HCI 560X - Learn to Speak AI

Summer 2020
Tuesday and Thursday: 2:10 – 3:50 p.m.
All sections taught online.
Iowa State University
Ames, Iowa 50011

Instructor: Vladimir Sukhoy

Office:	Howe Hall, Room 2624E
Phone:	630-445-1332 (email preferred)
Email:	sukhoy@iastate.edu
Web Page:	`http://sukhoy.public.iastate.edu/`
Office Hours:	4 p.m. – 5 p.m. after each lecture, or by appointment

Delivery: online and via recorded videos of lectures.

Course Description: This class covers the main concepts from the design and analysis of algorithms in Artificial Intelligence. During the summer term, we will quickly recap the necessary mathematical knowledge and use it to learn how several popular learning algorithms work. These examples will help illustrate technical issues that specialists in these fields have to address. You will work in small groups to compare the performance of two or more standard learning algorithms using a data set of your choice and describe the results using the material covered in the lectures.

At the end of this class, you will be able to understand technical articles that focus on evaluating standard learning algorithms. You will also have the skills to run standard algorithms on data sets and communicate the results of your experiments to technical audiences that include engineers and computer scientists. Finally, you will acquire an understanding of problems that learning algorithms were designed to solve and will also be able to identify tasks that may not be suitable for these methods.

This course is designed to teach you to be a well-informed user of existing artificial intelligence algorithms. It will not focus on the development of new learning algorithms. There will be no in-depth analysis of convergence, mathematical optimization, choices of programming languages and platforms, etc. The theoretical background will be restricted to the minimum set of concepts that are necessary to understand the problems solved by learning algorithms and the issues that arise due to the need to compute their solutions. The course does require basic computer literacy, i.e., the ability to find information on the Internet, download and install new software, and the ability to work with spreadsheets, e.g., Microsoft Excel.

Topics to be Covered: Discrete and continuous probability distributions. Markov models. n-gram models. Underfitting and overfitting. The curse of dimensionality. The computational complexity of algorithms. Supervised learning. Unsupervised learning. Reinforcement learning. Reproducibility of learning algorithms.

Classifiers: k-nearest-neighbors, artificial neural networks, support vector machines.
Regression models: polynomial curve fitting, k-nearest-neighbors regression.
Clustering algorithms: hierarchical clustering, k-means, and Gaussian mixtures (expectation minimization).
Dimensionality reduction: principal component analysis, feature hashing.
Artificial neural networks: the back-propagation algorithm, deep learning.
Evaluation techniques: precision and recall, receiver operating characteristic (ROC) curves, cross-validation.

Readings: The textbook for this class is "Introduction to Machine Learning" by Ethem Alpaydin, 3rd edition (2nd edition is also OK). In addition to the textbook, the lectures will be based on other sources, most of which can be downloaded from the Internet. All readings outside the required textbooks will be made available on the Canvas page for this course.

Organization and Format*: The course will be held during the 12 weeks of the summer term. It will be organized as a seminar. The expectation is that the students read all material assigned to each lecture before the meeting so that they are ready to discuss it during the meeting. Also, there will be four homework assignments and a course project.

* The instructor can modify any aspect of this class in any way and for whatever reason or for no reason.

Prerequisites: Graduate and advanced undergraduate students can take this graduate class. It is recommended that the students have been exposed to at least 2 of the following fields: linear algebra, statistics, computational perception, human computer interaction, computer vision, signal processing, software engineering, streaming audio or video, and web development.

Programming skills are useful but **not essential**. Data analysis skills with spreadsheet software (e.g., Microsoft Excel) are recommended, but not required because the necessary material will be introduced during the lectures. A subscription to Microsoft Office, which includes Microsoft Excel, is available to all Iowa State students free of charge. Basic computer literacy, i.e., the ability to install and run new software on your machine, download and upload files, etc. is an **essential prerequisite** for this course. Another crucial requirement is your interest in learning algorithms and the desire to understand how they work.

Required Technology:

1. A reliable high-speed Internet connection (for online students) or regular and dependable access to Iowa State University computer system (for on-campus students).

2. For online students only: access to a computer with a microphone and audio capability.

3. All students: reliable and regular access to a machine that can run Parallels Desktop, VirtualBox, VMWare Workstation, VMWare Workstation Player, VMWare Fusion, or another modern x86-64 virtualization platform that supports a 64-bit guest system distributed in the standard OVF format. The virtual machine used in class requires about 1 gigabyte of RAM and 20 gigabytes disk space. These requirements are only for the virtual machine itself. Your system will need additional RAM and disk space for to run everything else. **Most modern desktops and laptops satisfy this requirement.**

   This technology is required for running the implementations of machine learning algorithms were packaged for this course in a Linux virtual machine. This approach enables exposing students with basic computer literacy to learning algorithms without requiring an IT or computer programming background to set up their implementations. The VM will be able to process only relatively small data sets, i.e., no more than several tens of millions of numerical features across all data instances. Its coverage will be restricted to data sets that have a fixed number of numerical attributes. Text, images, and other types of non-numerical data will need to be converted into numerical form to make the data suitable for the course-provided software. Detailed instructions for using the VM will be provided during the first week of the course.

