise a repeatable analytical strategy to explore a machine learning problem, test the strategy and evaluate its effectiveness, and revise the strategy as needed
- Formulate a conclusion and assess/justify its validity
- Improve critical reading, discussion, and presentation skills

**Course Context**
This course meets a number of learning outcomes connected to program objectives for the BS/IS program, which in turn connect to larger iSchool and University of Illinois learning goals.

**Program Learning Outcomes**

- Understand fundamental mathematical and programming tools for solving problems of information modeling, expression, and transformation
- Apply critical analytical skills to information issues

**iSchool Goal**
This course meets the following goal:

- Maintain global leadership in education for the information professions

**University of Illinois Campus-Wide Learning Goals**
*This course meets the following goals:*

1. *Intellectual Reasoning and Knowledge*
2. *Creative Inquiry and Discovery*
3. *Social Awareness and Cultural Understanding*

**Course Materials (Required)**
Raschka, S. & Mirjalili, V. (2019). *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2* (3rd ed.). Packt Publishing.

**Assignments and Methods of Assessment**
See the detailed list of topics and assignments at the end of this document for the schedule of assignments.

- Weekly homework: 35%
- In-class quizzes: 10%
- Exams: 35%
- Class project: 20%

Homework assignments will be mainly based on PrairieLearn (https://prairielearn.org). The exact submission deadlines for the homework assignments are shown on the PrairieLearn site. A rough reference of the due dates can be found in the **Week-by-Week Topic and Assignment Schedule** section. Please see the **Late Policy** section if you cannot meet the assignment submission deadlines.

The in-class quizzes and exams will also be taken on PrairieLearn. There will be 6 in-class quizzes randomly scheduled throughout the semester. Your overall quiz grade will be calculated based on your best 4 out of the 6. The content of the quizzes will closely resemble the homework assignments, despite being much shorter. There will be two exams scheduled in the Computer-Based Testing Facility (CBTF), respectively in the middle and towards the end of the semester.

The class project will start with a proposal document submission describing the dataset(s), the research question, and the proposed method of analysis. The proposal will likely describe the application of a regression or classification technique from class, including expected outcomes. The class project will carry out the research in the proposal. Students will give a presentation near the end of the course describing the research question, dataset, and comparing the expected and actual results. There will also be a project report of 500–1,000 words. The proposal contributes 25% of the overall project grade, while the presentation and report evenly split the other 75%.

Note: Students are permitted to use and adapt the work of others (including AI tools such as ChatGPT) for homework assignments, provided that the following guidelines are followed:

1. Use of other people's material must not infringe the copyright of the original author, nor violate the terms of any licensing agreement. Know and respect the principles of fair use with respect to copyrighted material.
2. Students must scrupulously attribute the original source and author of whatever material has been adapted for the assignment. Summarize the changes or adaptations that have been made. Make plain how much of the assignment represents original work.

**Late Policy**
Assignments will be accepted late for 80% of points until the end of Week 14 in this semester, but in-class quizzes, exams, and the project will not be accepted late without prior approval.

We recognize that unforeseen situations might occasionally disrupt your schedule. As such, we also offer "late day" credits for assignments (**but not applicable to in-class quizzes, exams, and the project**). These late days are designed for exceptional circumstances (like illness, computer issues, or power outages), so there should be ample allowance for you to get through the term.

You are granted **four (4)** late days to use throughout the course. Late days are all-or-nothing, meaning no fractional use. For example, if the deadline was 11 am on Monday and you submitted the assignment at 11:10 am on Monday, you will have to use 1 full late day to offset the 20% penalty; if you submitted the assignment at 11:10 am on Tuesday, you will have to use 2 full late days to offset the 20% penalty. **No assignment can be turned in more than 2 days late without penalty.** Therefore, you cannot use late days on an assignment that is more than 48 hours late.

You don't need to provide a justification or request approval to use your late days. At the end of the semester, we will distribute a Google form for you to specify how you allocated the late days across assignments. Your grades will be adjusted accordingly.

**Incomplete Grades**
An exceptional request for an incomplete grade is most often granted to students encountering a medical emergency or other extraordinary circumstances beyond their control. Students must request an incomplete grade from the instructor. The instructor and student will agree on a due date for completion of coursework. The student must submit an Incomplete Form signed by the student, the instructor, and the student's academic advisor to the front office: https://uofi.app.box.com/s/sx7arobhr0gfw12teaetmp1qq32ifdrd
Please see the Student Code for full details: https://studentcode.illinois.edu/article3/part1/3-104/

**Grading Scale**
The course grading scale follows:

94–100 = A
90–93.999 = A-
87–89.999 = B+
83–86.999 = B
80–82.999 = B-
77–79.999 = C+
73–76.999 = C
70–72.999 = C-
67–69.999 = D+
63–66.999 = D
60–62.999 = D-
59.999 and below = F

Close grade numbers are not rounded up because that simply changes the cutoff. For example, if 89.9 is just barely close enough to 90 to round up to A-, but 89.8 is not, then the cutoff is

effectively 89.9, and should be specified as such in the syllabus. At a certain point, there must be a cutoff; for the sake of fairness and preparedness, this course uses the cutoffs specified above.

