INFO5505: APPLIED MACHINE LEARNING FOR DATA SCIENCE

Spring, 2021

<table>
<thead>
<tr>
<th>Instructor:</th>
<th>Junhua Ding, Ph.D.</th>
<th>Time:</th>
<th>Wednesday: 5:30pm - 8:20pm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email:</td>
<td><a href="mailto:junhua.ding@unt.edu">junhua.ding@unt.edu</a></td>
<td>Place:</td>
<td>Zoom meeting/NTDP K150</td>
</tr>
<tr>
<td>Office:</td>
<td>Discovery Park E292G</td>
<td>Phone:</td>
<td>(940)565-2186</td>
</tr>
</tbody>
</table>

Course Pages

1. Course management: Canvas: [https://unt.instructure.com/](https://unt.instructure.com/)
2. Lecture delivering: Zoom: [https://unt.zoom.us](https://unt.zoom.us)

Office Hours:

Textbooks


Prerequisite

1. General programming experience is assumed. The Python programming language will be used for this class and reviewed as appropriate.
2. Basic linear algebra and probability knowledge is needed.

Software

The following software and computing environments might be used in this course.

- TensorFlow, Google Colab, Python Jupyter Notebook, JupyterLab, Python Frameworks such as NumPy, SciPy, Pandas; Microsoft Excel; MySQL.

Important Note

Due to the continuing impact of COVID-19, it is important that you read all Announcements posted on Canvas and UNT Website. While attendance is expected as outlined in this syllabus, it is important for all of us to be mindful of the health and safety of everyone in our community, especially given concerns about COVID-19. Please contact me if you are unable to attend class because you are ill, or unable to attend class due to a related issue regarding COVID-19. It is important that you communicate with me prior to being absent so I may make a decision about accommodating your request to be excused from class. If you are experiencing any symptoms of COVID-19 ([https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html](https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html)) please seek medical attention from the Student Health and Wellness Center (phone number: 940-565-2333 or email: askSHWC@unt.edu) or your health care provider PRIOR to coming to campus. UNT also requires you to contact the UNT COVID Hotline (phone number: 844-366-5892 or email: COVID@unt.edu) for guidance on actions to take due to symptoms, pending or positive test results, or potential exposure. While attendance is an important part of succeeding in this class, your own health, and those of others in the community, is more important. Face coverings are required in all UNT
facilities. Students are expected to wear face coverings during this class. If you are unable to wear a face covering due to a disability, please contact the Office of Disability Access to request an accommodation. UNT face covering requirements are subject to change due to community health guidelines. Any changes will be communicated via the instructor.

Objectives

Machine learning is one of the most fundamental knowledge areas for Data Science. Through a combination of advanced computing techniques, training techniques and architectural components of neural networks, it is now possible to create much more complex machine learning models to learn hierarchies of information in a way that is like the function of the human brain. This course trains data science students basic understanding of modern neural networks and their applications in data science problems including information retrieval, natural language processing, and computer vision. It starts with an introduction of concepts of machine learning and widely adopted machine learning algorithms including linear regression, logistic regression, clustering algorithms, decision trees, naive Bayes Learning, and Support Vector Machine. This course then introduces building blocks of neural networks including fully connected layers, convolutional and recurrent layers. Students will use these building blocks to define complex modern architectures in Google TensorFlow and Keras frameworks using Python programming language. This course finally introduces the applications of machine learning to computer vision with Convolution Neural Networks (CNN), natural language processing with Recurrent Neural Network (RNN) such as Long Short-Term Memory (LSTM), and information retrieval with RNN and CNN.

The objectives of this course:

1. Be able to explain basic concepts and learning process of machine learning models.
2. Be able to use tools to build machine learning applications with machine learning models.
3. Be able to explain neural networks (deep and otherwise) compare to other machine learning models.
4. Be able to select machine learning models and framework to build real-world data science solutions.
5. Be able to select a deep neural network to solve a particular data science problem.
6. Be able to conduct verification, validation, and optimization of machine learning systems.

Topics

The following is the tentative schedule with the covered topics. Actual schedule may be adjusted according to progress:

1. Introduction to Machine Learning
2. Linear Regression with Multiple Variables
3. Logistic Regression
4. Regularization
5. Cluster Analysis
6. Decision Trees
7. Random Forests
8. Naive Bayes Algorithm
9. Support Vector Machine
10. Expectation-Maximization (EM) Algorithm
11. Neural Network: Representation and Learning
12. Backpropagation and Other Training Techniques
14. Convolutional Neural Networks (CNN)
15. Recurrent Neural Network (RNN) and LSTM Network
16. LSTM for Natural Language Processing
17. RNN and CNN for Information Retrieval
18. CNN for Classification of Large Scale Biomedical Image Data

Grading Policy
Students are required to attend the class on time, complete all assignments, quizzes, exams, and readings on time. Grading will be based on assignments to be assigned as the course proceeds.

