PENNSTATE HERSHEY College of Medicine

General Information							
Course Title:	**Data Mining With Application to Genomic Data**		Course Designation: **PHS 597**			Credits: **3**	
Semester:	**Fall**		Year:	**2016**			
Department:	**Public Health Sciences**						
Director:	**Dajiang Liu, PhD**	Tel #	**4178**	Email:	**Dajiang.liu@psu.edu,**	Office Rm #	**HCAR2020; Tel: x4178**
Time :	**1:00 – 2:30pm**	Days:	**Tuesday and Thursday 08/23/16---12/08/16**			Location:	**ASB3400M**

Course Information

Description and/or Overview:

This course covers basics for statistical learning, with an emphasis on its application to genomic data. As a first course in statistical learning, we will largely follow the book of "An Introduction of Statistical Learning with Applications in R". We will cover methods on classification, resampling methods, linear models with regularization (e.g. LASSO), additive models, classification and regression trees, random forests, support vector machines and basics of unsupervised learning. An emphasis will be on its applications to cutting edge genetics and genomics problems. Areas of genomic applications will include (but not limited to) variant annotation, genetic association analysis, variant calling and filtering from next generation sequencing.

Goals and/or Objectives:

This course is designed to be the first course of a sequel of statistical learning course for high dimensional statistical learning. We will cover a broad class of methods in statistical learning and discuss their applications in cutting edge genomics research. The students will be introduced to current research topics in genomics where data mining methods play a critical role.

Pre-requisites:

Students taking this course are expected to be familiar with basic statistical principle and ideally should have a solid grasp of basic mathematical statistics (on the level of Casella and Berger). In addition, a course on multivariate statistics is a plus.

Requirements; course-specific policies and expectations:

Students will need access to a laptop computer, and will be required to bring it to class on a regular basis. The computer must have access to wireless internet from the classroom.

Required Texts and Resources:

Required Text:
1. The Elements of Statistical Learning. *Authors: Hastie T., Tibshirani R. and Friedman J.*
2. An Introduction of Statistical Learning with Applications in R. *Authors: James, G., Witten, D., Hastie, T., Tibshirani, R*
Both textbooks are free to download.

Electronic Links:

ANGEL, the Penn State course management system, will be used to post files with the course materials, such as lectures, articles, homework assignments, exams, etc.

Attendance Policy:

Students are expected to attend class regularly. Students should consult with the instructor if they anticipate missing more than one class. Cell phones and pagers should be turned off during class time in order not to disrupt the class.

Examination Policy:

Students are expected to perform their own work on the take-home assignments and not consult with classmates**.**

Grading Criteria:

There will be homework assignments, and class projects.
Homework: 70% and Projects: 30%

Individual grades will not be uploaded until the student completes the confidential CourseEval survey for evaluating the course and the instructor(s). CourseEval surveys will be initiated during the last week of class instruction.

General Information					
Course Title:	**Data Mining with Application to Genomic Data**		**Course Designation**:	**PHS597**	
Course Director:	**Dajiang Liu, PhD;**				
Time :	**2:00 to 3:15pm**	**Days:**	**M & W**	**Location**	**ASB3400M**

Date	**Lecture #**	**Instructor Last, first**	**Instruction Type (Lecture or lab)**	**Projected Lecture Topic - This list is an approximate guide to lecture topics. Titles and content are subject to change**
08/23/16	1	Liu, Dajiang	lecture	Introduction to statistical Learning
08/25/16	2	Liu, Dajiang	lecture	Introduction to statistical Learning
08/30/16	3	Liu, Dajiang	lecture	Overview of Concurrent Genomic Research
09/01/16	4	Liu, Dajiang	lecture	Linear Models

09/06/16	5	Liu, Dajiang	lecture	Linear Models
09/08/16	6	Liu, Dajiang	lecture	Overview of Concurrent Genomic Research
09/13/16	7	Liu, Dajiang	lecture	Classification
09/15/16	8	Liu, Dajiang	lecture	Linear Discriminant Analysis (LDA)
09/20/16	9	Liu, Dajiang	lecture	Applications of LDA in Genomics
09/22/16	10	Liu, Dajiang	lecture	Resampling Methods I – Validation Methods
09/27/16	11	Liu, Dajiang	lecture	Resampling Methods II – Bootstrap
09/29/16	12	Liu, Dajiang	lecture	Linear Model Selection I – Shrinkage Methods
10/04/16	13	Liu, Dajiang	lecture	Linear Model Selection II – Dimension Reduction and High Dimensional Inference
10/06/16	14	Liu, Dajiang	lecture	Variable Selection and Prediction in Genomics
10/11/16	15	Liu, Dajiang	lecture	Nonlinear Models I - Spline Regression
10/13/16	16	Liu, Dajiang	lecture	Nonlinear Models II – Generalized Additive Models
10/18/16	17	Liu, Dajiang	lecture	Nonlinear Models in Genomics
10/20/16	18	Liu, Dajiang	lecture	Tree-based Methods – Decision Trees
10/25/16	19	Liu, Dajiang	lecture	Tree-based Methods – Random Forests
10/27/16	20	Liu, Dajiang	lecture	Overview of Tree Based Methods in Genomics
11/01/16	21	Liu, Dajiang	lecture	Support Vector Machines I
11/03/16	22	Liu, Dajiang	lecture	Support Vector Machines II
11/08/16	23	Liu, Dajiang	lecture	Exemplar Application of SVM in Genomics – Variant Filtering from Next Generation Sequencing
11/10/16	24	Liu, Dajiang	lecture	Unsupervised Learning I – principle Component Analysis
11/15/16	25	Liu, Dajiang	lecture	Unsupervised Learning II – Clustering Methods
11/17/16	26	Liu, Dajiang	lecture	Clustering Methods in Phylogenetics and Population Genetics
11/22/16		Thanksgiving	Sleep	Random Topics or Makeup for Classes
11/24/16		Thanksgiving	Eating Turkey	Eating Turkey and Prepare for Black Friday

11/29/16	27	Liu, Dajiang	Lecture	Random Topics or Makeup for Classes
12/01/16	28	Liu, Dajiang	lecture	Random Topics or Makeup for Classes
12/06/16	29		Lecture	Random Topics or Makeup for Classes
12/08/16	30	Liu, Dajiang		Random Topics or Makeup for Classes

Academic Integrity

Academic Integrity at Penn State is defined by Faculty Senate Policy 49-20 as “the pursuit of scholarly activity in an open, honest and responsible manner”. The University's Code of Conduct states that “all students should act with personal integrity, respect other students' dignity, rights and property, and help create and maintain an environment in which all can succeed through the fruits of their efforts.

Academic integrity includes a commitment not to engage in or tolerate acts of falsification, misrepresentation or deception. Such acts of dishonesty violate the fundamental ethical principles of the University community and compromise the worth of work completed by others”. Academic dishonesty (including, but not limited to cheating, plagiarism, or falsification of information) will not be tolerated and can result in academic or disciplinary sanctions such as a failing (F) grade in the course.