CSE537 Artificial Intelligence, Spring 2016
Professor Anita Wasilewska

web page: http://www.cs.stonybrook.edu/~ cse634/

Meets Tuesday Thursday 7:00 pm - 18:50 pm

Place JAVITS 109

Professor Anita Wasilewska

e-mail address: anita@cs.stonybrook.edu

Office phone number: 632 8458

Office location: Computer Science Department building, office 208

Office Hours Tuesday, Thursday 2:30 pm - 3:30 pm and by appointments.

PLEASE NOTE the course webpage address is http://www.cs.stonybrook.edu/~ cse634/

TA consult the webpage

TA Office Hours t.b.a

Course Goal and Description

Artificial Intelligence is a broad and well established field and AI textbooks seem to be getting longer and longer and and often narrowly specialized. Our course attempts t to provide a **concise** and accessible introduction to the field. It is designed to give a broad, yet in-depth overview of different fields of AI. We will examine the most recognized techniques and algorithms in a **rigorous detail**. We will also explore trends, areas, and developments of the field in form of students' **Research Presentaitons** based on newest research and applications.

Course Structure and Books

The course is divided into **three parts**, the third one reserved for students presentations.

Part 1: General Introduction]

This part is based on parts of two **additional books** for the course:

Book 1

LOGICAL FOUNDATION of ARTIFICIAL INTELLIGENCE
Michael, F. Genesereth, Nils, J. Nilsson
Morgan Kaufman Publishers

This is the most classical book and unfortunately out of print. I will provide **PDF copies** of the first section of the book that I cover in my LECTURE NOTES

Description: The first section of the book introduces the logicist approach to AI–discussing the representation of declarative knowledge and featuring an introduction to the process of conceptualization, the syntax and semantics of predicate calculus, and the basics of other declarative representations such as frames and semantic nets. This section also provides a simple but powerful inference procedure, resolution, and shows how it can be used in a reasoning system.

About the Author Nils J. Nilsson's long and rich research career has contributed much to AI. He has written many books, including the classic Principles of Artificial Intelligence. Dr. Nilsson is Kumagai Professor of Engineering, Emeritus, at Stanford University. He has served on the editorial boards of Artificial Intelligence and Machine Learning and as an Area Editor for the Journal of the Association for Computing Machinery. Former Chairman of the Department of Computer Science at Stanford, and former Director of the SRI Artificial Intelligence Center, he is also a past president and Fellow of the American Association for Artificial Intelligence

Book 2

MANAGING UNCERTAINTY IN EXPERT SYSTEMS
Jerzy W. Grzymala-Busse
Kluwer Academic Publishers

This is very well written and comprehensible book with a lot of exercises. I will provide **PDF copies** of relevant chapters covered in my LECTURE NOTES

Part 2: Machine Learning

This part is based on the **required book** for the course:

DATA MINING Concepts and Techniques
Jiawei Han, Micheline Kamber
Morgan Kaufman Publishers, 2006,11

We will cover MACHINE LEARNING chapters of the BOOK.

I will use MY OWN LECTURE NOTES based on these Chapters.

Amazon has plenty of possibilities of buying the Book- not expensive. It is A CLASSIC BOOK and it is useful to have it.

There are a free downloads of PDFs of the Book

Here is a full set of Book Second Edition Slides;

http://web.engr.illinois.edu/ hanj/bk2/slidesindex.htm

Description This book explores the concepts and techniques of data mining, a promising and flourishing frontier in database systems and new database applications. Data mining, also popularly referred to as knowledge discovery in databases (KDD), is the automated or convenient extraction of patterns representing knowledge implicitly stored in large databases, data warehouses, and other massive information repositories. Data mining is a multidisciplinary field, drawing work from areas including database technology, artificial intelligence, machine learning, neural networks, statistics, pattern recognition, knowledge based systems, knowledge acquisition, information retrieval, high performance computing, and data visualization.

Part 3: Student's Research Presentations

Grading

During the semester you have to complete the following.

1. TWO mid- semester TESTS - total for two - (200 pts)

2. Research Presentation (60pts);

3. Students Presentations evaluation reports (10 points);

4. Final Project (30pts)

Final grade computation

Attention: **NONE of the GRADES will be curved!**

During the semester you can earn 300pts or more (in the case of extra points).

The % grade will be determine in the following way: # of earned points divided by 3 = % grade.

