Information Sciences and Technology Department

AIT 726: Natural Language Processing
(with Emphasis on Deep Learning)

Course Syllabus
Spring 2021 - Section DL1
Lindi Liao, Ph.D.
dlia2@gmu.edu

For each section, a customized syllabus with information specific to that section will be made available to registered students via the Blackboard Learning System.

Objectives

This is an advanced course in natural language processing (NLP). It builds on the introductory Natural Language Processing class (AIT590/IT491) and goes on to explore deep learning methods in NLP. Topics covered include feedforward neural nets as applied to NLP and deep networks for NLP applications. Class reviews language structure and studies context free grammars, dependency parsing, semantic role labeling along with applications such as sentiment classification and information extraction.

Prerequisites

1) AIT590 or IT491 or already having a similar NLP experience (see AIT590 Syllabus)
2) Proficiency in Python
3) Probability and Statistics
4) College Calculus, Linear Algebra
   You should be comfortable taking (multivariable) derivatives and understanding matrix/vector notation and operations.
5) Highly Desirable: Machine Learning and Algorithm Design

Special Topics Description

This course emphasizes on advanced NLP theory & research as well as industry-ready hands-on technical skills. It includes the following topics but extends their wings of knowledge to some advanced techniques, especially including the latest promising NLP-related research and technologies:

- Deep Learning, Neural Networks, Deep Natural Language Semantics
- Vector Representations, Word Embeddings, Word2Vec, Skip-Gram
- Naïve Bayes, Logistic Regression
- Text Classification, Sentiment Classification, Spam Filtering
- Fake Reviews Detection, Social Media for Public Health Impact
- Feedforward Neural Networks
  - Text Classification, Sentiment classification, Language Modelling
- Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU)
  - Sequence Classification, Sequence labelling, Named Entity Recognition (NER), etc.
- Context-Free Grammar (CFG), Combinatory Categorial Grammar (CCG)
- Dependency Parsing (Transition-Based, Graph-Based), Literature Based Discovery (LBD)
- Semantic Role Labeling (SRL), Shallow Semantic Parsing, Large-Scale QA-SRL Parsing
- Ethic in NLP, Algorithmic Justice, Computational Approaches to Media/Race Bias
- NLTK, PyTorch, TensorFlow, Keras, AWS, Google Colab and other programming tools or platforms
- Advanced NLP Algorithm Design & NLP Project Management

**Course Learning Activities and Grade Distribution**

This course includes programming assignments, lecture reports, paper discussion & presentation, a term project, and online discussion. Programming assignments, lecture reports, and term project need to be done in teams. There is no penalty for late submission of programming assignments before the date as shown in the class schedule on Blackboard.

Your grade will be based on the following breakdown*:

<table>
<thead>
<tr>
<th>Item</th>
<th>Quantity</th>
<th>Percent of Grade</th>
<th>Extra Credit</th>
<th>On-Time Submission (Extra Credit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Participation in Online Activities</td>
<td>∞</td>
<td>8%</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>** Programming Assignments</td>
<td>4</td>
<td>36%</td>
<td>PA1 2pts PA3 2pts</td>
<td>Total 4 points: 1 point each</td>
</tr>
<tr>
<td>** Lecture Reports</td>
<td>4</td>
<td>12%</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>** 1 Team-based Paper Presentation 1 Individual Paper Presentation</td>
<td>2</td>
<td>12%</td>
<td>2 points</td>
<td>No</td>
</tr>
<tr>
<td>** Term Project</td>
<td>1</td>
<td>32%</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>** TOTAL</td>
<td></td>
<td>100% + 10% extra credit = 110%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Subject to revision before and throughout the course.
** Teamwork (up to 4)

**Textbook and Required Materials**

For this course, we will be using two required textbooks.

- **Textbook JM3**
  - *Speech and Language Processing*, by Jurafsky and Martin (3rd edition). Other editions are not a good substitute.
- **Textbook NLP with Python**

**Faculty and Staff**

AIT726, Instructor for DL1: Lindi Liao, Ph.D. Email: dliao2@gmu.edu

AIT726, Course Coordinator: Ozlem Uzuner, Ph.D. Email: ouzuner@gmu.edu

AIT726, GTA for Section DL1: Aishwarya Varala Email: avarala@gmu.edu

**Course Duration**

Dates: GMU Academic Calendar: [https://registrar.gmu.edu/calendars/](https://registrar.gmu.edu/calendars/)

Total Duration: **16 weeks**

**Tentative Course Outline**

See the class schedule on Blackboard for a detailed list of topics that will be covered in this course. There will be readings assigned from the two books and from external sources. Readings will be announced in advance and will need to be completed before the corresponding class meeting.