**Attendance/Participation Policy**
The iSchool expects students to attend all classes except in cases of emergency. See the Student Code regarding attendance: https://studentcode.illinois.edu/article1/part5/1-501/

**Academic Integrity**
The iSchool has the responsibility for maintaining academic integrity so as to protect the quality of education and research in our school and to protect those who depend on our integrity. Consequences of academic integrity infractions may be serious, ranging from a written warning to a failing grade for the course or dismissal from the University. See the student code for academic integrity requirements: https://studentcode.illinois.edu/article1/part4/1-401/

**Statement of Inclusion**
https://diversity.illinois.edu/about/senate-diversity-resolution/
As the state's premier public university, the University of Illinois Urbana–Champaign's core mission is to serve the interests of the diverse people of the state of Illinois and beyond. The institution thus values inclusion and a pluralistic learning and research environment, one which we respect the varied perspectives and lived experiences of a diverse community and global workforce. We support diversity of worldviews, histories, and cultural knowledge across a range of social groups including race, ethnicity, gender identity, sexual orientation, abilities, economic class, religion, and their intersections.

**Religious Observances**
In keeping with our Statement of Inclusion and Illinois law, the University is required to reasonably accommodate its students' religious beliefs, observances, and practices in regard to admissions, class attendance, and the scheduling of examinations and work requirements. Religious Observance Accommodation Request form:
https://cm.maxient.com/reportingform.php?UnivofIllinois&layout_id=19

**Accessibility Statement**
To obtain accessibility-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES you may visit 1207 S. Oak St., Champaign, call 333-1970 (V/TTY), or e-mail a message to disability@illinois.edu.

**Week-by-Week Topic and Assignment Schedule (subject to revision)**
All assignments (including weekly homework assignments and project deliverables) will be released on PrairieLearn. The weekly assignments will be released each Friday morning (except for exam weeks), and are due by the end of the next Friday. The readings and assignments are intended to reinforce what is discussed in class afterward, but you may benefit from reading ahead before class as well.

| **Weekly Topics** | **Readings (R) / Assignments (A) due** |
|---|---|
| *Week 1*<br>Course introduction: Machine learning, artificial intelligence, and data science | R – Chapter 1 (pp. 1–17) |
| *Week 2*<br>Classification concepts with k-nearest neighbors | R – Chapter 3 *k-nearest neighbors* (pp. 103–107)<br>A – Week 1 homework |
| *Week 3*<br>Decision tree concepts | R – Chapter 3 *decision tree learning* (pp. 90–100)<br>A – Week 2 homework |
| *Week 4*<br>Cross-validation | R – Chapter 4 *partitioning a dataset* (pp. 121–124)<br>R – Chapter 6 *using k-fold* (pp. 195–201)<br>A – Week 3 homework |
| *Week 5*<br>Regression concepts with linear models | R – Chapter 10 *introducing linear regression* (pp. 315–318)<br>R – Chapter 10 *exploring housing* (pp. 318–325)<br>R – Chapter 10 *using regularized methods for regression* (pp. 337–339)<br>R – Chapter 10 *turning a linear regression model into a curve – polynomial regression* (pp. 3339–341)<br>A – Week 4 homework |
| *Week 6*<br>Logistic regression and linear SVM | R – Chapter 3 *modeling class probabilities via logistic regression* (pp. 60–67, 72–75)<br>R – Chapter 3 *maximum margin classification with support vector machines* (pp. 79–84)<br>A – Week 5 homework |
| *Week 7*<br>Regression with k-nearest neighbors and trees | R – Chapter 10 *implementing regression* (pp. 325–331)<br>R – Chapter 10 *evaluating the performance* (pp. 334–337)<br>R – Chapter 10 *dealing with nonlinear relationships* (pp. 345–350)<br>A – Week 6 homework |
| *Week 8*<br>Exam 1 | A – Exam 1<br>A – Week 7 homework |
| *Week 9*<br>Evaluating machine learning accuracy and fairness in depth | R – Chapter 6 *fine-tuning models* (pp. 207–211)<br>R – Chapter 6 *looking at metrics* (pp. 211–222)<br>A – Class project proposal due |

| **Weekly Topics** | **Readings (R) / Assignments (A) due** |
|---|---|
| *Week 10*<br>Feature selection and dimensionality reduction | R – Chapter 4 *selecting meaningful features* (pp. 127–140)<br>R – Chapter 4 *assessing feature importance* (pp. 141–143)<br>R – Chapter 5 *unsupervised dimensionality reduction via principal component analysis* (pp. 145–159)<br>A – Week 9 homework |
| *Week 11*<br>Clustering | R – Chapter 11 *grouping objects* (pp. 353–367)<br>A – Week 10 homework |
| *Week 12*<br>Deep neural networks concepts | R – Chapter 12 *modeling complex functions* (pp. 383–393)<br>R – Chapter 12 *training neural networks* (pp. 412–423)<br>R – Chapter 13 *activation functions* (pp. 462–470)<br>(Optional) R – Chapter 12 *implementing a multilayer perceptron* (pp. 400–412)<br>A – Week 11 homework |
| *Week 13*<br>Language model concepts | R – https://machinelearningmastery.com/what-are-large-language-models/<br>A – Week 12 homework |
| *Week 14*<br>Class project presentations | A – Class project report due |
| *Week 15*<br>Exam 2 | A – Exam 2 |