NYU Paris

CSCI-UA 9473,

Introduction to Machine Learning

Instruction Mode: Hybrid

Time Zone: (UTC+1)

SPRING 2021

We know that you may be taking courses at multiple locations this semester. If you are enrolled in this course 100% remotely and are not a Go Local/Study Away student for this course site, please make sure that you've completed the online academic orientation via NYU Classes so you are aware of site specific support structure, policies and procedures. **Please contact the site academic staff (**Marion Aller **ma5461@nyu.edu)** if you have trouble accessing the NYU Classes site.

If you are attending in person, you will be assigned a seat on the first day and are expected to use that seat for the entire semester due to NYU COVID-19 safety protocol.

Instructor Information

- Augustin Cosse,
- Office hours: NYU Paris, Thursday, 2.15pm - 3.45pm (UTC+1)
- Main Office: 57 Boulevard Saint-Germain, 75005 Paris, France

Course Description

- CSCI-UA 9473
- Introduction to Machine Learning
- Course website : http://www.augustincosse.com/teaching
- Class meeting days and times:
    - Meeting days: (Lectures) UTC+1 Tuesday/Thursday 9.00am - 10.15am
      (Recitations) UTC+1 Thursday 10.30am - 12.00pm
      A Zoom link will be sent the day before each lecture
    - Classroom number: 410

  - Classroom building: NYU Paris, 57 Boulevard Saint-Germain, 75005 Paris

- Exam schedule (to be confirmed) : Midterm: Around March 25, Final: Around May 12

- Course description: Machine Learning is getting more and more important these days with applications ranging from autonomous driving to computer assisted medicine, including weather or financial forecasting. In this class we will study the mathematical foundations of the current machine learning algorithms. We will cover the main models from both supervised learning including linear and nonlinear regression and classification (kernel methods, support vector machine, neural networks) and unsupervised learning (including clustering, gaussian mixtures, self organizing maps, principal and independent component analysis and non linear dimensionality reduction) We will review basic concepts in probability and statistics. We will discuss Bayesian vs frequentist statistics and model/parameter inference, as well as sampling methods. Finally, we will also discuss the important question of model assessment and selection. The class will alternate between Lab sessions and lectures. Students also engage in a semester long project of their choice. A poster session is usually scheduled at the end of the semester (see for example https://twitter.com/augustinmcosse?lang=en as well as http://www.augustincosse.com/mlspring2019)

- Prerequisite
  - MATH-UA Calculus I
  - CSCI-UA 0201 Computer Systems Organisation OR Computer Architecture: CS-UY 2214 or CENG-SHU 202
  - Linear Algebra: MATH-UA 140 or MATH-UH 1022 or MATH-SHU 140 or MATH-UY 2034 OR
  - Fundamentals of Linear Algebra: MATH-UH 1023 or MATH-UY 2034 or MATH-SHU 265
- Pre-requisite or co-requisite
  - CSCI-UA 310 basic Algorithms
- Recommended:
  - MATH-UA 235 Probability and Statistics
  - MATH-UA 234 Mathematical Statistics

Course Overview and Goals

See above.

Students who complete this course successfully will be able to:

- Code basic machine learning algorithms in Python using libraries/toolkits such as

- NumPy
- Matplotlib
- Panda
- Scikit-learn
- Keras (TensorFlow)
- OpenAI Gym
- ...

- Use and write Jupyter notebooks
- Use version control through Github.
- Download and use Popular machine learning datasets as well as new original data from various sources (consumer and financial websites, data spreadsheets,… and online public data platforms such as Kaggle)
- Understand the main challenges in machine learning including the curse of dimensionality as well as regularization, the difficulty of (large scale) non convex parameter estimation problems, the relation between training and generalization…
- Understand the distinction between supervised and unsupervised learning, as well the the interest and difficulties of both approaches.
- Understand and apply the main statistical tools enabling parameter estimation including Maximum Likelihood and Maximum a Posteriori.
- Understand and use the major algorithms from supervised learning such as linear regression and classification as well as their extension to non linear problems through Kernel methods.
- Understand and design Neural networks (including backpropagatio) through both Scikit-learn and Keras. Understand the relation between training and generalization as well as the relation between network size and approximation error (i.e. Universal approximation Theorem).
- Use neural networks on simple deep learning tasks such as image classification/recognition.
- Understand and implement the main algorithms from unsupervised Learning including clustering algorithms (K-means, K-medoid, EM), latent linear and non linear variable models, Gaussian mixture models,…
- Understand some of the most recent ideas from deep learning (such as Generative Adversarial Networks and Variational AutoEncoders) and some of the associated open questions.
- Understand the main ideas used in reinforcement learning (RL) and implement some basic RL algorithms using the Gym toolkit from openAI.
- Conduct independent ML related research. Design, present and defend some scientific work.

