ROLLINS SCHOOL OF PUBLIC HEALTH
EMORY

DEPARTMENT: Biostatistics and Bioinformatics

COURSE NUMBER: 740 SECTION NUMBER: 0 SEMESTER: Fall 2012

CREDIT HOURS: 2

COURSE TITLE: Bioinformatic Machine Learning

INSTRUCTOR NAME: Tianwei Yu

INSTRUCTOR CONTACT INFORMATION:

EMAIL: tianwei.yu@emory.edu

PHONE: 404-727-7671

SCHOOL ADDRESS OR MAILBOX LOCATION: GCRB 334

OFFICE HOURS: TBA

COURSE DESCRIPTION (3-4 Sentences)

This course covers some popular supervised and unsupervised machine learning techniques in Bioinformatics and general high-dimensional data research. The topics covered fall into three categories – classification, clustering and dimension reduction.

Prerequisites:

BIOS 540 or permission of instructor

EVALUATION

The grade assignment will be based on:

Participation in class and discussions (10%);

Three homeworks (10% each);

Final project (60%). The instructor will assign a high throughput dataset. Each student should write a thorough data analysis report in research article style.

ACADEMIC HONOR CODE

The RSPH requires that all material submitted by a student in fulfilling his or her academic course of study must be the original work of the student.

LEARNING OBJECTIVES OR COMPETENCIES OF THE COURSE

After taking the course, the students are expected to have working knowledge in three areas: (1) Classification, topics including Bayesian decision theory, linear machines, support vector machines, boosting, random forests, bump hunting, and model generalization; (2) Clustering, topics including hierarchical clustering, center-based clustering, model-based clustering; (3) Dimension reduction, topics including principal component analysis, independent component analysis, projection pursuit, sliced inverse regression.

Textbook:

The elements of statistical learning. Hastie, Tibshirani & Friedman.

Other references:

Pattern classification. Duda, Hart & Stork.
Data clustering: theory, algorithms and application. Gan, Ma & Wu.
Applied multivariate statistical analysis. Johnson & Wichern.

Schedule:

Lecture 1, Overview
Lecture 2, Bayesian decision theory
Lecture 3, Density estimation
Lecture 4, Linear machine
Lecture 5, Support vector machines
Lecture 6, Tree and forest
Lecture 7, Bump hunting
Lecture 8, Boosting
Lecture 9, Generalization of models
Lecture 10, Similarity measures, hierarchical clustering
Lecture 11, Center-based and model-based clustering
Lecture 12, PCA, SIR
Lecture 13, ICA, PLSDA

LEARNING OBJECTIVES OR COMPETENCIES FOR THE DEPARTMENT OR PROGRAM TO WHICH THE COURSE CONTRIBUTES

This is an advanced course of Bioinformatics. The purpose is to prepare students for research in machine learning that focus on omics data. It is highly recommended for students who are interested in Bioinformatics research.