The % grade which is **translated** into letter grade as follows.

100 - 90 % is A range; A (100-96%) A- (95- 90%), 89 - 80 % is B range: B- (80 - 82%), B (83 -85%), B+ (86 -89%) , 79 - 70 % is C range: C- (70- 72%), C (73-75%), C+(76-79%), 69 - 60 % is D range and F is below 60%.

RESEARCH PRESENTATION DESCRIPTION

Presentation **main goal is to teach others** the material. It has to be a detailed, lecture type presentation.

Students are free to choose their own subjects.

It can be an overview of an AI subject not covered in the class, or an **extension** of a subject covered in class.

It can be in depth presentation of a software involving AI techniques, of a product that uses an AI techniques

Search the WEB, look in for the lectures, research papers, books- anything goes. It has to be YOUR own vision of what you think is **interesting** in or about AI **today**!

Bring what you find interesting and share with us

Students presentations are a VERY IMPORTANT part of the whole course design. You will bring us up date with AI technology, research, implementations, and trends and explore areas not covered by me or **extend** subjects I covered.

SOME POSSIBLE PRESENTATION SUBJECTS are:

Statistical Methods 1: Statistical Prediction, Prediction by Regression, other purely statistical methods.

Statistical Methods 2: Advances Neural Networks.

Statistical Methods 3: Advances in Bayesian Classification.

Statistical Methods 4: Cluster Analysis. A Categorization of major Clustering methods.

Evolutionary Computing: Genetic algorithms as optimization, Genetic algorithms as classification. Other evolutionary computing methods.

NEW ADVANCES like for for example:

Web Mining: an overview of methods and problems

Text Mining: an overview of methods and problems

Visualization: as an area of new applications of Machine Learning

Natural Language: NL Processing, NL Understanding- overview of methods and problems

Quantum Computing; theory and hope for the future

GENERAL PRINCIPLES OF THE PRESENTATIONS

First slide must contain: the title (subject) of your presentation, your **Team members** names, student ID, professor name, course number and the title.

Second slide must contain ALL sources you used for the your presentation. The course book, or lecture notes is included. In the case of the book the reference you have to put are title of the chapter, sections and pages numbers.

Third slide is an OVERVIEW of your presentation.

Remember to include **a source** of any picture, of slides copied from a source or any DIRECT citation on the bottom of each of your slides where it appears. REFERENCES are very important. You must be clear about the distinction between the information from a source and your own statements.

Presentation slides have to be e-mailed to the Professor and TA before the presentation.

You can also e-mail a working copy of you want to have Professor's feed back before you deliver a final version.

You receive 0-10pts for the organization of your presentation.

It is a presenter responsibility to STRUCTURE the presentation to fit the time framework.

Remember to leave some time for students (and professor) questions and discussion.

Practice and time your presentation before you present it in the class.

Presentations Teams

Each student has to deliver a a 30 minutes long presentation on a **chosen topic of AI** as a member of a **Presentation Team**

It is students responsibility to form the **Presentation Teams**

Each team has to have a designated **Team Leader** in order to communicate with Professor and course TA.

Please **e-mail TA** as soon as possible, the following:

1. list names and e-mails of your **Team** members denoting who is the **Team Leader**

2. TITLE and a one short paragraph long description of your team presentation. This is a PRELIMINARY proposal and can be changed later.

TA will assign a **Team Number** to each team and e- mail it to each TEAM LEADER to be used for future correspondence.

You have to use your **Team Number** when reserving the **presentation date**

Students Presentations will be available on the course webpage for other students to help them to write their **final presentations reports**.

Of course students **should** attend the presentations to **learn** the material and evaluate the presentation delivery

I will **collect** their **preliminary reports** written in class during the presentations

By having access to already delivered (and improved, if needed) presentations students will be able to comprehend better the material and hence to judge better other students work and write their **final presentation report**

The **final presentation report** is invalid without submitted in class preliminary report

Presentations evaluation reports (10pts)

Each student has to evaluate 10 presentations (1pts each) and submit the evaluation report. Students evaluation reports are to be FIRST written during presentations and submitted to Professor at the end of the class, then can be improved - and resubmitted after the presentations that were evaluated are published on the web page.

Each report must include:

1. one page description-summary (own words!) of the presentation content,

2. your own evaluation of the presentation.

Evaluation forms are on the course web page.

