

Math 5637: Topics in Risk Modeling

INSTRUCTOR	Chunsheng Ban
OFFICE	MW 124B
EMAIL	ban.1@osu.edu
PHONE	614-292-5331
OFFICE HOURS	TBD

- DESCRIPTION

An introduction to some basic concepts and methods in statistical learning with emphasis on the mathematics behind these concepts and methods.

The course covers most topics of Exam SRM administered by SOA and exams MAS-I and MAS-II administered by CAS.

- COURSE OBJECTIVES

The objectives of the course include understanding of basic concepts and main methods in statistical learning, such as supervised and unsupervised learning, the lasso and sparse regression, classification and regression trees, and boosting and support vector machines. Several computer labs based on the software package R will be assigned to expose students to applications.

- PREREQUISITE

Linear algebra (Math 2568 or equivalent), probability (Math 4530 or Stat 4201 or equivalent), and statistics (Stat 4202 or equivalent); or by department permission.

- RECOMMENDED TEXTS

1. *An Introduction to Statistical Learning with Applications in R*
by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

<https://link-springer-com.proxy.lib.ohio-state.edu/book/10.1007/978-1-4614-7138-7>

2. Class Notes by Chunsheng Ban
3. *Mathematics for Machine Learning*
by Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong
4. *The Elements of Statistical Learning – Data Mining, Inference, and Prediction*
by Trevor Hastie, Robert Tibshirani, Jerome Friedman
<https://web.stanford.edu/~hastie/ElemStatLearn/>
5. *Predictive Modeling Applications in Actuarial Science*
Volume I: Predictive Modeling Techniques
Edited by E. Frees, R. Derrig, G. Meyers

The course topics will mainly follow 1 with mathematical supplement in 2 and 3.

- TOPICS

- Introduction: supervised/unsupervised learning, loss function, regression and classification
- Model Accuracy: training and test error, bias-variance trade-off, KNN

- Linear Regression: simple and multiple linear regression, hypothesis testing, variable selection, model fit
- Introduction to GLM: general linear regression and link function
- Classification: logistic regression, linear and quadratic discriminant analysis
- Resampling: cross-validation, bootstrap,
- Model Selection: variable selection, ridge regression and Lasso, dimension reduction, principal component analysis, partial least squares
- Nonlinear Regression: polynomial regression, regression splines, smooth splines, generalized additive model
- Tree Based Methods: decision trees, pruning, classification and regression trees, error rate, bagging, random forests, boosting
- Neural Networks
- Support Vector Machines: maximal margin classifier, support vector classifier, kernel functions, support vector machines
- Introduction to Unsupervised Learning: PCA in unsupervised learning, K-means and hierarchical clusterings

- **GRADE**

The course grade will be based on

- Homeworks and labs, 20%
- Two midterm exams, 50%
- Final exam, 30%

Course grade will be determined by the total percentage obtained, roughly as 90–100 for an A, 80–89 for a B, 65–79 for a C, and 50–64 for a D.

- **SCHEDULE**

A tentative weekly schedule is attached. This schedule and material covered may be changed without notice. It is the students responsibility to keep track of these changes. Changes may be announced in class verbally, through Carmen, or through email.

- **ACADEMIC MISCONDUCT**

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term “academic misconduct” includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-48.7). For additional information, see the Code of Student Conduct at <http://studentaffairs.osu.edu/csc/> .

- **Disability Services Statement**

Students with disabilities that have been certified by Student Life Disabilities Services (SLDS) will be appropriately accommodated and should inform the instructor as soon

as possible of their needs. SLDS contact information: slds@osu.edu; 614-292-3307; 098 Baker Hall, 113 W. 12th Avenue.

Week 1	Introduction to Statistical Learning
Week 2	Model Accuracy, Introduction to R
Week 3	Simple and Multiple Linear Regression
Week 4	Considerations in Regression Model
Week 5	GLM, Logistic Regression
Week 6	Linear Discriminant Analysis
Week 7	Cross-Validation
Week 8	Bootstrap
Week 9	Linear Model Selection and Regularization, Principal Component
Week 10	Nonlinear Regression and Splines
Week 11	General Additive Models, Decision Trees
Week 12	Support Vector Machines
Week 13	Neural Networks, Unsupervised Learning/PCA
Week 14	K-Means and Hierarchical Clusterings