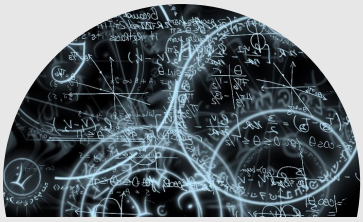

Introduction to Statistical Learning

MATH 624-C01 (Fall 2025)

Instructor Info

Dr. Faezeh Soleimani (She/Her)

Office: RB 435

E-mail: fsoleimani@bsu.edu

Office H: Wed 11 am-noon (EST)

My BSU Profile

Course Info

Format: Asynchronous Online

Platform: Coursera

Important Note

It may take up to 24 hours to respond to your emails during weekdays (Mon-Fri) and up to 48 hours during weekends.

In your email, please include the course name/number (MATH 624).

I will use Zoom to hold online meetings for the class.

Overview

Statistical learning refers to a set of tools for modeling and understanding complex data sets. It is a recently developed area in statistics and blends with parallel developments in computer science and, in particular, machine learning. With the explosion of "Big Data" problems, statistical learning has become a very hot field in many scientific areas as well as marketing, finance, and other business disciplines and people with statistical learning skills are in high demand.

Prerequisites

MATH 620/DSCI 602 or permission of the department chairperson.

Recommended Textbook & Technology

- An Introduction to Statistical learning with Applications in R, (2013) by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. Link
- Elements of Statistical Learning: Data Mining, Inference and Prediction, (2011) by Trevor Hastie, Robert Tibshirani and Jerome Friedman. Link
- Technology: R/RStudio will be utilized in this course. You may download R from this site and download a user-friendly interface RStudio from this site for free.

Course Content/Description

- Supervised learning: simple and multiple linear regression, logistic regression, linear/quadratic discriminant analysis, resampling methods and bootstrap, model/variable selection and regularization, tree-based methods.
- Unsupervised learning: principal component analysis, canonical correlation, clustering methods.

Course Objectives

Students will be able to:

- Explain fundamental concepts of statistical learning, including supervised and unsupervised learning.
- Conduct exploratory data analysis (EDA) to discover the structure and patterns and find relationships among different variables within a dataset.
- Explain various algorithms used in statistical learning, such as linear regression, logistic regression, decision trees, ensemble methods like random forests, and clustering algorithms.
- Implement statistical learning techniques using the R programming language, including data preprocessing, model fitting, and evaluation.
- Evaluate the performance of different models using techniques like cross-validation and evaluation metrics such as mean square error, accuracy, and ROC curves.
- Apply techniques for feature selection and dimensionality reduction to improve model performance and interpretability.
- Interpret the results of statistical learning models, including understanding model coefficients, feature importance, and prediction intervals.
- Apply statistical learning techniques to real-world datasets, gaining practical experience in solving regression, classification, and clustering problems.

Course Rationale

With the explosion of data in almost every field including biology, health sciences, finance and business, environmental sciences and agronomy, the problem of storing, managing, analyzing and extracting meaningful knowledge from these data poses a substantial challenge. A statistician's job now becomes to extract important patterns and trends from the bulk of data. Our graduates need such state of the art skills in order to compete in today's competitive job market. This course is intended to provide an introduction to statistical learning tools that are essential parts of a statistician's expertise. This course is required for the statistics program and may be of interest to students in actuarial science, computer science and mathematics.

Grading Scheme

20%	Auto-graded Quizzes: They will be assigned every week with some exceptions. These quizzes are given to test concepts in the lectures and/or in the lab demo videos for programming code expressions.
20%	Lab Assignments: These assignments will be assigned every week with some exceptions and are given to assess your programming skills.
10%	Graded Peer-reviewed Discussion Boards: A total of 10 discussion entries will be assigned. Students are required to assess their peers' discussions and assign grades according to a predefined rubric.
10%	Graded Reflective Peer-reviewed Reading Assignments (Papers): You need to read and summarize papers on specific topics. A total of three papers will be assigned. Paper I covers classification and regression methods, Paper II covers variable selection and dimension reduction, and Paper III covers tree-based classification methods.
20%	Midterm Project: A midterm project is scheduled during a three-week period. This Project covers the topics discussed under supervised learning methods.
20%	Final Project: A final project is scheduled during a three-week period. This project covers both supervised and unsupervised learning.

Important Note

Students are responsible for all material submitted correctly for assignments.Please make sure you submit the correct and readable version of your documents.

Grade Categories and Weights

[93, 100]	A
[90,93)	A-
[87,90)	B+
[83,87)	B
[80,83)	B-
[77,80)	C+
[73,77)	C
[70,73)	C-
[67,70)	D+
[63,67)	D
[60,63)	D-
[0,60)	F

Attendance Policy

Given the asynchronous online format of the course, attendance is NOT obligatory. Nevertheless, to enhance your academic success, it is crucial that you organize your schedule to allocate 2-3 hours per week for coursework for each credit you are enrolled in.

Feedback and Grading Policy

- If you have any questions please contact us using ds.support@bsu.edu (for MSDS students) or cs.support@bsu.edu (for MSCS students). I will typically respond to messages within 24 hours of receiving them during weekdays (Monday-Friday) and within 48 hours on the weekend.
- Emails must include the COURSE NUMBER AND SECTION NUMBER in the subject line in order to be answered.
- When asking for help to fix the code error, please state the error clearly.
- It is my policy that appropriate evaluation of your academic performance is an integral part of your learning experience. In the absence of mistakes, fraud, bad faith or incompetence, I will be the key decision-maker on the assignment of grades. Please make your submissions readable and organized clearly.
- Some of the assignments are auto-graded, I will make every effort to grade non auto-graded assignments no later than 7 business days of submission.

