NYU Paris
CSCI-UA 9473,
Introduction to Machine Learning

Instructor Information
- Augustin Cosse,
- augustin.cosse@ens.fr, acosse@nyu.edu
- Ecole Normale Supérieure, Département de Mathématiques et Applications,
  45 Rue d’Ulm, Paris, FR.
  C16 / Espace Cartan, ground floor

Course Information
- CSCI-UA 9473
- Introduction to Machine Learning

Course description: Machine learning is an exciting and fast-moving field of computer science with many recent consumer applications (e.g., Microsoft Kinect, Google Translate, Iphone’s Siri, digital camera face detection, Netflix recommendations, Google news) and applications within the sciences and medicine (e.g., predicting protein-protein interactions, species modeling, detecting tumors, personalized medicine). This course introduces undergraduate computer science students to the field of machine learning. Students learn about the theoretical foundations of machine learning and how to apply machine learning to solve new problems. Assuming no prior knowledge in machine learning, the course focuses on two major paradigms in machine learning which are supervised and unsupervised learning. In supervised learning, we learn various methods for classification and regression. Dimensionality reduction and clustering are discussed in the case of unsupervised learning.

The class will alternate between Lab sessions and lectures and an independent research project will be conducted by each student on a topic of his/her choice towards the end of the semester.

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

Class meeting days and times:
- Meeting days: Tuesday/Thursday 5.15pm - 6.45pm
- Class room number: To be determined
- Class room building: NYU Paris, 57 Boulevard Saint-Germain, 75005 Paris

Course Overview and Goals

Upon Completion of this Course, students will be able to:

- Program machine learning algorithms in python using libraries such as
  - Scikit-learn
  - NumPy
  - Matplotlib
  - Panda
  - ...

- Download and use Popular machine learning datasets as well as new original data from various sources (consumer and financial websites, data spreadsheets,...)

- Understand the various challenges of machine learning including the curse of dimensionality as well as regularisation, the difficulty of (large scale) non convex parameter estimation problems, the relation between training and generalisation...

- Understand the distinction between supervised and unsupervised learning, as well the the interests and difficulties of both approaches.

- Understand and apply the major statistical tools enabling parameter estimation including Maximum Likelihood Estimator 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 an implement Neural networks (including backpropagation). Understand the relation between training and generalisation as well as the relation between size and function approximation.

- Understand and implement the main algorithms from unsupervised Learning including clustering algorithms, latent linear and non linear variable models, Gaussian mixtures,...

- Understand the most recent ideas from deep learning and some of the associated open questions.

- Conduct research on a personal project related to machine learning and implement an innovative new approach at solving question related to this problem

Course Requirements
Class Participation

Students are encouraged to ask questions both during the class and during office hours.

Assignments

There will be 4 assignments in total after each main chapter of the class. The assignments will be given at the end of the classes and will be collected about 2 weeks after the start date.

Personal project

Students will be given the opportunity to work on a personal project, studying a specific question in machine learning. They will be asked to implement an algorithm solving a question related to the project of their choice and/or to summarize the latest scientific results relating to that project.

Assigned Readings

- All the information will be contained in the notes which can be downloaded from the website of the class http://www.augustincosse.com/ml2018
- The Python examples from the notes can be found on GitHub https://acosse.github.io/IntroMLFall2018/
- The exercises for the programming labs will also be uploaded as python notebooks on the GitHub webpage https://acosse.github.io/IntroMLFall2018/
- Interesting 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:

<table>
<thead>
<tr>
<th>Assignments/Activities</th>
<th>% of Final Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignments</td>
<td>30%</td>
</tr>
<tr>
<td>Final Exam</td>
<td>60%</td>
</tr>
<tr>
<td>End of Semester Project</td>
<td>10%</td>
</tr>
</tbody>
</table>

Letter Grades

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

<table>
<thead>
<tr>
<th>Letter Grade</th>
<th>Points</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>16-20</td>
<td>Outstanding</td>
</tr>
</tbody>
</table>
The detailed schedule will be updated regularly (depending on the progress made) on http://www.augustincosse.com/ml2018. The password will be given during the first class.