Course Requirements

Class Participation

You are expected to attend class in person or remote synchronously. Your active participation in class and attendance will be reflected in this part of the course requirements. The camera should be turned on for the duration of the course.

Class Participation

Students are always encouraged to ask questions and to take an active role in class activities and discussions. They are also strongly encouraged to attend office hours, especially when they feel that they are lacking a clear understanding of some of the concepts covered during the lectures.

Assignments

There should be a number of assignments following the main chapters covered during the lectures. The assignments will be given at the end of the lectures, or at the end of the week, and will be collected about 2 weeks after the start date. Assignments will either require to apply the material covered during previous lectures and programming sessions or to read and apply new material in preparation for future lectures.

Personal project

Students will be asked to work on a personal project, investigating a question of their choice in machine learning. They will be asked to implement a related algorithm and/or to summarize the related scientific literature. Finally they will be asked to design a poster summarizing their work and will get the opportunity to present and defend their work during a dedicated poster session towards the end of the semester.

Assigned Readings

- All the information will be contained in the notes which can be downloaded from the website http://www.augustincosse.com/mlspring2021
- Examples of assignments and programming sessions can be found on GitHub: https://github.com/acosse
- Relevant papers and/or book chapters available free of charge on the internet might also be assigned throughout the semester

Grading of Assignments

The grade for this course will be determined according to the following formula:

Assignments/Activities	% of Final Grade
Assignments	30%
MidTerm	30%

Final	30%
End of Semester Project	10%

Letter Grades

Letter grades for the entire course will be assigned as follows:

Letter Grade	Points	Description
A	16-20	Outstanding
A-	15	Excellent
B+	14	Very Good
B	13	Good
B-	12	Satisfactory
C+	11	Above Average
C	10	Average
C-	9	Below Average
D+	8	Unsatisfactory
D	7	Low Pass
D-	6	Low Pass
F	5	Fail

View Grades

Grades will be available on the NYU Classes site.

Course Schedule

The detailed schedule will be updated regularly (depending on the progress made) on http://www.augustincosse.com/teaching. The notes are available on the same website.

Tentative topics and Assignments

Week/Date	Topic	Reading	Assignment Due
Week 1, 01/26 - 01/28	Introduction, reminders on Probability, Multivariate differential calc., Inference and decision theory	Sections 1 to 4 in the notes	
Supervised			
Week 2, 02/02 - 02/04	Linear and Logistic Regression, regularization Compressed sensing, Linear Classification	Sections 5, 7, 8	Assignment 1 start date
Week 3, 02/09 - 02/11	Lab I : Introduction to Python, linear Classification and Linear Regression		
Week 4, 02/16 - 02/18	Non linear classification, Kernel methods SVMs	Sections 7, 10, 11	Assignment 1 due date Assignment 2 start date
Week 5, 02/23 - 02/25	Neural Networks Optimization, Stochastic Optimization	Section 12	

	Deep Learning		
Week 6, 03/02 - 03/04	Lab 2: Non Linear Regression and Classification + Deep Learning		Assignment 2 due date, Assignment 3 start date
Unsupervised			
Week 7, 03/09 - 03/11	**MidTerm Exam** Clustering + linear Latent variable models (part I)	Sections 13, 14	Assignment 3 due date, Assignment 4 start date **Project Final choice**
Week 8, 03/16 - 03/18	Linear latent variable models (part II), PCA, ICA, GMM, EM Algorithm, Non Linear Latent Variable models (part I)	Sections 13, 14	
Week 9, 03/23 - 03/25	Non linear latent variable models and manifold learning (part II)	Section 15	Assignment 4 due date. Assignment 5 start date
Week 10, 03/30 - 04/21	Lab 3: Unsupervised Learning	Part II	
Week 11, 04/06 - 04/08	Generalization, Complexity and VC Dimension		Assignment 5 due date.
Week 12, 04/13 - 04/15	Probabilistic models HMM Bayesian networks/ Advanced Topics: Reinforcement Learning	Section 16	
Week 13, 04/20 - 04/22	Lab 4, Exam Review, wrap up		**Posters deadlines**
Week 14, 04/27- 04/29	Projects presentations		

