ESE-670: Topics in Electrical Sciences  
Topic: Data Science and Machine Learning in MOOCs  
(Massive Open Online Courses)  
Spring 2019  
Syllabus (Final Version)

1. Instructor  
Wendy Tang, wendy.tang@stonybrook.edu,  
Office Hours  
Tue: 2-4 pm in Light Engineering Rm 239, (631) 632-8404  
Thr: 2-4 pm in Admin 407 (631) 632-7308  

*Hours may change. Please check Blackboard for most up-to-date information.*

2. Course Description  
This is a seminar course in which we explore how to use MOOCs (Massive Open Online Courses) to learn data science and machine learning skills. The Class will go over Introduction to Data Science with Python presented by Dr. Christopher Brook at the University of Michigan. There will be individual / group explorations of other MOOC courses in Machine Learning in the Coursera Platform. A project where critical thinking skills coupled with data sciences and machine learning skills is included.

3. Objectives  
The course intends to give students a broad understanding and exposure of the latest development of MOOCs in data science, machine learning, and other technical areas. Students will learn data sciences and machine learning skills as well as how to accomplish life-long learning via open resources such as MOOCs. The course is a seminar course. Hence, it is intended to be flexible and discussion based rather than lecture oriented. For the most parts, students learned the technical content from watching videos of MOOCs, assignments, and through discussions in class and online. The course also provides an opportunity to explore group/collaborative learning online.
4. **Class Meeting**

The course is offered in hybrid mode. Students watch videos on the MOOC platform asynchronously. In some of the weeks, we’ll have **face-to-face meetings** in class (Tuesday, 4-5:20 pm in Humanities 3016) and in other weeks, **class meetings are replaced by asynchronous online discussions**. For a list of weekly meeting modes, please see Schedule (item 5).

5. **Schedule**

<table>
<thead>
<tr>
<th>Week</th>
<th>Meeting Mode</th>
<th>Topics</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1- Jan 29</td>
<td>In-Class</td>
<td>Introduction to Python</td>
<td>Week 1 of C1</td>
</tr>
<tr>
<td>Week 2- Feb 5</td>
<td>Online</td>
<td>Data Cleaning &amp; Processing</td>
<td>Week 2 of C1</td>
</tr>
<tr>
<td>Week 3-Feb 12</td>
<td>In-Class</td>
<td>Data Cleaning &amp; Processing</td>
<td>Week 3 of C1</td>
</tr>
<tr>
<td>Week 4-Feb 19</td>
<td>Online</td>
<td>Pandas, Statistical Techniques</td>
<td>Week 4 of C1</td>
</tr>
<tr>
<td>Week 5-Feb 26</td>
<td>In-Class</td>
<td>Statistical Techniques</td>
<td>Explore C2-C7</td>
</tr>
<tr>
<td>Week 6-March 5</td>
<td>Online</td>
<td>Topics in C2-C7</td>
<td>Explore C2-C7</td>
</tr>
<tr>
<td>Week 7-March 12</td>
<td>In-Class</td>
<td>Topics in C2-C7</td>
<td>Explore C2-C7</td>
</tr>
<tr>
<td>Spring Break</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 8-March 26</td>
<td>Online</td>
<td>Project Preparation</td>
<td>Project Preparation</td>
</tr>
<tr>
<td>Week 9-April 2</td>
<td>In-Class</td>
<td>Project Preparation</td>
<td>Project Preparation</td>
</tr>
<tr>
<td>Week 10-April 9</td>
<td>Online</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 11 – April 16</td>
<td>In-Class</td>
<td>Project Presentation</td>
<td></td>
</tr>
<tr>
<td>Week 12 – April 23</td>
<td>Online</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 13 – April 30</td>
<td>In-Class</td>
<td>Project Presentation</td>
<td></td>
</tr>
<tr>
<td>Week 14 – May 7</td>
<td>In-Class</td>
<td>Project Presentation/</td>
<td></td>
</tr>
</tbody>
</table>

6. **Grading**

The grade will be based on participation, assignments, and one project in which a report and oral presentations are expected.

- Discussions (in-class and online) -25%
- Assignments - 25%
- Project (Report & Presentation)-50%
7. Course Materials

Course Materials are videos and resources from the Coursera Platform. Students need to sign up for the Coursera Certificate ($50) for the first course (C1). This cost is equivalent to that of a textbook. For the subsequent courses, students can choose the audit option (no payment). The first course is to learn Python programming. The certificate option allows students to have access to solutions of the graded assignments.

C1: Introduction to Data Science w Python University of Michigan
Week 1: Intro [3 hours]
Week 2: [3 hours]: Data Cleaning and Processing
Week 3: [3 hours]: pandas
Week 4: [6 hours]: Statistical Techniques

C2: Applied Machine Learning in Python University of Michigan
Week 1: Fundamentals of Machine Learning - Intro to SciKit Learn [8 hours]
Week 2: Supervised Machine Learning - Part 1 [9 hours]
Week 3: Evaluation [7 hours]
Week 4: Supervised Machine Learning - Part 2 [10 hours]

C3: Machine Learning Foundations: A Case Study Approach - University of Washington
Week 1: Welcome [2 hours]
Week 2: Regression: Predicting House Prices [2 hours]
Week 3: Classification: Analyzing Sentiment [2 hours]
Week 4: Clustering and Similarity: Retrieving Documents [2 hours]
Week 5: Recommending Products [2 hours]
Week 6: Deep Learning: Searching for Images [2 hours], Closing Remarks [1 hours]

C4: Machine Learning: Regression - University of Washington
Week 1: Welcome [1], Simple Linear Regression [3]
Week 2: Multiple Regression [3]
Week 3: Assessing Performance [2]
Week 4: Ridge Regression [3]
Week 5: Feature Selection & Lasso [3]

C5: Machine Learning w Big Data by UC San Diego
Week 1: Welcome [24 min], Introduction to Machine Learning w Big Data [3 hour]
Week 2: Data Exploration [3 hours], Data Preparation [3 hours]
Week 3: Classification [4 hours]
Week 4: Evaluation of Machine Learning Models [3 hours]
Week 5: Regression, Cluster Analysis, and Association Analysis

C6: Machine Learning by Stanford

Week 1: Introduction, Linear Regression w One Variable
Week 2: Linear Regression w Multiple Variables [Octave/Matlab Tutorial]
Week 3: Logistic Regression [2 hours], Regularization [4 hours]
Week 4: Neural Networks: Representations [5 hours]
Week 5: Neural Networks: Learning [5 hours]
Week 6: Advice for Applying Machine Learning [5 hours]
Machine Learning System Design [1 hour]
Week 7: Support Vector Machines [5 hours]
Week 8: Unsupervised Learning [1 hour], Dimensionality Reduction [4 hours]
Week 9: Anomaly Detection [2 hours], Recommender Systems
Week 10: Large Scale Machine Learning [1 hour]
Week 11: Application Example: Photo OCR [1 hour]

C7: The Data Scientist’s Toolbox

Week1: Intro [2 hours]
Week 2: Installing the Toolbox p1 hour]
Week 3: Conceptual Issue [1 hour]
Week 4: Course Project [2 hours]

8. Learning Outcomes – Upon completion of the course, students are expected to
   • Have a broad understanding of machine learning and data science
   • Be able to learn technical concepts independently via online resources
   • Be able to work collaboratively in an online environment
   • Be able to communicate and learn in an online environment

However, it is NOT the intent of the course to provide a structured comprehensive technical
course in machine learning/data sciences. There are other courses for that purpose. In other
words, the course may not cover all essential areas in machine learning. However, the
course will provide the basic background and students will learn the skills to pursue data
sciences and machine learning skills from online resources. It is the intent of the course to
provide students with life-long learning skills on technical area.

9. Disability

If you have a physical, psychological, medical or learning disability that may impact your
course work, please contact Disability Support Services, 128 ECC Building (631) 632-6748.
They will determine with you what accommodations are necessary and appropriate. All
information and documentation are confidential. Students who require assistance during
emergency evacuation are encouraged to discuss their needs with their professors and
Disability Support Services. For procedures and information, go to the following web site:
http://www.egr.sunysb.edu and search Fire Safety and Evacuation and Disabilities.
10. Academic Honesty
Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Any suspected instance of academic dishonesty will be reported to the Academic Judiciary. For more comprehensive information on academic integrity, including categories of academic dishonesty, please refer to the academic judiciary website at http://www.stonybrook.edu/uaa/academicjudiciary/

11. Conduct
The University at Stony Brook expects students to maintain standards of personal integrity that are in harmony with the educational goals of the institution; to observe national, state, and local laws and University regulations; and to respect the rights, privileges, and property of other people. Faculty are required to report disruptive behavior that interrupts faculty’s ability to teach, the safety of the learning environment, and/or students ability to learn. Judicial Affairs.

Online Communication Guidelines and Learning Resources:
*Maintain Professional Conduct Both in the Classroom and Online:* The classroom is a professional environment where academic debate and learning take place. I will make every effort to make this environment safe for you to share your opinions, ideas, and beliefs. In return, you are expected to respect the opinions, ideas, and beliefs of other students—both in the face-to-face classroom and online communication. Students have the right and privilege to learn in the class, free from harassment and disruption. The course follows the standards set in the Student Code of Conduct, and students are subject to disciplinary action for violation of that code. If your behavior does not follow the course etiquette standards stated below, the grade you receive for a posting may suffer. I reserve the right to remove any discussion messages that display inappropriate language or content.

*Online Post Etiquette:*
- Offensive language or rudeness will not be tolerated. Discuss ideas, not the person.
- Avoid cluttering your messages with excessive emphasis (stars, arrows, exclamations).
- If you are responding to a message, include the relevant part of the original message in your reply, or make sure to refer to the original's contents so as to avoid confusion;
- Be specific and clear, especially when asking questions.
- Use standard punctuation and capitalization. Using all UPPERCASE characters gives the appearance of shouting and makes the message less legible;
- Remember that not all readers have English as their native language, so make allowances for possible misunderstandings and unintended discourtesies.

*My Role as the Instructor in Discussion Boards:* As the instructor, I will serve as a “guide” in terms of the Discussion Board. While I will not respond to every post, I will read what is posted, and reply when necessary. Expect instructor posts in the following situations:
● To assist each of you when it comes to making connections between discussion, lectures, and textbook material.
● To fill in important things that may have been missed.
● To re-direct discussion when it gets “out of hand”.
● To point out key points or to identify valuable posts.

I’ll visit the discussion board at least twice a week on Tuesday and Thursdays. However, I may not be in the discussion board everyday.