**Tentative topics and Assignments**

<table>
<thead>
<tr>
<th>Week/Date</th>
<th>Topic</th>
<th>Reading</th>
<th>Assignment Start/ Due dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1, 09/03 - 09/09</td>
<td>Introduction, reminders on Probability, Inference and decision theory</td>
<td>Chapter 1, 2, 3</td>
<td></td>
</tr>
<tr>
<td>Week/Date</td>
<td>Topic</td>
<td>Reading</td>
<td>Assignment Start/Due dates</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------------------------------------------------------</td>
<td>------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td><strong>Supervised</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week 2, 09/10 - 09/16</strong></td>
<td>Linear and Logistic Regression, regularization Compressed sensing, Linear Classification</td>
<td>Chapter 6, 7</td>
<td>Assignment 1 start date</td>
</tr>
<tr>
<td><strong>Week 3, 09/17 - 09/23</strong></td>
<td>Lab I : Introduction to Python, linear Classification and Linear Regression</td>
<td>Part Ia</td>
<td></td>
</tr>
<tr>
<td><strong>Week 4, 09/24 - 09/30</strong></td>
<td>Non linear classification, Kernel methods SVMs</td>
<td>Chapter 8</td>
<td>Assignment 1 due date</td>
</tr>
<tr>
<td><strong>Week 5, 10/01 - 10/07</strong></td>
<td>Neural Networks Optimization, Stochastic Optimization Deep Learning</td>
<td>Chapter 9, 4, 18</td>
<td>Assignment 2 start date</td>
</tr>
<tr>
<td><strong>Week 6, 10/08 - 10/14</strong></td>
<td>Lab 2: Non Linear Regression and Classification + Deep Learning</td>
<td>Part Ib</td>
<td>Assignment 2 due date, Assignment 3 start date</td>
</tr>
<tr>
<td><strong>Unsupervised</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week 7, 10/15 - 10/21</strong></td>
<td>Clustering + linear Latent variable models (part I)</td>
<td>Chapter 11, 12</td>
<td></td>
</tr>
<tr>
<td><strong>Week 8, 10/22 - 10/28</strong></td>
<td>Linear latent variable models (part II), PCA, ICA, GMM, EM Algorithm, Non Linear Latent Variable models (part I)</td>
<td>Chapter 12, 13</td>
<td>Independent project choice</td>
</tr>
<tr>
<td><strong>Week 9, 10/29 - 11/4</strong></td>
<td>Non linear latent variable models and manifold learning (part II)</td>
<td>Chapter 13</td>
<td>Assignment 3 due date, Assignment 4 start date</td>
</tr>
</tbody>
</table>
Course Materials

- See the Assigned Readings. The material for the finals 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/ml2018](http://www.augustincosse.com/ml2018)

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 pursuing a career in data science and machine learning (either in academia or in industry)

- M. Vidyasagar, *Learning and Generalization, with applications to Neural Networks*, Springer, 2003

<table>
<thead>
<tr>
<th>Week/Date</th>
<th>Topic</th>
<th>Reading</th>
<th>Assignment Start/ Due dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 10, 11/4 - 11/11</td>
<td>Lab 3: Unsupervised Learning</td>
<td>Part II</td>
<td></td>
</tr>
<tr>
<td>Week 11, 11/12 - 11/18</td>
<td>Generalization, Complexity and VC Dimension</td>
<td>Chapter 16</td>
<td></td>
</tr>
<tr>
<td>Week 12, 11/19 - 11/25</td>
<td>Probabilistic models</td>
<td>Chapters 14, Part V</td>
<td>Assignment 4 due date</td>
</tr>
<tr>
<td>Week 13, 11/26 - 12/5</td>
<td>Lab 4, Exam Review, wrap up</td>
<td></td>
<td>Project due date</td>
</tr>
<tr>
<td>Week 14, 12/6 - 12/13</td>
<td>Final Exam</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

**Attendance and Tardiness**

- Study abroad 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 is mandatory, and unexcused absences will be penalized with a two percent deduction from the student’s final course grade for every week’s worth of classes missed. Students are responsible for making up any work missed due to absence. Repeated absences in a course may result in harsher penalties including failure.

- Unexcused absences affect students’ grades: unexcused absences will be penalized with a 2% deduction from the students’ final course grade.

- Absences are excused only for illness, religious observance, and emergencies.

**Illness:** For a single absence, students may be required to provide a doctor’s note, at the discretion of the Associate Director of Academics. In the case of two consecutive absences, students must provide a doctor’s note. Exams, quizzes, and presentations will not be made up without a doctor’s note.

**Religious Observance:** 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 instructor and the Academic Office in writing via email one week in advance before being absent for this purpose. If exams, quizzes, and presentations are scheduled on a holiday a student will observe, the Associate Director, in coordination with the instructor, will reschedule them.

**Please note:** If you are unable to attend class, you are required to email your professors directly to notify them.
Late Assignment

Late submission or work will be accepted only with justifiable reasons of health or family emergency.

Academic Honesty/Plagiarism

At NYU, a commitment to excellence, fairness, honesty, and respect within and outside the classroom is essential to maintaining the integrity of our community.

Plagiarism: presenting others' work without adequate acknowledgement of its source, as though it were one’s own. Plagiarism is a form of fraud. We all stand on the shoulders of others, and we must give credit to the creators of the works that we incorporate into products that we call our own. Some examples of plagiarism:

- a sequence of words incorporated without quotation marks
- an unacknowledged passage paraphrased from another's work
- the use of ideas, sound recordings, computer data or images created by others as though it were one’s own
- submitting evaluations of group members’ work for an assigned group project which misrepresent the work that was performed by another group member
- altering or forging academic documents, including but not limited to admissions materials, academic records, grade reports, add/drop forms, course registration forms, etc.

For further information, students are encouraged to check www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/academic-integrity-for-students-at-nyu.html

Disability Disclosure Statement

Academic accommodations are available for students with disabilities. Please contact the Moses Center for Students with Disabilities (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.

Instructor Bio

Augustin Cosse received the BS and MS in Engineering and applied mathematics at the University of Louvain, Belgium in 2009 and 2011. He then obtained a PhD in applied math 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 NY before joining the department of Mathematics at Ecole Normale Superieure, Paris, in 2018.
His research interests include applied analysis as well as inverse problems, convex optimization, high dimensional probability and statistics and machine learning.