FINAL PROJECT

You work on the Project in the same **Teams** as you formed for the Research Presentations

PROJECT DESCRIPTION

THE PROJECT GOAL is to use Internet based Classification Tools to **build two classifiers:** one descriptive and one statistical and **discuss** the results and **compare** these to approaches on the basis of obtained results.

1. Descriptive Classifier

Use a TOOL to generate sets of DISCRIMINANT RULES describing the content of the data.

You can choose one you like, or use WEKA:
http://www.cs.waikato.ac.nz/~ ml/weka/index.html)

2. Non-Decsriptive Classifier

Use Neural Networks to build your Classifier

Choose your own tool: here are some suggestions:

http://www.mathworks.com/products/neural-network/?requestedDomain=www.mathworks.com

http://www.simbrain.net/

PROJECT DATA is provided on the course web page.

This is a real life classification data with TYPE DE ROCHE (Rock Type) as a CLASS attribute. There are 98 records with 48 attributes and 6 classes.

Classes are:

C1 : R. Carbonatees AND R. Carbonatees impures

C2 : Pyrate

C3 : Charcopyrite

C4 : Galene

C5 : Spahlerite

C6 : Sediments terrigenes

Most important attributes (as determined by the expert) are: **S, Zn, Pb, Cu, CaO+MgO, CaO, MgO, Fe2O3**

This is a real life experimental data and it contains a lot of missing data (no value).

The project has to follow all steps of **Learning Process**

Data Preparation that includes attributes selection, cleaning the data, filling the missing values, etc...

Data preprocessing : must use at least 2 methods of data discretization, and compare the final results obtained after each of them.

Learning Proper : for each experiment describe below use a classification tool for rules generation applied to the TWO sets of preprocessed data and compare the results.

Experiments ; you have to perform 3 experiments (all on the same preprocessed data)

Experiment 1 : use all records to find rules for the full classification; i.e. rules describing all classes **C1- C6** simultaneously.

Experiment 2 : use all records to find rules contrasting class **C1** with all others

Experiment 3 : repeat Experiments 1, 2 for all records with the **most important attributes** as defined thy the expert only.

Write a detailed Project Description with methods, motivations, results and submit via e-mail to TA and Professor. Please, submit also a **HARD COPY** the Professor.

Course Contents and Schedule

Part One: GENERAL INTRODUCTION

We will cover the following subjects included in NILSSON Book and in LECTURE NOTES

1. AI history and applications.

2. Knowledge Representation and Inference.

3. Short overview of EXPERT SYSTEMS Design and Technology.

4. Overview of Propositional and Predicate Logic; Predicate languages and basic Laws of Quantifiers

5. Automated theorem proving 1: Propositional Resolution.

6. Automated theorem proving 2: Predicate Resolution.

MIDTERM 1 Date to be determined

Part Two: MACHINE LEARNING

We will follow the DATA MINING book very closely and in particular we will cover the following chapters and subjects. The order does not need to be sequential.

We will use my own **Lecture Notes** based on the BOOK and I will also post the Book Slides as a reference

Chapter 1 Introduction. General overview: what is Data Mining, which data, what kinds of patterns can be mined.

Chapter 2 Data preprocessing: data cleaning, data integration and transformation, data reduction, discretization and concept hierarchy generation.

Chapter 5 Mining Association Rules in Large Databases. Transactional databases and Apriori Algorithm

Chapter 6 Classification and prediction.

1. Decision Tree Induction ID3, C4.5

2. Neural Networks

3. Bayesian Classification

4. Classification based on Concepts from Association rule mining

5. Genetic algorithms

Chapter 7 Cluster Analysis

MIDTERM 2 Date to be determined

Part Three: STUDENTS PRESENTATIONS

FINAL PAPER is is due anytime before Finals Week.

FINAL PROJECT is is due anytime before Finals Week.

Required Syllabi Statements: The University Senate has authorized that the following required statements appear in all teaching syllabi on the Stony Brook Campus.

Americans with Disabilities Act: If you have a physical, psychological, medical or learning disability that may impact your course work, please contact Disability Support Services, ECC(Educational Communications Center) Building, Room 128, (631)632-6748. They will determine with you what accommodations, if any, are necessary and appropriate. All information and documentation is confidential.

Academic Integrity: 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. Faculty is required to report any suspected instances of academic dishonesty to the Academic Judiciary. Faculty in the Health Sciences Center (School of Health Technology & Management, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures.