BUAS546

Data Mining and Modern Analytics

BUAS546

Data Mining and Modern Analytics

Instructor Contact Details

Lecturer-in-charge: Sha ZHAO (TBD)
Email: wlwyxy_29@zju.edu.cn
Office location: Huajiachi Campus, Zhejiang University, Hangzhou, China
Consultation Time: Book appointment by sending email to: wlwyxy_29@zju.edu.cn

Teaching Times, Modes and Locations

Course Duration: 23 Jun 2025 to 11 Jul 2025
Modes: Online/Face-to-face
Location: Anywhere via online, or Huajiachi Campus, Zhejiang University via face-to-face

Academic Level

Undergraduate

Credit Points:

The course is worth 6 units of credit point.

Credit Hours

The number of credit hours of this course equals to the credits of a standard semester- long course.

Contact Hours

The course contains a total of 53 contact hours, which consists of orientation, lectures, seminars, quiz, discussion, research, case study, small tests, assignments, on-site field trip(s), in-class and after-class activities, revision, self-study, and final exam. Students will receive an official transcript which is issued by Zhejiang University when completing this course.

Enrolment Requirements

Eligibility requires enrollment in an overseas university as an undergraduate or postgraduate student, proficiency in English, and pre-approval from the student's home institution.

Course Description:

This unit provides an accessible overview of the field of statistical learning, an essential toolset for making sense of the vast and complex data sets that have emerged in fields ranging from biology to finance to marketing to astrophysics in the past twenty years. This unit also presents some of the most important modeling and prediction techniques, along with relevant applications. Topics include linear regression, classification, resampling methods, shrinkage approaches, tree-based methods, support vector machines, clustering, deep learning, survival analysis, multiple testing, and more.

Prerequisite:

N/A

Learning Resources

● Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning with Applications in R, 1st edition, Springer. 2017

Learning Objectives

By the end of this course, you should be able to:

● Formulate business decision problems as predictive machine learning problems.

● Extract business insights from large volumes of data using machine learning and data mining methods.

● Recognize how machine learning can help organizations optimize business processes and make effective decisions at scale.

Course Delivery:

• Online Lecture mode includes lectures, seminars, quiz, discussion, research, case study, small tests, assignments, online field trip(s), in-class activities, revision, and final exam.

• Face-to-face Lecture mode includes lectures, seminars, quiz, discussion, research, case study, small tests, assignments, on-site field trip(s), in-class and after-class activities, revision, and final exam.

The following course will be taught in English. There will also be guest speakers and optional field trips available for students who would like to enhance their learning experience. All courses and other sessions will be run during weekdays.

Topics and Course Schedule:

WK	Topic	Activities
1	Introduction: What is Statistical Learning?	Lecture; Tutorial
1	Introduction to R	Lecture; Tutorial
1	Simple Linear Regression	Lecture; Tutorial
1	Multiple Linear Regression	Lecture; Tutorial
1	Classification: Logistic Regression	Lecture; Tutorial
1	Classification: Linear Discriminant Analysis	Lecture; Tutorial
1	Classification: A Comparison of Classification Methods	Lecture; Tutorial
2	Resampling Methods: Cross-Validation	Lecture; Tutorial
2	Linear Model Selection and Regularization: Subset Selection	Lecture; Tutorial
2	In-class Test	Closed book
2	Linear Model Selection and Regularization: Best Subset Selection	Lecture; Tutorial
2	Linear Model Selection and Regularization: Stepwise Selection	Lecture; Tutorial
2	Linear Model Selection and Regularization: Choose the Optimal Model	Lecture; Tutorial
2	Linear Model Selection and Regularization: lab 1: Subset Selection methods	Lecture; Tutorial
3	Moving Beyond Linearity: Regression Splines	Lecture; Tutorial
3	Tree-Based Methods: Classification Trees	Lecture; Tutorial
3	Tree-Based Methods: Bagging, Random Forests, Boosting	Lecture; Tutorial
3	Principal Components Analysis	Lecture; Tutorial
3	Revision	Tutorial
3	Final exam	Closed book

Assessments:

Class participation	15%
In-class Test	15%
Assignments	20%
Final exam	50%

Grade Descriptors:

HD	High Distinction	85-100
D	Distinction	75-84
Cr	Credit	65-74
P	Pass	50-64
F	Fail	0-49

High Distinction 85-100

• Treatment of material evidences an advanced synthesis of ideas Demonstration of initiative, complex understanding, and analysis.
• Work is well-written and stylistically sophisticated, including appropriate referencing, clarity, and some creativity where appropriate.
• All criteria addressed to a high level.

Distinction 75-84
• Treatment of material evidences an advanced understanding of ideas Demonstration of initiative, complex understanding and analysis Work is well-written and stylistically strong.
• All criteria addressed strongly.

Credit 65-74
• Treatment of material displays a good understanding of ideas
• Work is well-written and stylistically sound, with a minimum of syntactical errors.
• All criteria addressed clearly.

Pass 50-64
• Treatment of material indicates a satisfactory understanding of ideas Work is adequately written, with some syntactical errors.
• Most criteria addressed adequately.

Fail 0-49
• Treatment of ideas indicates an inadequate understanding of ideas Written style inappropriate to task, major problems with expression.
• Most criteria not clearly or adequately addressed.

<u>Academic Integrity</u>

Students are expected to uphold the university's academic honesty principles which are an integral part of the university's core values and principles. If a student fails to observe the acceptable standards of academic honesty, they could attract penalties and even disqualification from the course in more serious circumstances. Students are responsible for knowing and observing accepted principles of research, writing and any other task which they are required to complete.

Academic dishonesty or cheating includes acts of plagiarism, misrepresentation, fabrication, failure to reference materials used properly and forgery. These may include, but are not limited to: claiming the work of others as your own, deliberately applying false and inaccurate information, copying the work of others in part or whole, allowing others in the course to copy your work in part or whole, failing to appropriately acknowledge the work of other scholars/authors through acceptable referencing standards, purchasing papers or writing papers for other students and submitting the same paper twice for the same subject.

This Academic Integrity policy applies to all students of the Zhejiang University in all programs of study, including non-graduating students. It is to reinforce the University's commitment to maintain integrity and honesty in all academic activities of the University community.

<u>Policy</u>

The foundation of good academic work is honesty. Maintaining academic integrity upholds the standards of the University. The responsibility for maintaining integrity in all the activities of the academic community lies with the students as well as the faculty and the University. Everyone in this community must work together to ensure that the values of truth, trust and justice are upheld.

Academic dishonesty affects the University's reputation and devalues the degrees offered. The University will impose serious penalties on students who are found to have violated this policy. The following penalties may be imposed:

- ✓ Expulsion
- ✓ Suspension
- ✓ Zero mark /fail grade
- ✓ Marking down
- ✓ Re-doing/re-submitting of assignments or reports, and
- ✓ Verbal or written warning.

.