Grades will be computed as follows:

- 8 Project Assignments (Python programming for different data science tasks): 80%
- 5 Quizzes/1 Comprehensive Exam: 20%

Grading Scale: A: 90-100; B: 80-89; C: 70-79; D: 60-69; F: 59 or below.

All assignments, projects, quizzes and term paper requirements are post in Canvas, and students are required to uploaded their work into Canvas. Late submission of quizzes won’t be graded, but late submission of other work will be graded with reduced points. The final grade is calculated based on grade points of assignments, project, quizzes, and exam.

Incompletes
A grade of incomplete (I) will be given only for a justifiable reason (such as a serious illness or military service) and only if you are passing the course. It is your responsibility to contact the instructor to request an incomplete and discuss requirements for completing the course. If you do not remove the incomplete within the timeframe agreed upon with the instructor or within one calendar year, you will receive a grade of an F. Please refer to http://essc.unt.edu/registrar/academic-record-incomplete.html for more information.

Withdrawal
A grade of withdraw (W) or withdraw-failing (WF) will be given depending on your participation and grades to date. If you simply disappear and do not file a formal UNT withdrawal form, you may receive a grade of an F.

Teaching Philosophy
This course will be taught in traditional lectures although the lectures are delivered via video conference system Zoom. Attending the class on time, and on-class discussions are strongly encouraged. Although software tools such as Python Notebook, and TensorFlow will be needed to complete the assignments, few lectures will be given on how to use the tools.

This course employs lecture capture technology to record class sessions. Students may occasionally appear on video. The lecture recordings will be available to you for study purposes and may also be reused in future course offerings.

Technical Assistance
UIT Help Desk: http://www.unt.edu/helpdesk/index.htm. The University of North Texas provides student technical support in the use of Zoom and Canvas and supported resources. The student help desk may be reached at:

- Email: helpdesk@unt.edu
- Phone: 940.565-2324
- In Person: Sage Hall, Room 130

Hours

- Monday-Thursday 8am-midnight
- Friday 8am-8pm
- Saturday 9am-5p
- Sunday 8am-midnight

Canvas technical requirements: https://clear.unt.edu/supported-technologies/canvas/requirements

Minimal Technical Skills and Resources Needed

Each student should be able to access computers that have environment for running and developing Python programs with databases Microsoft Excel and mySQL. Google Colab (https://colab.research.google.com/) is a recommended environment for developing course projects and running sample codes from the textbook.

Communication

Students can email their questions to the instructor and the teacher assistant (TA). They are also encouraged to talk to the instructor and TA during the office hours. Emails are normally respond within 24 hours, and all assignments, quizzes, projects, and papers should be graded within 10 days after the submission deadline.

ADA Statement

The University of North Texas makes reasonable academic accommodation for students with disabilities. Students seeking reasonable accommodation must first register with the Office of Disability Accommodation (ODA) to verify their eligibility. If a disability is verified, the ODA will provide you with a reasonable accommodation letter to be delivered to faculty to begin a private discussion regarding your specific needs in a course. You may request reasonable accommodations at any time, however, ODA notices of reasonable accommodation should be provided as early as possible in the semester to avoid any delay in implementation. Note that students must obtain a new letter of reasonable accommodation for every semester and must meet with each faculty member prior to implementation in each class. Students are strongly encouraged to deliver letters of reasonable accommodation during faculty office hours or by appointment. Faculty members have the authority to ask students to discuss such letters during their designated office hours to protect the privacy of the student. For additional information see the Office of Disability Accommodation website at http://www.unt.edu/oda. You may also contact them by phone at 940.565.4323.

Class Policy

- Prohibition of Discrimination, Harassment, and Retaliation: As members of the UNT community, we have all made a commitment to be part of an institution that respects and values the identities of the students and employees with whom we interact. UNT does not tolerate identity-based discrimination, harassment, and retaliation. According UNT Policy 16.004, UNT prohibits discrimination and harassment because of race, color, national origin, religion, sex, sexual orientation, gender
identity, gender expression, age, disability, genetic information, veteran status, or any other characteristic protected under applicable federal or state law in its application and admission processes; educational programs and activities; employment policies, procedures, and processes; and university facilities. The University takes active measures to prevent such conduct and investigates and takes remedial action when appropriate.

- **Attendance Policy:** You are expected to attend class via Zoom. You are responsible for announcements and assignments given in class. If you miss a class, it is up to you to obtain notes and any other information that was provided in the class. Those who do not attend class or review the recorded lectures in a timely manner can count on doing poorly in this course.

- **Academic Integrity Policy:** Academic Integrity Standards and Consequences. According to UNT Policy 06.003, Student Academic Integrity, academic dishonesty occurs when students engage in behaviors including, but not limited to cheating, fabrication, facilitating academic dishonesty, forgery, plagiarism, and sabotage. A finding of academic dishonesty may result in a range of academic penalties or sanctions ranging from admonition to expulsion from the University.

