**PHYS3600ID – Machine Learning for Physics and Astronomy**

**Fall 2018**

**Instructor**

Dr. Viviana Acquaviva (Physics)

[vacquaviva@citytech.cuny.edu](mailto:vacquaviva@citytech.cuny.edu)

**Guest Lecturers**

Mr. Kartheik Iyer, Physics PhD candidate, Rutgers, the State University of New Jersey

Mr. Andy Lawler, Data Scientist and Statistics PhD candidate, Baylor University

Dr. Ashwin Satyanarayana, Assistant Professor, CST department, City Tech

**Course philosophy**

This course is very hands-on and we assume that you are here because you have a genuine interest in machine learning methods and in understanding how they are applied to real-world research problems. We will guide you and we are happy to help but we also expect you to be independent thinkers and to take initiative – make a Stack Overflow account and don’t be afraid to ask questions in class, to your fellow students and online!

**Class/assignment rules**

We encourage you to talk to each other in class and beyond but your assignments need to be the result of your own work. Identical or very similar assignments are not acceptable. This is valid also for longer coding assignments and reports. **Using online sources as inspiration for coding assignments is allowed but sources should be cited.** **Using large chunks of text from outside sources in reports is not allowed and will be considered plagiarism.**

We expect students to attend class. If you miss several classes, your grade will suffer. If you have extenuating circumstances, please talk to us and we’ll figure out something.

**Class material**

We will distribute course material (lecture slides, notebooks, worksheets) on an ongoing basis using Blackboard. You will also be asked to upload homework (worksheets) and reports to Blackboard.

**Resources**

This course uses Open Education Resources. Rather than a proper textbook, since the material is new and ever-changing, we have a collection of useful resources.

**On programming:**

These are good introductory lectures and we will use some of this material in the first few weeks:

https://github.com/jrjohansson/scientific-python-lectures

https://github.com/jakevdp/WhirlwindTourOfPython/

**On Data Science:**

<https://github.com/jakevdp/PythonDataScienceHandbook>

(all content is available free of charge in the form of Python notebooks, and the book can be purchased on Amazon).

**On practical implementation of Machine Learning algorithms in Python:**

<http://scikit-learn.org/stable/user_guide.html>

**Software**

We used Jupyter notebooks, which require Python, and we have them installed in the computer lab, but if you would like to install them on your personal machine, you can find easy-to-follow instructions here:

<http://jupyter.readthedocs.io/en/latest/install.html>

(refer to “Instally Jupyter with Anaconda” section)

**Grading Policy**

The course will be graded according to: weekly worksheets (30%), three special homework sets (reports) (25%), two in-class quizzes (20%), and final project (25%).

**Class Schedule**

| Week | Topic | Tools introduced | HW due  **Worksheets due weekly starting from week 2** |
| --- | --- | --- | --- |
| 1 | Introduction; course overview; first steps in Python | Basic Python commands and tutorials; Jupyter notebooks |  |
| 2 | Basic statistics and programming recap | Numpy |  |
| 3 | Plotting; Fitting data and building models I | Package: Matplotlib. |  |
| 4 | Fitting data and building models II; Dark Energy example |  |  |
| 5 | Introduction to machine learning terminology and concepts; decision trees | Package: Scikit-learn | HW due: Special report essay on the accelerated expansion of the Universe |
| 6 | Evaluating performance; Diagnostics;  Learning curves;  SVMs |  |  |
| 7 | Parameter optimization; Cross-validation |  |  |
| 8 | The Higgs boson discovery example; Review |  |  |
| QUIZ |  |  |  |
| 10 | Linear and logistic regression |  | HW due: Worksheet/report on particle physics data challenge |
| 11  Guest lecture  (Mr Kartheik Iyer) | Advanced machine learning algorithms: Gaussian Processes |  |  |
| 12  Guest lecture (Dr. Ashwin Satyanarayana) | Unsupervised learning algorithms; Clustering; Dimensionality reduction (PCA) |  | Final project topics due |
| 13 | Ensemble algorithms; algorithm comparison |  |  |
| 14  Guest lecture (Mr. Andy Lawler) | Machine learning applications outside academia: the insurance industry |  | HW due: Worksheet/report on guest lectures topics;  Final projects outlines due |
| 15 | Quiz; Final project presentations |  |  |