Exam 1 Guide

IST 5535, Spring 2021 Chen

The exam 1 is a 50-minute online closed-book exam. The exam will be from 12:30 PM to 1:45 PM on Mar 9th. 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 Mar 9th 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 1

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

Module 1: Introduction to Machine Learning

Reading:

- Book Chapters 1, 2
- Online Article: Statistics – Understanding the Levels of Measurement
- Chapter 2 exercises 1, 2, 3, 4, 5, 6, 7, 9, 10

Content to Cover:

- Understand important concepts related to machine learning.
- Be able to explain the practical definition of machine learning.
- Be able to distinguish between parametric and non-parametric methods.
- Be able to distinguish between different types of machine learning.
- Be able to distinguish between regression and classification problems.
- Be able to describe real-life applications of regression and classification.
- Understand the tradeoff between prediction accuracy and model interpretability.
- Understand dataset and be able to distinguish among different scales of measurement including nominal, ordinal, interval, and ratio.
- Understand methods used to assess model accuracy.
- Understand and be able to explain bias-variance trade-off.
- Be able to justify the performance of a flexible or less flexible machine learning method in various conditions.
- Understand Bayes optimal classifier.
- Understand and be able to conduct KNN.

Module 2: Getting Started with R

Reading:

- An Introduction to R (Chapters 1, 2, 3, 4, 5, 6, 9,10)

Content to Cover:

- Learn basic R programming knowledge.
- Get familiar with RStudio, be able to use it for BA and ML projects.
- Be able to distinguish and apply basic data structures including vector, matrix, array, data frame, and list in R.
- Understand the concepts of control structures including sequence, selection, and iteration and be able to use them in R programming.

- Be able to define functions for code reuse.
- Be able to use R to manipulate, summarize, and explore datasets.

Module 3: Linear Regression

Reading:

- Book chapter 3
- Chapter 3 exercises 1, 2, 3, 4, 6, 7, 13, 14

Content to Cover:

- Be able to compare explanatory modeling and predictive modeling.
- Understand linear regression coefficient estimation and the ways of assessing the accuracy of coefficient estimates and the accuracy of the model.
- Be able to explain null and alternative hypotheses for regression analysis.
- Be able to interpret regression analysis result.
- Understand methods dealing with qualitative predictors in linear regression.
- Understand interaction terms in linear regression. Be able to explain synergy effect in marketing.
- Understand non-linear relationship fit using polynomial regression.
- Understand potential problems of linear regression.
- Be able to explain collinearity and multicollinearity.
- Be able to explain the difference between KNN classifier and KNN regression methods.
- Be able to compare linear regression and KNN regression.
- Be able to use R to generate a simulation dataset. Understand the purpose of set.seed() function in R.
- Be able to use R to conduct linear regression analysis and use diagnostic plots to check potential issues in linear regression such as detecting potential outliers and high leverage points.

Module 4: Classification

Reading:

- Book chapter 4
- Chapter 4 exercises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12

Content to Cover:

- Understand logistic regression, linear discriminant analysis, and quadratic discriminant analysis.

- Be able to formally specify and explain logistic regression.
- Be able to interpret logistic regression result.
- Be able to explain confounding effect.
- Be able to explain the assumptions of both LDA and QDA.
- Understand performance measures including sensitivity, specificity, false positive rate, false negative rate, and AUC.
- Understand confusion matrix and be able to manually calculate accuracy, sensitivity, specificity, and balanced accuracy from a confusion matrix without providing you formulas.
- Understand null classifier.
- Understand the problem of imbalanced dataset.
- Understand the impact of prediction threshold on performance measures.
- Be able to compare logistic regression, linear discriminant analysis, quadratic discriminant analysis, and KNN.
- Be able to explain the curse of dimensionality.
- Be able to explain confusion matric and calculate overall accuracy, sensitivity, and specificity given a positive or negative class.
- Be able to use R to conduct logistic regression, linear discriminant analysis, and quadratic discriminant analysis.
