



AS.440.625.81

Machine Learning in Statistics

The class is conducted online



Professor

Ilya Rahkovsky

irahkov1@jhu.edu



Teaching Assistant

Shan (Kelly) Gao

sgao30@jhu.edu

A Syllabus may change during the semester. In case of change, the students will be promptly notified and the updated version of syllabus will be uploaded to the blackboard.

Course Description: This course focuses on the use of machine learning methods for in-sample and out-of-sample prediction. The topics include regression, classification, random trees (forests, boosting, and pruning), regularization, neural networks, support vector machines, model selection and ensemble learning. Prerequisite 440.606 Econometrics.

Course Access: The course material is located on [Github repo](#). Please install Github and create Github account on [Github](#). Clone the linked repository and you will see lecture code and HW assignments appear on Github folder.

There are two ways to run code in the class:

1. Google Colab: <https://colab.research.google.com> Go to upload and upload the code file to the Colab and it's ready to run.

Advantages: You don't have to install Python on your PC or worry about Python libraries. You can access your code from multiple computers. Google will make sure the code is able to run in 98

Disadvantages: The code will run 2-7 times slower than on your own PC (the exact difference depends on your computer speed). You will not learn important skills how to use Python on local machine and optimize different versions of Python libraries.

2. Install Anaconda on your machine from [Anaconda](#)

Install package either from `Mac_Lin_requirements.txt` or from `Win_requirements.txt` (files are in the Github repos)

To install package you run in terminal:

```
pip install package_name
```

If you move working folder in the terminal to the `ml_stat_class` you run bulk install of all Python libraries:

```
pip install -r Mac_Lin_requirements.txt
```

Or

```
pip install -r Win_requirements.txt
```

The homework answers, lecture video and exams are located/will appear in the Lessons folder in Blackboard.

The class requires a lot of programming on Python, so try to learn it as fast as possible. When you don't know how to code something the easiest solution is to Google your problem or look at <https://stackoverflow.com>

Learning Objectives:

1. Being comfortable using **Python** to perform a variety of Machine Learning procedures.
2. Learn basic concepts of ML: bias-variance trade-off, model selection, validation, and testing.
3. Learn common ML models: regression, classification, Support vector machines, decision trees, neural networks.

Text(s): *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2st Edition

Author(s): Aurélien Géron

ISBN-13: 978-1492032649

Grade Distribution:

Midterm Exam	24%
Final Exam	32%
Homework	44%

Letter Grade Distribution:

>= greater than 93 - 100	A	78 - 82	B
88 - 92	A-	70 - 77	B-
83 - 87	B+	0 - 70	C

Course Policies:

- **General**

- **Important!** The best way of communication is Blackboard forum. You can get quick answers there either from a teaching assistant (Roamn Maranets) or from your professor (Ilya Rahkovsky). If you want to ask a private question please use our emails. Feel free to send email to both your profesor and TA to maximize the speed of responce. There is a significant positive correlation between frequency of communications and grade. You don't want to lose points because your misunderstood the assignment or lecture concepts. Please contact me to clear it up.
- Exams are open book, notes performed on a personal laptop.
- Problem sets are posted on Blackboard on every Tuesday and are due the following Tuesday at 11:59 PM EST.
- Late assignment submissions are not accepted.
- For each assignment the students should upload two files to Blackboard:
 - * A file with written calculations or iPython code/Google Colab. The HW and exams need to be solved in Python unless specified otherwise. Please mark your answers clearly in the code, for example: The answer to each question should be clearly written like: "The answer for Question 4a is 0.5".
 - * A document with final answers in Excel, Word or PDF. For example:

```
accuracy = 5
print(f'Answer for the question 5A. The estimates accuracy is accuracy'
      Problem 1, part a. precision = 0.84, recall = 0.44
      Problem 1, part b. Evidence of model X outfitting because of low ROC statistics on
      the validation data
```
- Students can discuss class material, but they must work independently on all assignments. Significant correlations in answers of any two students will be investigated by the instructor and often will result in a punishment of all students submitting identical or very similar answers.

What To Expect in this Course

This course is 15 weeks in length and includes individual, group, and whole group activities in a weekly cycle of instruction. Each week begins on a Tuesday and ends on the following Tuesday. Please review the course syllabus thoroughly to learn about specific course outcomes and requirements. Be sure to refer to the Checklist each week, which provides a week-at-a-glance and shows targeted dates for the completion of activities.

Course Policies

Grading The instructor will aim to return assignments to you within 5-7 days following the due date, depending on the length of the assignment. You will receive feedback under the My Grades link on the left hand menu of your course.

Participation Requirements

You are expected to log into Blackboard regularly throughout the week. It is your responsibility to read all announcements and discussion postings within your assigned forums. You should revisit the discussion multiple times over the week to contribute to the dialogue.

Online Etiquette

In this course, online discussion will primarily take place in our online discussion board. In all textual online communication, it is important to follow proper rules of online etiquette... communicating with others in a proper and respectful way. For helpful tips, please see these Ground Rules for Online Discussions.

Course Protocols and Getting Help Amendments to the Course

Changes to the course will be posted in the Announcements section of your course. Please check announcements every time that you log into your online course.

Course Communication

You should communicate often with your classmates and the instructor. The majority of communication will take place within the Discussion forums. When you have a question about an assignment or a question about the course, please contact your instructor, or post your question in the course's "Syllabus & Assignment Question" forum.

Email Communication

For questions regarding course activities and assignments that would be of general interest to other students, please post those in the Discussion forum. If you have a question regarding course activities and assignments of a personal nature, please send an email message to the instructor and observe the following guidelines:

Include the title of the course in the subject field (e.g., JHU Insert Name of Course).