Academic Dishonesty: cheating, plagiarism, and other academic misconducts will not be tolerated and will behandled according to the ISU's academic dishonesty procedures, which are posted here: `http://catalog.iastate.edu/academic_conduct/#academicdishonestytext`

Accommodations: Iowa State University complies with the Americans with Disabilities Act and Sect 504 of the Rehabilitation Act. If you have a disability and anticipate needing accommodations in this course, please contact your instructor to set up a meeting within the first two weeks of the semester or as soon as you become aware of your need. Before meeting with your instructor, you will need to obtain a SAAR form with recommendations for accommodations from the Student Accessibility Services (`http://new.dso.iastate.edu/dr/student`) located in Room 1076 on the main floor of the Student Services Building. Their telephone number is 515-294-7220 or email accessibility@iastate.edu. Retroactive requests for accommodations will not be honored.

Religious Accommodation: If an academic or work requirement conflicts with your religious practices and/or observances, you may request reasonable accommodations. Your request must be in writing, and your instructor or supervisor will review the request. You or your instructor may also seek assistance from the Dean of Students Office (`http://www.dso.iastate.edu/`) or the Office of Equal Opportunity and Compliance (`http://www.eoc.iastate.edu/`).

Dead Week: This class follows the Iowa State University Dead Week policy as noted in section 10.6.4 of the Faculty Handbook (`http://www.provost.iastate.edu/resources/faculty-handbook`).

Harassment and Discrimination: Iowa State University strives to maintain our campus as a place of work and study for faculty, staff, and students that is free of all forms of prohibited discrimination and harassment based upon race, ethnicity, sex (including sexual assault), pregnancy, color, religion, national origin, physical or mental disability, age, marital status, sexual orientation, gender identity, genetic information, or status as a U.S. veteran. Any student who has concerns about such behavior should contact his/her instructor, Student Assistance (`http://www.dso.iastate.edu/sa/`) at 515-294-1020 or email dso-sas@iastate.edu, or the Office of Equal Opportunity and Compliance (`http://www.eoc.iastate.edu/`) at 515-294-7612.

Accessibility Statement: Iowa State University is committed to assuring that all educational activities are free from discrimination and harassment based on disability status. Students requesting accommodations for a documented disability are required to work directly with staff in Student Accessibility Services (SAS) to establish eligibility and learn about related processes before accommodations will be identified. After eligibility is established, SAS staff will create and issue a Notification Letter for each course listing approved reasonable accommodations. This document will be made available to the student and instructor either electronically or in hard-copy every semester. Students and instructors are encouraged to review contents of the Notification Letters as early in the semester as possible to identify a specific, timely plan to deliver/receive the indicated accommodations. Reasonable accommodations are not retroactive in nature and are not intended to be an unfair advantage. Additional information or assistance is available online at `www.sas.dso.iastate.edu`, by contacting SAS staff by email at accessibility@iastate.edu, or by calling 515-294-7220. Student Accessibility Services is a unit in the Dean of Students Office located at 1076 Student Services Building.

Homework Assignments: There will be four homework assignments. You will have between 10 and 14 days to finish each task. Each homework can be completed using only Microsoft Excel and the virtualized software that will be made available for download during the course. The purpose of this software is to expose preinstalled versions of popular machine learning libraries through a unified user interface.

Each of the four homeworks will require downloading CSV files from the course homepage or other web sites, processing them using standard machine learning algorithms (e.g., using the course-provided virtual machine), analyzing the results in Microsoft Excel. For each homework, there will be an HTML template that will need to be filled out to summarize the results of your experiments.

It is possible to complete each homework without using the course-provided software, e.g., by writing programs that solve problems or by using third-party software packages. In these cases, it is expected that a student describes the solution in a way that enables the course instructor to reproduce it without undue experimentation. For example, if a student writes a program that solves a homework problem, then the source code for this program and a brief description of how to run it need to be included in the solution. The grade for the assignment may be affected by the description quality and by the reproducibility of the results.

There will also be a separate introductory homework called 'Homework 0' that will be made available during the first week of the course. It will count for 1% of your grade. Its purpose is to help students become familiar with the software that will be used during the remaining part of the course. That is, homework 0 will require only installing the software and taking its screenshots to get the credit.

