



北京理工大学国际特色课程

Beijing Institute of Technology Global Courses

## **CSE9506 - APPLIED DATA MINING AND STATISTICAL LEARNING**

### *Syllabus*

**July. 1 - July. 19, 2024**

---

Term Duration: July. 1 - July. 19, 2024

Credit Points: 4

Level: Postgraduate

Instructor Name: TBA

Home Institution: Beijing Institute of Technology

Lecture Hour: 12:30-15:30

---

### ***Course Description***

---

The course is designed to provide students with a strong foundation in both theory and application of statistical learning techniques. The course starts with the basics of statistical learning and the different types of machine learning methodologies. Students will then learn about linear regression and compare it with non-parametric methods such as K-Nearest Neighbors. The curriculum includes various linear methods for classification, resampling methods for model assessment, and techniques for model selection and regularization. Advanced topics include moving beyond linearity to methods such as additive models, trees, and boosting, as well as adaptive basis function and generalized additive models. The course will also cover neural networks and the principles of support vector machines and flexible discriminants.

## ***Course Aims:***

---

Upon successful completion of this course, students should be able to:

1. understand the fundamental concepts and different types of machine learning techniques;
2. apply linear regression and contrast it with K-Nearest Neighbors for predictive modeling;
3. utilize linear methods for classification and understand the role of resampling in model validation;
4. implement model selection strategies and regularization techniques to enhance model performance;
5. develop proficiency in neural networks, support vector machines, and flexible discriminant analysis for advanced data mining applications;
6. develop the ability of using various machine learning techniques to solve specific problems in the real business environment.

## ***Language of Instruction***

---

English

## ***Required Textbook***

---

### ***An Introduction to Statistical Learning: with Applications in R***

**Author:** Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

**Publisher:** Springer

**ISBN:** 9781461471370

### ***The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2ed Edition***

**Authors:** Trevor Hastie, Robert Tibshirani, Jerome Friedman

**Publisher:** Springer

**ISBN:** 9780387848570

### ***Probabilistic Machine Learning: An Introduction***

**Authors:** Kevin P. Murphy

**Publisher:** The MIT Press

**ISBN:** 9780262046824

Other materials provided by the course lecturer.

## Course Hours

---

This course requires 48 hours of contact including 42 hours of lectures and one 6-hour field trip. Lectures are from Monday to Friday.

## Prerequisite Course

---

Students are expected to have taken **Statistical Methods** or to possess a thorough knowledge of the topics covered in the mentioned course.

## Course Schedule

---

Week	Day	Lecture	Topic	Assignment/ Notes
Week 1	Day 1	Lecture 1	Statistical Learning; Types of Machine Learning	Gareth James (Chap 2); Kevin P. Murphy (Chap 1)
	Day 2	Lecture 2	Linear Regression; Comparison of Linear Regression with K-Nearest Neighbors	Gareth James (Chap 3)
	Day 3	Lecture 3	Linear Methods for Classification	Trevor Hastie (Chap 4)
	Day 4	Lecture 4	Classification	Gareth James (Chap 4)
	Day 5	Lecture 5	Resampling Methods	Gareth James (Chap 5)
Week 2	Day 6	Lecture 6	Linear Model Selection and Regularization	Gareth James (Chap 6); Trevor Hastie (Chap 7)
	Day 7	Lecture 7	Moving Beyond Linearity	Gareth James (Chap 7)
	Day 8	Lecture 8	Additive Models, Trees, and Related Methods	Trevor Hastie (Chap 9) Gareth James (Chap 8)
	Day 9	Lecture 9	Generative Models for Discrete Data	Kevin P. Murphy (Chap 3)
	Day 10	Industrial Visit		
Week 3	Day 11	Lecture 10	Boosting and Additive Trees; Adaptive Basis Function Models; Generalized Additive Models	Trevor Hastie (Chap 10); Kevin P. Murphy (Chap 16)

Day 12	Lecture 11	Neural Networks	Trevor Hastie (Chap 11)
Day 13	Lecture 12	Support Vector Machines and Flexible Discriminants	Trevor Hastie (Chap 12); Gareth James (Chap 9)
Day 14	Guided Revision		
Day 15	Final Exam		

**Note:** Students will be notified if the schedule of the field trip changes according to the situation.

### *Grading Policy*

---

Method	Percentage
Participation	10%
Mid-Semester Test	20%
Group Project	25%
Final Exam	45%
Total	100%

**Participation:** Students are expected to attend all course sessions punctually. Absences will impact the attendance grade. We will have some case studies, quizzes and group discussion in class and the performance will be considered as partial attendance points or bonus.

**Mid-Semester Test:** The test will span a duration of 2 hours and will assess students on the material covered in the first half of the course. Students are expected to demonstrate their knowledge and analytical skills through this examination.

**Group Project:** Students will be evaluated on a collaborative project where they apply statistical learning techniques to a real-world dataset.

**Final Exam:** The final exam will be a comprehensive assessment encompassing all course content. This exam will have a duration of 3 hours, and students will be evaluated on their understanding and application of the material covered throughout the course.

### *Academic Honesty*

---

Academic honesty is not only a fundamental part of learning and teaching, but also a core value that this course embraces. Behaviors of academic dishonesty, as outlined hereinafter, are unacceptable and will be penalized:

- a) Plagiarism where students present work for assessment, publication or otherwise that is not their own, without appropriate attribution or reference to the original source. Plagiarism can include:
- i) paraphrasing or copying published and unpublished work without a reference;
  - ii) adopting the ideas or concepts of others, including the structure of an existing analysis without due acknowledgement by way of reference to the original work or source.
- b) Collusion, where students present work as independent work when it has in fact been produced in whole or in part with others unless prior permission for joint or collaborative work has been given by the Course Coordinator. Collusion can include:
- i) a student inappropriately assisting with or accepting assistance with the production of an assessment task;
  - ii) submitting work which is the same or substantially similar as another student's work for the same assessment task.
- c) Cheating, where a student acts in such a way as to seek to gain unfair advantage or assist another student to do so. Cheating can include:
- i) submitting falsified, copied or improperly obtained data relating to results of practicum, field trips or other work as if they were genuine; submitting an assessment task with the intention of deceiving or misleading the instructor about the student's contribution to the work;
  - ii) submitting an assessment task written or answered for the student by another person or which the student has copied from another person;
  - iii) submitting the same or a substantially similar piece of work for assessment in two different courses (except in accordance with approved study and assessment schemes);
  - iv) a student falsely indicating that they have been present at an activity where attendance is required;
  - v) completing an assessment task outside the conditions specified for that task.
- d) Cheating in Examinations means engaging in dishonest practice or breaching the rules regarding examinations, which can include:
- i) communicating in any way during an examination with any person who is not an examination supervisor inside or outside the examination venue;
  - ii) giving or accepting assistance from any person who is not an examination supervisor whilst in the examination venue;
  - iii) reading, copying from or otherwise using another student's work in an examination or knowingly allowing a student to do so;
  - iv) possessing, referring to or having access to any material or device containing information directly or indirectly related to the subject matter under examination, other than that explicitly approved by the Course Coordinator;
  - v) acquiring, or attempting to acquire, possess or distribute examination materials or information without approval;
  - vi) permitting another person to attend an examination on a student's behalf or attending an examination on behalf of another student;
- e) Other dishonest acts including but not limited to:
- i) altering or falsifying any document or record for the purposes of gaining academic advantage;
  - ii) offering or giving money or any item or service to a University staff member or any other person to gain academic advantage for the student or another person;
  - iii) inventing references.