**CS 5751: Introduction to Machine Learning (4)**

**Catalog Description:**
Survey of methods in machine learning including supervised and unsupervised methods. Topics covered may include clustering, decision trees, neural networks, support vector machines, genetic algorithms and reinforcement learning. Theoretical concepts associated with machine learning.

**Textbook:** Tom Mitchell, *Machine Learning*, McGraw-Hill, 1997.

**Course Goals:**
This course presents a comprehensive introduction to the field of machine learning as well as significant insight into the related field of data mining. The overall goal of the course is for students to develop the background knowledge and hands-on experience that would allow them to take a graduate survey course in new machine learning research or to work at a data mining or machine learning company. One major goal of the course is to give students significant knowledge and experience with a variety of classifier learning methods such as decision trees, neural networks and instance-based learning techniques that form the core of machine learning. Students also acquire significant knowledge regarding control learning methods (such as reinforcement learning) and in unsupervised learning methods (such as clustering techniques). Further, students are expected to gain increased proficiency with the statistical techniques employed in machine learning algorithms. Students are also expected to gain increased proficiency in programming languages, software design and team implementation as part of the process of designing their program solutions. Finally, students are expected to gain increased proficiency with experimental methodology as part of the process of validating and analyzing the programming solutions they develop for machine learning problems.

**Prerequisites by Course & Topic**
CS 3512: Computer Science Theory – basic CS theory, proof methods, via CS 2511, software engineering, object-oriented design, software testing and debugging, team implementation
Stat 3611: Introduction to Probability and Statistics – basic probability, Bayes theorem, random variables, mathematical expectation, binomial and normal distributions, basic statistics

**Major Topics Covered in the Course**
- Data Set Construction and Representation
- Inductive Learning (Classification)
- Analytical Learning
- Control Learning
- Unsupervised Learning (Clustering)
- Hybrid Methods
- Emerging Techniques
- Statistical and Theoretical Basis for Learning

**Course Outcomes**
1. Familiarity with the foundations of machine learning.
   a. Understand the statistical basis of learning methods.
   b. Understand basic issues of statistical testing of learning methods and solutions
   c. Understand basic issues in classifier generation.
   d. Familiarity with basic issues of unsupervised learning.
   e. Familiarity with basic issues of control learning.
2. Proficiency with basic classifier generation.
   a. Proficiency with a variety of classifier methods including decision trees, neural networks, naïve bayes learning, nearest neighbor methods.
   b. Familiarity with evolutionary methods such as genetic algorithm learning.
3. Familiarity with analytic learning methods.
   a. Familiarity with explanation-based learning methods.
   a. Proficiency with hybrid learning methods involving domain theories and adaptive learning methods.

4. Familiarity with unsupervised learning.
   a. Familiarity with basic clustering methods.
5. Familiarity with control learning.
   a. Proficiency with reinforcement learning methods and implementation.
   b. Familiarity with temporal difference learning
6. Ability to work successfully on a team.
   a. Meet with team members to design project goals, delegate responsibilities, and complete projects.
   b. Communicate with team members.
   c. Integrate code generated by different team members.
7. Further proficiency with software design in high level languages.
   a. Design and develop various machine learning algorithms as part of a machine learning package.
   b. Familiarity with testing methodology for evaluating learning methods.

**Relationship to Program Outcomes**

CS 5751 requires prior completion of systems analysis and design and probability and statistics. It contributes to meeting the following program outcomes:

*1. Students understand the mathematics and statistics that underlie scientific applications.*
Students deepen their understanding of statistical techniques as they must employ statistical methods used in machine learning as part of solving their programming assignments, homework and exam questions. Course outcomes 1-5 map to this program outcome.

*2. Students can design, develop, and analyze significant software systems.*
Students must complete significant programming assignments which require them to understand significant problem statements and to design appropriate solutions for those problem descriptions. This process greatly increases their awareness of good programming practices and software design. Course outcomes 1-5 and 7 map to this program outcome.

*3. Students understand the fundamentals of computer organization and architecture, data structures and related algorithms, and programming languages.*
Students implement their programming projects in C++, increasing their proficiency in this language. Students must understand complex code and data structures and to implement further complex data structures and algorithms as part of the solutions to their programming assignments. Course outcomes 1-5 and 7 map to this program outcome.

*4. Students can apply computer science principles and practices to a variety of problems.*
Students gain experience developing solutions to machine learning problems employing standard methods such as decision trees, neural networks and clustering. Course outcomes 1-5 and 7 map to this program outcome.

*5. Students can work independently and also work effectively in teams.*
This course enables students to gain experience in working as a member of a team and producing a successful team project and report. Students also work individually on homework assignments. Course outcome 6 maps to this program outcome.

*9. Students understand the scientific method and can apply this mode of inquiry in a laboratory setting.*
As part of the process of testing their solutions, students are required to perform appropriate empirical testing methods using the scientific method in order to assess their solutions, both in terms of quality and for the capabilities of their solutions. Course outcome 7 maps to this program outcome.

**Assessment Plan for Course:**
This course is assessed every third year by the instructor and a course assessment document covering all of the course outcomes and their effect on the program outcomes is prepared.

**Estimate CSAB Category Content**

| | CORE | ADVANCED | | CORE | ADVANCED |
|---|---|---|---|---|---|
| Data Structures | | 1 | Computer Organization and Architecture | | |
| Algorithms | | 2 | Concept of Programming Languages | | |

| Software Design | | 1 |
|---|---|---|

**Coordinator/Prepared by:** R. Maclin