**Tentative Course Schedule***

<table>
<thead>
<tr>
<th>Week</th>
<th>Main Topic</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction to NLP and deep learning</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Naïve Bayes and Sentiment classification</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Logistic Regression</td>
<td>Lecture report 1 due Assignment 1 out</td>
</tr>
<tr>
<td>4</td>
<td>Vector Representations</td>
<td>Term project topics due</td>
</tr>
<tr>
<td>5</td>
<td>Neural Networks and Language Models</td>
<td>Assignment 1 due Assignment 2 out</td>
</tr>
<tr>
<td>6</td>
<td>Deep Learning Resources</td>
<td>Term project checkpoint 1</td>
</tr>
<tr>
<td>7</td>
<td>Sequence Processing with Recurrent Neural Nets</td>
<td>Assignment 2 due</td>
</tr>
<tr>
<td>8</td>
<td>Formal Grammars of English</td>
<td>Term project checkpoint 2 Assignment 3 out</td>
</tr>
<tr>
<td>9</td>
<td>Dependency Parsing</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Semantic Role Labeling</td>
<td>Term project checkpoint 3 Lecture report 2 due</td>
</tr>
<tr>
<td>11</td>
<td>Language and Vision</td>
<td>Assignment 3 due Assignment 4 out</td>
</tr>
<tr>
<td>12</td>
<td>Literature Based Discovery</td>
<td>Lecture report 3 due</td>
</tr>
</tbody>
</table>
* A **detailed schedule for classes, topics, and assignment due dates will be published on Blackboard.** As many factors may affect the development and progress of a class, the instructor reserves the right to alter the schedule as may be required to assure attainment of course objectives. The schedule is subject to revision before and throughout the course.

*Registered students should see the Blackboard Learning System for the latest class schedule.*

**Class Participation**

This course is designed for mixed "**synchronous**" and "**asynchronous**" delivery (https://masononline.gmu.edu/course-delivery-methods):

- *Synchronous*: Classes are held on a set schedule, and students virtually attend an instructor-led session on a regular basis.
- *Asynchronous*: Students can study at their own pace, accessing instructional materials online in Blackboard at any time, though a set schedule with due dates is still there.

For this course, there are several **important online meetings** ("**Synchronous**" mode) as scheduled on Blackboard Collaborate Ultra. All students are required to virtually attend the instructor-led sessions on a regular basis.

All assignments, assessments, class announcements, schedules, files and presentations will use Blackboard.

Additionally, students need to contribute actively and participate in **online discussions** on Blackboard for grading.

**Communication, Writing and Submissions**

**Communication:** Course announcements will be made through Blackboard.

**Writing:** All discussions, and assignments for this course must be in standard English. Do not use slang or texting abbreviations (i.e., lol). Capitalize and use complete sentences in your discussion responses and in your paper. You can use bulleted lists if they make sense as a way to convey the information. Emoticons are acceptable as long as they are not overused and help with communication.

Before submitting work, be sure to proof read your writing and make sure that any references that you include are correct.

**Submission of Work:** All work for this class must be submitted as the assignment states.
ASA Style Guide: ASA Style Guides are easy to locate using an internet search. The following link is one that should work well for this class, you can access it by clicking here: http://personal.monm.edu/jkessler/ASA-Style.htm.

Academic Honesty
An important component in learning is taking on tasks, assignments and exams in an honest effort to do your best possible work. You are expected to turn in and do original work.

Grading Guidelines and Grade Scale
Grading Guidelines
Some grade components are evaluated subjectively
A: consistently above and beyond the course/assignment requirements
B: meets and occasionally exceeds the course/assignment requirements
C: minimally meets the course/assignment requirements
F: fails to meet the course/assignment requirements

Grades will be awarded in accordance with the Mason Grading System for graduate students. See the university catalog for policies: http://catalog.gmu.edu for more information.

Grading Scale
- The grading scale for this course, is:

  97 – above  A+  Passing
  93 – 96%   A   Passing
  90 – 92%   A-  Passing
  87 – 89%   B+  Passing
  83 – 86%   B   Passing
  77 – 82%   B-  Passing
  70 – 76%   C   Passing
  0 – 69%    F   Failing

NOTE: Study success takes constant effort!
Instructor will double check all students’ coursework graded by GTA at the end of the course.
Raw scores may be adjusted by the Instructor to calculate final grades.

Students are responsible for checking the currency of their grade books. Grade discrepancies must be brought to instructor’s attention within one week of assignment submission and 48 hours of exam submission.

No make-up for any activity, unless arranged in advance. Only in special cases, such as medical problems and family emergency, make-ups and late assignments may be allowed with verifiable proof.

Final grades will be posted to PatriotWeb, which is the only vehicle for students to obtain those grades. A student with a "hold" on his/her PatriotWeb account will be unable to access final grades until the hold has been removed by the Registrar.
Etiquette and Disabilities

Please observe proper “etiquette” and “netiquette” – courteous and appropriate forms of communication and interaction – within this course. This means no personal attacks, obscene language, or intolerant expression. All viewpoints should be respected.

Giving Feedback: This course is designed along the principles of synergy and collaborative learning. Therefore, it is important that all students understand how to provide quality feedback to their peers. Here are a few tips for providing, positive, constructive, and useful feedback to peers.

- Be empathetic and remember that this environment is a safe place for making mistakes
- Use nonjudgmental language and phrases that do not attack an individual. One way of doing this is to ask the individual to discuss his/her process for making the final decision.
- Use specific questions, examples, and references as a way of making your point.
- Make your feedback useful by providing suggestions that the individual can understand and use to improve her/his work.

Disabilities: Please message me if you have a disability so we can discuss ways to help you succeed in the course. If you need accommodations that would affect the terms of this syllabus, you will need to provide documentation of your disability.