Week 15, 05/04 - 05/06	**Revisions**		
Week 16 05/11 - 05/13	Finals		

Course Material

- See the Assigned Readings. The material for the final will be discussed at the end of the semester depending on the topics that were covered during the lectures and labs. All the topics covered in class will be summarized in the notes http://www.augustincosse.com/mlspring2021

Optional Textbooks & Materials

The books listed below are not required for the class but are listed as additional resources for those who are interested in getting additional details/clarifications with respect to the material covered during the lectures. PDF versions can be found online for most of those books. Versions of those books will be available at the library.

- C. M. Bishop, *Pattern Recognition and Machine Learning,* Springer 2006.
- Suvrit Sra, Sebastian Nowozin, Stephen J. Wright, *Optimization for Machine Learning,* MIT Press, 2012
- M. J. Kearns, M. Vazirani*, An introduction to Computational Learning Theory,* MIT press, 1994
- M. Vidyasagar, *Learning and Generalization, with applications to Neural Networks,* Springer, 2003
- T. Hastie, R. Tibshirani, J Friedman, *The elements of statistical Learning,* Springer Series in Statistics*, 2017*
- K. P . Murphy*, Machine Learning, A probabilistic perspective,* MIT Press 2012.
- I. GoodFellow, Y. Bengio, A. Courville, *Deep Learning,* MIT Press, 2016 *http://www.deeplearningbook.org*

Resources

- **Access your course materials**: NYU Classes (nyu.edu/its/classes)
- **Databases, journal articles, and more**: Bobst Library (library.nyu.edu)
- **Assistance with strengthening your writing**: NYU Writing Center (nyu.mywconline.com)
- **Obtain 24/7 technology assistance:** IT Help Desk (nyu.edu/it/servicedesk)

Course Policies

Hygiene/Physical Distancing policies

- Students will be assigned/choose a seat on the first day of class. For NYU COVID-19 Safety protocols, please use the same seat for the duration of the semester.

Attendance and Tardiness

Studying at Global Academic Centers is an academically intensive and immersive experience, in which students from a wide range of backgrounds exchange ideas in discussion-based seminars. Learning in such an environment depends on the active participation of all students. And since classes typically meet once or twice a week, even a single absence can cause a student to miss a significant portion of a course. **To ensure the integrity of this academic experience, class attendance at the centers or online through NYU Classes if the course is remote synchronous/blended, is expected promptly when class begins. Unexcused absences will affect students' semester participation grade. If you have scheduled a remote course immediately preceding/following an in-person class, you may want to discuss where at the Academic Center the remote course can be taken** Students are responsible for making up any work missed due to absence. Repeated absences in a course may result in failure.

Students are responsible for making up any work missed due to absence. This means they should initiate email and/or office hour discussions to address any missed lectures and assignments and arrange a timeline for submitting missed work.

Classroom Etiquette/Expectations

Things to consider:

- Please be mindful of <u>your</u> microphone and video display during synchronous class meetings. Ambient noise and some visual images may disrupt class time for you and your peers.
- If you are not using your cell phone to follow the lesson, cell phones should be turned off or in silent mode during class time.
- Make sure to let your classmates finish speaking before you do.
- Please do not eat during class and minimize any other distracting noises (e.g. rustling of papers and leaving the classroom before the break, unless absolutely necessary)
- If deemed necessary by the study away site (ie COVID related need), synchronous class sessions may be recorded and archived for other students to view. This will be announced at the beginning of class time.

- Students should be respectful and courteous at all times to all participants in class. Consider using the chat function or “raise hand” function in order to add your voice to class discussions especially if leaving the video on presents challenges.

Final Exams

Final exams must be taken at their designated times. Should there be a conflict between final exams, please bring it to the attention of the site Academic representative as soon as this is known to facilitate alternate arrangements. Final exams may not be taken early, and students should not plan to leave the site before the end of the finals period.

Late Assignment

1) Written work due in class must be submitted during the class time to the professor.
2) Late work should be emailed to the faculty as soon as it is completed.
3) Students who arrive to class late for an exam do not have automatic approval to take extra time to complete the exam.
4) Students who miss an exam (including the final) without previously arranged permission will receive a zero on that exam.
5) Assignments due during finals week that are submitted more than 3 days without previously arranged extensions will not be accepted and will receive a zero. Any exceptions or extensions for work during finals week must be discussed with the Site Director.

Incomplete Grade Policy

An “incomplete” is a temporary grade that indicates that the student has, for good reason, not completed all of the course work. This grade is not awarded automatically nor is it guaranteed; rather, the student must ask the instructor for a grade of “incomplete,” present documented evidence of illness, an emergency, or other compelling circumstances, and clarify the remaining course requirements with the instructor.

In order for a grade of “incomplete” to be registered on the transcript, the student must fill out a form, in collaboration with the course instructor and the academic administration at the site; it should then be submitted to the site’s academic office. The submitted form must include a deadline by which the missing work will be completed. This deadline may not be later than the end of the following semester.

Academic Honesty/Plagiarism

As the University's policy on "Academic Integrity for Students at NYU" states: "At NYU, a commitment to excellence, fairness, honesty, and respect within and outside the classroom is essential to maintaining the integrity of our community. By accepting membership in this community, students take responsibility for demonstrating these values in their own conduct and for recognizing and supporting these values in others." **Students at Global Academic Centers must follow the University and school policies.**

The presentation of another person’s words, ideas, judgment, images, or data as though they were your own, whether intentionally or unintentionally, constitutes an act of plagiarism.

NYU X takes plagiarism very seriously; penalties follow and may exceed those set out by your home school. All your written work must be submitted as a hard copy AND in electronic form to the lecturer. Your lecturer may ask you to sign a declaration of authorship form.

It is also an offense to submit work for assignments from two different courses that is substantially the same (be it oral presentations or written work). If there is an overlap of the subject of your assignment with one that you produced for another course (either in the current or any previous semester), you MUST inform your professor.

For guidelines on academic honesty, clarification of the definition of plagiarism, examples of procedures and sanctions, and resources to support proper citation, please see:

NYU Academic Integrity Policies and Guidelines

NYU Library Guides

Religious Observances

Students observing a religious holiday during regularly scheduled class time are entitled to miss class without any penalty to their grade. This is for the holiday only and does not include the days of travel that may come before and/or after the holiday.

Students must notify their professor and the local Academics team in writing via email at least 7 days before being absent for this purpose.

Inclusion, Diversity, Belonging and Equity

NYU is committed to building a culture that respects and embraces diversity, inclusion, and equity, believing that these values – in all their facets – are, as President Andrew Hamilton has said, "...not only important to cherish for their own sake, but because they are also vital for advancing knowledge, sparking innovation, and creating sustainable communities." At NYU PARIS, we are committed to creating a learning environment that:

• fosters intellectual inquiry, research, and artistic practices that respectfully and rigorously take account of a wide range of opinions, perspectives, and experiences; and

• promotes an inclusive community in which diversity is valued and every member feels they have a rightful place, is welcome and respected, and is supported in their endeavours.

Moses Accommodations Statement

Academic accommodations are available for students with documented and registered disabilities. Please contact the Moses Center for Student Accessibility (+1 212-998-4980 or mosescsd@nyu.edu) for further information. Students who are requesting academic accommodations are advised to reach out to the Moses Center **as early as possible in the semester for assistance**. Accommodations for this course are managed through the site sponsoring the class once you request it.

Instructor Bio/About Your Instructor

Augustin Cosse received the BS and MS in Engineering and applied mathematics at the University of Louvain, Belgium in 2009 and 2011 respectively. He then obtained a PhD in applied mathematics and electrical engineering which was funded by the Belgian National Science Foundation (FNRS). He was a visiting student at MIT between 2013 and 2014, a visiting fellow at Harvard (IACS) between 2014 and 2015 and visited the University of Chicago, Galton school of Statistics, between 2015 and 2016. He completed a one year Postdoc at the Courant Institute and Center for Data Science in NYC before joining the department of Mathematics at Ecole Normale Superieure, Paris, as a postdoctoral researcher in 2017. In parallel, he joined NYU Paris as a lecturer in Computer Science in 2018.

His research interests include applied analysis and inverse problems, convex optimization, high dimensional probability and statistics, machine learning and theoretical computer science.