  Students caught cheating or plagiarizing will receive a “0” for that particular assignment or exam (or specify alternative sanction, such as course failure). Additionally, the incident will be reported to the Office of Student Rights and Responsibilities, which may impose for further penalty. According to the UNT catalog, the term “cheating” includes, but is not limited to: (a). use of any unauthorized assistance in taking quizzes, tests, or examinations; (b). dependence upon the aid of sources beyond those authorized by the instructor in writing papers, preparing reports, solving problems, or carrying out other assignments; (c). the acquisition, without permission, of tests or other academic material belonging to a faculty or staff member of the university; (d). dual submission of a paper or project, or resubmission of a paper or project to a different class without express permission from the instructor(s); or (e). any other act designed to give a student an unfair advantage. The term “plagiarism” includes, but is not limited to: (a). the knowing or negligent use by paraphrase or direct quotation of the published or unpublished work of another person without full and clear acknowledgment; and (b). the knowing or negligent unacknowledged use of materials prepared by another person or agency engaged in the selling of term papers or other academic materials.

- **ADA Policy:** UNT makes reasonable academic accommodation for students with disabilities. Students seeking accommodation must first register with the Office of Disability Accommodation (ODA) to verify their eligibility. If a disability is verified, the ODA will provide a student with an accommodation letter to be delivered to faculty to begin a private discussion regarding one’s specific course needs. Students may request accommodations at any time, however, ODA notices of accommodation should be provided as early as possible in the semester to avoid any delay in implementation. Note that students must obtain a new letter of accommodation for every semester and must meet with each faculty member prior to implementation in each class. For additional information see the ODA website at disability.unt.edu.

- **Emergency Notification & Procedures:** UNT uses a system called Eagle Alert to quickly notify students with critical information in the event of an emergency (i.e., severe weather, campus closing, and health and public safety emergencies like chemical spills, fires, or violence). In the event of a university closure, please refer to Blackboard for contingency plans for covering course materials.

- **APA Style:** When doing the assignment, it is important to provide details of all the sources of information that you have used to prepare your work. All written assignments should follow APA (American Psychological Association) style to ensure that all sources are cited completely, correctly, and with consistency. The purpose of APA style is to (a). give credit to the author whose ideas or research you have used, (b). provide the exact location for sources of information used in the text of your paper, and (c). enable your reader to verify information you have provided or to explore your topic in greater depth. Consult the Publication Manual of the American Psychological Association, 6th edition.
• **Acceptable Student Behavior:** Student behavior that interferes with an instructor's ability to conduct a class or other students’ opportunity to learn is unacceptable and disruptive and will not be tolerated in any instructional forum at UNT. Students engaging in unacceptable behavior will be directed to leave the classroom and the instructor may refer the student to the Center for Student Rights and Responsibilities to consider whether the student’s conduct violated the Code of Student Conduct. The university’s expectations for student conduct apply to all instructional forums, including university and electronic classroom, labs, discussion groups, field trips, etc. The Code of Student Conduct can be found at deannofstudents.unt.edu/conduct.

• **Access to Information - Eagle Connect:** Students access point for business and academic services at UNT is located at: my.unt.edu. All official communication from the University will be delivered to a student's Eagle Connect account. For more information, please visit the website that explains Eagle Connect and how to forward e-mail: eagleconnect.unt.edu/

• **Student Evaluation Administration Dates:** Student feedback is important and an essential part of participation in this course. The student evaluation of instruction is a requirement for all organized classes at UNT. The survey will be made available during weeks 13, 14 and 15 [insert administration dates] of the long semesters to provide students with an opportunity to evaluate how this course is taught. Students will receive an email from ”UNT SPOT Course Evaluations via IASystem Notification” (no-reply@iasystem.org) with the survey link. Students should look for the email in their UNT email inbox. Simply click on the link and complete the survey. Once students complete the survey they will receive a confirmation email that the survey has been submitted. For additional information, please visit the SPOT website at http://spot.unt.edu/ or email: spot@unt.edu.

• **Sexual Assault Prevention:** UNT is committed to providing a safe learning environment free of all forms of sexual misconduct, including sexual harassment sexual assault, domestic violence, dating violence, and stalking. Federal laws (Title IX and the Violence Against Women Act) and UNT policies prohibit discrimination on the basis of sex, and therefore prohibit sexual misconduct. If you or someone you know is experiencing sexual harassment, relationship violence, stalking, and/or sexual assault, there are campus resources available to provide support and assistance. UNTs Survivor Advocates can assist a student who has been impacted by violence by filing protective orders, completing crime victims compensation applications, contacting professors for absences related to an assault, working with housing to facilitate a room change where appropriate, and connecting students to other resources available both on and off campus. The Survivor Advocates can be reached at SurvivorAdvocate@unt.edu or by calling the Dean of Students Office at 940-565-2648. Additionally, alleged sexual misconduct can be non-confidentially reported to the Title IX Coordinator at oeo@unt.edu or at (940) 565 2759.