Technical Information

It is your responsibility to become familiar with the class interface and seek help if necessary. BSU and Coursera provide resources to help you. A lack of Coursera knowledge is not an acceptable excuse for late or incomplete work.

- Technical Equipment: For you to be successful in this type of course, access to high speed internet is essential.
- Technical Assistance: If you have technical questions, you can contact Coursera Help Center or BSU HelpDesk.

Course Withdrawal

The course withdrawal period ends October 27th, 2025. Before this date, students can elect to receive a "W" for the course by completing and submitting the proper form. The instructor's permission is not required. For details, see this link OR Withdrawal Procedures in the current graduate catalog.

University Statement

We are committed to ensuring that all members of the community are welcome, through valuing the various experiences and worldviews represented at Ball State and among those we serve. We promote a culture of respect and civil discourse. If you need course adaptations or accommodations because of a disability, please contact me as soon as possible. Ball State's Disability Services office coordinates services for students with disabilities; documentation of a disability needs to be on file in that office before any accommodations can be provided. Disability Services can be contacted at 765-285-5293 or dsd@bsu.edu.

Freedom of Expression

In this course, we are committed to fostering a learning environment that values intellectual diversity, encourages free expression, and promotes open inquiry. As members of the Ball State Community, we treat each person in the Ball State community with civility, courtesy, compassion, and dignity and respect and learn from differences in people, ideas, and opinions. Please review Ball State University's Statement on Freedom of Expression, the resources on Ball State's Freedom of Expression webpage, and Ball State's Beneficence Pledge.

Academic Honesty

Ball State's Academic Ethics Policy is given here. Violating this policy can lead to disciplinary action up to and including expulsion from the university.

Student Academic Ethics Policy & Rights & Responsibilities

Ethics Policy, Rights and Responsibilities.

Course Schedule

Weeks	Topics	Q*/L**/DB***	Paper/Projects
W 1: 08/18-08/22	Introduction to Statistical Learning • What is Statistical Learning? • Supervised versus Unsupervised Learning • Regression versus Classification Problems	• Q1 • DB 1	-
W 2: 08/25-08/29	Introduction to Statistical Learning • Regression function f • Assessing Model Accuracy • Classification Setting • Introduction to Exploratory Data Analysis (EDA)	• Q2 • L1	-
W 3: 09/02-09/05	Linear Regression • Simple Linear Regression • Multiple Linear Regression • Other Considerations in the Regression Model	• Q3 • L2	-
W 4: 09/08-09/12	Classification with Logistic Regression • Overview of Classification • The Logistic Regression Model • Simple and Multiple Logistic Regression	• Q4 • L3 • DB 2	-
W 5: 09/15-09/19	Classification with Linear Discriminant Analysis • Linear Discriminant Analysis • Quadratic Discriminant Analysis • K-Nearest Neighbor (KNN)	• Q5 • L4 • DB 3	-
W 6: 09/22-09/26	Evaluating Model Performance • The Confusion Matrix • The ROC Curve	• Q6	• Paper I • Midterm Project
W 7: 09/29-10/03	Resampling Methods: Cross-Validation • The Validation Set Approach • Leave-One-Out Cross-Validation (LOOCV) • k-Fold Cross Validation • Cross-Validation on Classification Problems	• Q7 • L5 • DB 4	• Midterm Project
W 8: 10/08-10/10	Resampling Methods: The Bootstrap • Bootstrap in General • Uses of Bootstrap	• Q8	• Midterm Project
W 9: 10/13-10/17	Subset Selection • Best Subset Selection • Stepwise Selection	• Q9 • L6 • DB 5	-
W 10: 10/20-10/24	Shrinkage Methods • Ridge Regression • The Lasso	• Q10 • L7 • DB 6	• Paper II

Weeks	Topics	Q*/L**/DB***	Paper/Projects
W 11: 10/27-10/31	Dimension Reduction Methods • Principal Components Regression • Partial Least Squares	• Q11 • L8 • DB 7	-
W 12: 11/03-11/07	Tree-Based Methods • Regression Trees • Classification Trees	• Q12 • L 9 • DB 8	-
W 13: 11/10-11/14	Ensemble Methods • Bagging • Random Forests • Boosting	• Q13 • L10 • DB 9	-
W 14: 11/17-11/21	Unsupervised Learning: Principal Component Analysis • What are the Principal Components? • Interpretations of Principal Components	• Q14 • L11 • DB 10	• Paper III • Final Project
W 15: 11/24-11/25	Unsupervised Learning: Clustering • Hierarchical Clustering • Non-Hierarchical K-Means Clustering	• Q15 • L12	• Final Project
W 16: 12/01-12/05	No Lecture	-	• Final Project

* Q: Quizzes
** L: Lab Assignments
***DB: Discussion Boards

Important Dates

First Day of Classes	August 18, 2025
Labor Day	September 1, 2025 (No Classes)
Fall Break	October 6th – 7th, 2025 (No Classes)
Withdrawal Deadline	October 27th, 2025 (By 11:59 PM)
Thanksgiving	November 26th - 30th, 2025 (No Classes)
Last Day of Classes	December 8th, 2025