

Exam 2 Guide

IST 5535, Spring 2021 Chen

The exam 1 is a 75-minute online closed-book exam. The exam will be from 12:30 PM to 1:45 PM on April 15th. Try to start the exam on Canvas at least 5 minutes before the exam start time, since it will take some time to check your computer for Proctorio compatibility before you can actually take the exam.

If you would like to take a paper-and-pencil exam in S&T testing center (<https://testcenter.mst.edu/>). Please contact S&T testing center to reserve a time period from 12:30 PM to 1:45 PM on April 15th and let me know your choice asap.

Before the Exam:

- Students will use the proctor service provided by Proctorio to take the online exam. For a guidance of using the Proctorio service, refer to: <https://keeplearning.umsystem.edu/support/solutions/articles/11000084247-taking-proctorio-tests>
- In order to use the proctor service by Proctorio, students need to have a webcam and a microphone installed to your PC or laptop. You can only use **Google Chrome** to take the exam. Other browsers (such as Internet Explorer and Firefox) are not supported. You need to install Google Chrome first and then Proctorio Chrome Extension (refer to the link above).
- Please do NOT start exam unless you've already set up the Proctorio addon. You also need to have an ID with photo ready since Photo ID will be checked before you take the exam.
- It is recommended to restart your computer in advance of taking the exam. Before launching the exam, please close all other applications. Both of these steps will free up memory, processors, and bandwidth. This will allow for a smoother test experience.
- Please reserve a quiet room for your exam.

During the Exam:

- The exam will force full screen. Do not to use keyboard shortcuts or use keyboard + mouse shortcuts to magnify the browser or an image. If you need to increase the magnification you should use the icons in the floating Quiz Tools palette to the left of the test window.
- If you are forced out of the exam for any reason, you should immediately log back into Canvas, return to the exam, and re-enter it as soon as possible. You will need to run through the Proctorio pre-checks and setup again, but, if the exam timer has not expired, should be able to finish the test.
- If something goes wrong with Proctorio, click the Chat icon in the floating Quiz Tools palette at the left of the test.

Exam Rules:

- The exam is a closed book exam. Textbooks, notes, other Canvas pages, search engines, websites, and other references are **NOT** allowed to be used during the exam. You can **ONLY** use one computer with a single screen and look at the exam web page during the exam. Your exam process will be video-recorded by Proctorio.

- During the exam, electronic devices of any other kind will **NOT** be allowed. A scientific calculator and a basic whiteboard embedded on the exam webpage can be used in case you need them. **If you take the paper-and-pencil exam in testing center, please bring your calculator.**
- Cheating on examinations involves giving or receiving unauthorized help before, during or after an examination. Also includes allowing another student to view one's own examination.
- Violation of the above rules will result in student(s) receiving a "0" on this exam and subject to possible dismissal.

Exam Structure

1. The exam contains two major types of questions: multiple choice questions and essay questions.
2. For essay questions, you need to not only provide correct answers, but also provide detailed explanations and/or arguments for the answers. These questions are asked from three aspects including:
 - a) Conceptual understanding and articulation;
 - b) Meaning of R code blocks;
 - c) Understanding and interpretation of data analysis results.
3. Use the exam guide to guide your review of the course content, slides, reading materials, homework assignments, and in-class exercises.
4. During the exam, use your time properly. If you get stuck in one question, you need to move on and come back later.

Coverage of Exam 2

Note: The exam may cover other topics NOT explicitly mentioned in this guide.

Module 5: Resampling Method

Reading:

- Book Chapter 5

Content to Cover:

- Explain why we need to use resampling methods.
 - Explain and compare performance evaluation methods including validation set approach, k-fold cross-validation, and leave-one-out cross-validation.
 - Explain why k is usually 5 or 10 for k-fold cross-validation.
 - Explain the process of parameter tuning using cross-validation.
 - Explain the bootstrap process and some practical applications of this approach.
 - Use the train() method in caret package to quickly conduct k-fold cross-validation.
 - Implement a k-fold cross-validation and a repeated k-fold cross-validation in R, using your own code for more detailed control.
 - Implement bootstrap in R.
-

Module 6: Linear Model Selection and Regularization

Reading:

- Book Chapter 6

Content to Cover:

- Compare linear model selection methods with linear model regularization methods.
- Explain why it is not a good idea to include all predictors in a model.
- Explain when OLS may not work in terms of prediction performance and model interpretability.
- Explain and compare the three subset selection methods: best subset selection, forward stepwise selection, and backward stepwise selection.
- Explain why RSS and R² are not appropriate to be used for linear model selection.
- Choose and explain appropriate measures for model selection.
- Explain how ridge regression and the lasso work in order to regularize linear models. Explain why the regularization methods provide better results than OLS in some situations.
- Explain the objective functions of ridge regression and the lasso. Compare them with the objective function of OLS.

- Explain L1 and L2 penalties and their effects.
 - Explain the impact of the choice of lambda value on regularization methods.
 - Implement linear model selection and regularization methods in R. Be able to use cross-validation to tune the parameter lambda.
-

Module 7: Tree-Based Methods

Reading:

- Book Chapter 8

Content to Cover:

- Explain the regression tree and classification algorithms.
 - Explain the advantages and disadvantages of decision trees compared with other supervised machine learning methods.
 - Explain why pruning a decision may improve performance.
 - Explain ensemble learning methods and be able to explain three basic types of ensemble.
 - Explain bagging trees.
 - Explain random forests and compare random forests with bagging trees.
 - Implement tree-based methods in R. Use cross-validation to tune hyperparameters.
-