Final Project: The final project for this class should use learning algorithms to help solve a practical research or design problem or focus on evaluating two or more learning algorithms on a data set of your choice. You can join a small group (no more than 3 students) or work alone. Joining a group, however, is highly recommended because you can complement your skills with the skills of other students and divide the writing tasks. **Students are encouraged to select their project topics as soon as possible.** A written project proposal (5–10 pages) will be due on 7/7. The final project report (11–15 pages) will be due on 8/6. Each group will have to present the results of their project during the last two lectures in the course.

You are free to explore public sources of data sets, e.g., the Kaggle platform, data.gov, and other sources, to find suitable data sets. Some of them may be too large to be usable with the course-provided VM. In many cases it will be necessary to perform data conversion and preprocessing to make the data suitable for the VM. For the project, you can use any implementation of learning algorithms, i.e., you are not required to use the VM to obtain your results. Please check the data set license before making the final selection.

Collaboration Policy: You are encouraged to form groups and discuss the readings for this class. These groups can be larger than the project groups. You can discuss the homework assignments with other students who take this class. However, each student is expected to solve the homework problems on their own. Sharing of solutions, including, but not limited to, the Excel spreadsheets that analyze the results of learning algorithms, is not permitted.

IMPORTANT: Cheating, plagiarism, and other academic misconducts will not be tolerated and will be handled according to the ISU's academic dishonesty procedures, which are posted here:
`http://catalog.iastate.edu/academic_conduct/#academicdishonestytext`

Class Participation: The expectation is that you attend or view the offline recording of each lecture. If you forget to view a lecture recording, then you may miss what happened, including any announcements.

Asking questions during the live lecture is one way to satisfy the course participation requirement. Another way is to submit discussion posts related to the course topics. There will be six types of discussions: 1) "Introductions", 2) "Learning algorithm or dataset of the day", 3) "Homework assignments and suggestions for new problems", 4) "Possible topics for course projects", 5) "Help Forum", and 6) "Other discussions related to learning algorithms". Your goal is to submit at least one post or an extended response in at least one of these categories every week during the course except week 7 (project proposal presentations) and week 12 (final project presentations). **Multiple posts generated during the same week don't carry over to the next week.** They do generate **brownie points**, however.

Only one post in the "Introductions" board can count toward the participation score. A post counts toward the participation score only if it is not a restatement of already-posted information, is on-topic, and is not too short. A single paragraph that consists of 3–4 sentences or more that contain new information and do not simply state an opinion or express emotions is sufficient to satisfy the class participation requirement for **one week**. Please don't copy text from other sources to generate your posts, because the Academic Dishonesty policy applies to them similarly to all other assignments in this course. The discussion boards will be moderated by the instructor.

Grading: Your grade will be calculated as follows:

Homework Assignments:	56% ($4 \times 14\%$ each)
Final Project Proposal:	10%
Final Project:	18%
Project proposal presentation:	5%
Final project presentation:	5%
Class participation:	5%
Homework 0:	1%

Tentative Schedule

Week	Day/Date	Topic	Assignment
1	Tuesday 5/19	Introduction	Homework 0 out
	Thursday 5/21	K-Means Clustering and Technology Demo	
2	Tuesday 5/26	Probability Distributions and Gaussian Mixtures	Homework 0 due
	Thursday 5/28	TBD	Homework 1 out
3	Tuesday 6/2	KNN Classifier and Evaluation Techniques	Homework 2 out
	Thursday 6/4	Cross-Validation and ROC Curves	
4	Tuesday 6/9	Presentations of Project Ideas	
	Thursday 6/11	Support Vector Machines	Homework 1 due
5	Tuesday 6/16	Hidden Markov Models	Homework 3 out
	Thursday 6/18	Hidden Markov Models	Homework 2 due
6	Tuesday 6/23	Speech Processing and Recognition	
	Thursday 6/25	KNN Regression and Logistic Regression	
7	Tuesday 6/30	Project Proposal Presentations	
	Thursday 7/2	Project Proposal Presentations	Homework 3 due
8	Tuesday 7/7	Principal Component Analysis	Project Proposal due
	Thursday 7/9	Feature Hashing, Reproducibility in AI	Homework 4 out
9	Tuesday 7/14	Artificial Neural Networks: Back-Propagation Algorithm	
	Thursday 7/16	Deep Learning	
10	Tuesday 7/21	Reinforcement Learning	
	Thursday 7/23	Deep Reinforcement Learning	Homework 4 due
11	Tuesday 7/28	Easy VS Hard Problems for Learning Algorithms	
	Thursday 7/30	Overview of Video, Speech, and Text Processing	
12	Tuesday 8/4	Final Project Presentations	
	Thursday 8/6	Final Project Presentations	Final Report due