Syllabus for
Introduction to Machine Learning (MLearn 210)
Machine Learning Certificate Program

Bellevue / Online
Thursdays, Jan 12 – Mar 16, 2017: 6-9pm

Dave DeBarr
ddebarr@uw.edu; (425) 679-2428; Office Hours by Appointment

Course Description:
This course is designed to discuss fundamental concepts of statistical analyses, mathematical modeling, optimization techniques, and how they relate to a set of fundamental algorithms and concepts that are the foundation for machine learning. The topics for the course will include:

- Basics of probability and statistics
- Attribute types, distance measures and tools
- Introduction to optimization
- Discriminative models: supervised learning
- Generative models
- Model evaluation and parameter tuning
- Ensemble methods

Course Learning Objectives:
The student should be able to …

- Translate a real-world task into a machine learning problem
- Recognize common properties of machine learning models
- Describe common algorithms for constructing machine learning models
- Construct machine learning models using popular machine learning algorithms
- Select the best machine learning model for a task
- Assess the generalization performance of a machine learning model

Course Format:
The course will consist of both lecture and demonstration. Students will need access to a computer to complete weekly assignments. Please bring your computer to class.

Course Materials:
We will be using "An Introduction to Statistical Learning with Applications in R" as the text book for this course. It can be downloaded for free from the book's web site: www.StatLearning.com, or it can be purchased from Amazon: https://www.amazon.com/Introduction-Statistical-Learning-Applications-Statistics/dp/1461471370/

Technical Requirements:
The tools we will use in this class include R and related software packages.

Program Links:
Canvas Web Page (for Homework and Discussions): https://canvas.uw.edu/courses/1122352
Connect Link (for Online Attendance): http://uweoconnect.extn.washington.edu/mlearn210/

Course Topics by Date:

- **Jan 12, 2017**
  - **Chapter 1: Introduction**
  - **Chapter 2: Statistical Learning**
    - **What is statistical learning?**
    - **Assessing Model Accuracy**
    - **Lab: Introduction to R**

- **Jan 19, 2017**
  - **Chapter 3: Linear Regression**
    - **Simple Linear Regression**
    - **Multiple Linear Regression**
    - **Lab: Linear Regression**

- **Jan 26, 2017**
  - **Chapter 4: Classification**
    - **Logistic Regression**
    - **Linear Discriminant Analysis**
    - **Lab: Linear Discriminant Analysis, Quadratic Discriminant Analysis, and K Nearest Neighbor**

- **Feb 2, 2017**
  - **Chapter 5: Resampling Methods**
    - **Cross-Validation**
    - **The Bootstrap**
    - **Lab: Cross-Validation and the Bootstrap**

- **Feb 9, 2017**
  - **Chapter 6: Linear Model Selection and Regularization**
    - **Subset Selection**
    - **Shrinkage Methods**
    - **Dimension Reduction Methods**
    - **Lab 1: Subset Selection Methods**
    - **Lab 2: Ridge Regression and the Least Absolute Shrinkage and Selection Operator**
    - **Lab 3: Principal Components Regression and Partial Least Squares Regression**

- **Feb 16, 2017**
  - **Chapter 7: Moving Beyond Linearity**
    - **Polynomial Regression, Step Functions, and Basis Functions**
    - **Regression Splines and Smoothing Splines**
    - **Local Regression**
    - **Generalized Additive Models**
    - **Lab: Non-Linear Modeling**

- **Feb 23, 2017**
    - **Chapter 8: Tree-Based Methods**
        - **The Basics of the Decision Tree**
        - **Bagging, Random Forests, and Boosting**
        - **Lab: Decision Trees**

- **Mar 2, 2017**
    - **Chapter 9: Support Vector Machines**
        - **Maximal Margin Classifier**
        - **Support Vector Machines**
        - **Support Vector Machines with More than Two Classes**
        - **Lab: Support Vector Machines**

- **Mar 9, 2017**
    - **Chapter 10: Unsupervised Learning**
        - **Principal Component Analysis**
        - **Clustering Methods**
        - **Lab 1: Principal Component Analysis**
        - **Lab 2: Clustering**
        - **Lab 3: National Cancer Institute 60 Data Example**

- **Mar 16, 2017**
    - **Neural Networks**
    - **Genetic Algorithms**

Student Assessment:
There will be a brief take-home quiz each week covering the week's material (50% of the grade). There will be hands-on assignments each week as well (50% of the grade).

Policies and Values:
You must attend at least 6 of the 10 sessions to be given credit for the course. You must satisfactorily complete at least 80% of the assigned work to receive credit for the course. We value both academic and personal integrity, as well as respect for others and the free exchange of ideas.