Feel free to contact your instructor with comments, questions, and concerns. All email messages will be sent to you via your JHU email account, so you should be in the habit of checking that account every day or you should ensure that your JHU email account forwards messages to another account of your choice.

Tentative Course Outline: The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments.

Week	Content
Week 1, January 26	<ul style="list-style-type: none"> • Python • Assignment: Learn Python
Week 2, February 2	<ul style="list-style-type: none"> • Introduction to ML, Regression • Chapter 1-2 • Assignment: HW1 is due
Week 3, February 9	<ul style="list-style-type: none"> • Classification and Training Models • Assignment: Chapter 3, HW2 is due
Week 4, February 16	<ul style="list-style-type: none"> • Training ML models • Assignment: Chapter 4, HW3 is due
Week 5, February 23	<ul style="list-style-type: none"> • Support Vector Machines • Assignment: Read Chapter 5, HW4 is due
Week 6, March 2	<ul style="list-style-type: none"> • Regression Trees and Forests • Reading assignment: Read Chapter 6-7, HW5 is due
Week 7, March 9	<ul style="list-style-type: none"> • Midterm Exam is posted • No Lecture • HW6 is due
Week 8, March 16	<ul style="list-style-type: none"> • Dimensionality Reduction and Unsupervised Learning • Reading assignment: Read Chapters 8-9, • Midterm Exam is due.
Week 9, March 23	<ul style="list-style-type: none"> • HW 7 is due • Neural Networks • Read Chapter 10
Week 10, March 30	<ul style="list-style-type: none"> • Spring Break • NO CLASS
Week 11, April 6	<ul style="list-style-type: none"> • Optimizing Neural Networks • Read Chapter 11, HW8 is due
Week 11, April 13	<ul style="list-style-type: none"> • Recurrent Neural Networks • Read Chapter 15, HW9 is due
Week 12, April 20	<ul style="list-style-type: none"> • Natural Language Processing • Read Chapter 16 • HW10 is due
Week 13, April 27	<ul style="list-style-type: none"> • Reinforcement Learning (extra credit material) • HW11 is due • read chapter 18
Week 15, May 4	<ul style="list-style-type: none"> • Final Exam is posted
Week 16, May 11	<ul style="list-style-type: none"> • Final Exam is due

University Policies

General

This course adheres to all University policies described in the academic catalog. Please pay close attention to the following policies:

Students with Disabilities

Johns Hopkins University is committed to providing reasonable and appropriate accommodations to students with disabilities. Students with documented disabilities should contact the coordinator listed on the [Disability Accommodations](#) page. Further information and a link to the Student Request for Accommodation form can also be found on the [Disability Accommodations](#) page.

Ethics & Plagiarism

JHU Ethics Statement: The strength of the university depends on academic and personal integrity. In this course, you must be honest and truthful. Ethical violations include cheating on exams, plagiarism, reuse of assignments, improper use of the Internet and electronic devices, unauthorized collaboration, alteration of graded assignments, forgery and falsification, lying, facilitating academic dishonesty, and unfair competition. Report any violations you witness to the instructor. Read and adhere to JHU's [Notice on Plagiarism](#).

Dropping the Course

You are responsible for understanding the university's policies and procedures regarding withdrawing from courses found in the current catalog. You should be aware of the current deadlines according to the [Academic Calendar](#).

Getting Help

You have a variety of methods to get help on Blackboard. Please consult the resource listed in the "Blackboard Help" link for important information. If you encounter technical difficulty in completing or submitting any online assessment, please immediately contact the designated help desk listed on the [AAP online support](#) page. Also, contact your instructor at the email address listed in the syllabus.

Copyright Policy

All course material are the property of JHU and are to be used for the student's individual academic purpose only. Any dissemination, copying, reproducing, modification, displaying, or transmitting of any course material content for any other purpose is prohibited, will be considered misconduct under the [JHU Copyright Compliance Policy](#), and may be cause for disciplinary action. In addition, encouraging academic dishonesty or cheating by distributing information about course materials or assignments which would give an unfair advantage to others may violate [AAP's Code of Conduct](#) and the University's [Student Conduct Code](#). Specifically, recordings, course materials, and lecture notes may not be exchanged or distributed for commercial purposes, for compensation, or for any purpose other than use by students enrolled in the class. Other distributions of such materials by students may be deemed to violate the above University policies and be subject to disciplinary action.

Code of Conduct

To better support all students, the Johns Hopkins University non-academic [Student Conduct Code](#) has been integrated and updated to include all divisions of the University. In addition, it is important to note that all AAP students are still accountable for the [Code of Conduct](#) for Advanced Academic Programs.

Title IX Confidentiality and Mandatory Reporting

As an instructor, one of my responsibilities is to help create a safe and inclusive learning

environment on our campus. I also have mandatory reporting responsibilities related to my role as a Responsible Employee under the Sexual Misconduct Policy & Procedures (which prohibits sexual harassment, sexual assault, relationship violence and stalking), as well as the General Anti-Harassment Policy (which prohibits all types of protected status based discrimination and harassment). It is my goal that you feel able to share information related to your life experiences in classroom discussions, in your written work, and in our one-on-one meetings. I will seek to keep information you share private to the greatest extent possible. However, I am required to share information that I learn of regarding sexual misconduct, as well as protected status based harassment and discrimination, with the Office of Institutional Equity (OIE). For a list of individuals/offices who can speak with you confidentially, please see Appendix B of the [JHU Sexual Misconduct Policies and Laws](#).

For more information on both policies mentioned above, please see: [JHU Relevant Policies](#), Codes, Statements and Principles. Please also note that certain faculty and other University community members also have a duty as a designated Campus Safety Authority under the Clery Act to notify campus security of certain crimes, as well as a duty under State law and University policy to report suspected child abuse and/or neglect.