Dr. Christoph F. Eick

Review List 2025 Exam1 DS 1 COSC 3337

Thursday, October 2, 11:30a-12:45p in **104 SEC**

*Last updated: September 29 at 9p*

*Missing COSC 3337 Exams:* If you miss a course exam for reasons that are not covered by the Undergraduate Excused Absence Policy or you missed the exam and did not follow the procedure outlined in this policy, you will get a grade of ‘F’ for the missed exam (for more details, see COSC 3337 Syllabus).

The exam will be an online exam and use lock-down browsers. The exam will be “open books and notes” (but use of computers & internet is **not** allowed, but use of basic calculators is okay) and the use of ordinary calculators is also permitted. The will center on the following topics (at least 85% of the questions will focus on material that was covered in the lecture).

1. \*\*\*\*\*\*\*\* Exploratory Data Analysis (class transparencies including “interpreting displays” and discussion of Chapter3 in the first edition of the textbook; capability to apply EDA to a problem at hand (similar to Task1 1 centering on histograms, box plots, scatter plots, density plots and statistical summaries))
2. \*\*\*\* Basics of correlation, attribute normalization, Normal distribution,; additional reading material for this topics includes: <http://en.wikipedia.org/wiki/Correlation_and_dependence> . <http://en.wikipedia.org/wiki/Normal_distribution> , <http://en.wikipedia.org/wiki/Standard_score> ,

<https://en.wikipedia.org/wiki/68–95–99.7_rule>

1. \*\*\*\*\*\*\*Decision Trees, and General Topics for Classification, particularly decision tree induction algorithm, overfitting, classification model performance evaluation (covered class transparencies and textbook[[1]](#footnote-1) pages 117-157 (skip 3.3.5) and 162(starting with 3.5.4)-169
2. \*\*Nearest Neighbor Classifiers (class transparencies, textbook pages 208-212)
3. \*\*\*\*\*\* Neural Networks (class transparencies including the CNN slides from Pei-Chi’s lecture); Additional Reading Material: <https://medium.com/thedeephub/convolutional-neural-networks-a-comprehensive-guide-5cc0b5eae175> (please read; Exam1 will only ask very basic questions about this article, whereas Exams2 will ask more technical questions about this article)

You should have detailed knowledge concerning the following algorithms and approaches: Decision Tree induction algorithm, information and GINI gain computations, how nearest neighbor classifiers work, understanding NN architectures and backpropagation/how do they learn weights; CNN(architecture, how they differ from classical NNs; what they are mostly used for).

Relevant Lecture Slides:

II [Exploratory Data Analysis](http://www2.cs.uh.edu/~ceick/UDM/DS1-EDA.pptx) (covers [chapter 3 from the the First Edition of the Tan Book](http://www2.cs.uh.edu/~ceick/UDM/DA_Tan.pdf) (download as this material is not in the second edition);

IV Classification ([Introduction to Classification: Basic Concepts and Decision Trees](https://www2.cs.uh.edu/~ceick/UDM/dm_classification1.pptx), [Overfitting](https://www2.cs.uh.edu/~ceick/UDM/Overfitting.pptx), [kNN-Classifiers and Support Vector Machines](https://www2.cs.uh.edu/~ceick/UDM/oc1.pptx) (SVMs are not relevant for Exam1!), [Neural Networks](https://www2.cs.uh.edu/~ceick/UDM/NN25.pptx) (Pei-Chi's slides, discussing CNNs have been added on Sept. 25, 2025),

The exam will include 3-6 multiple choice questions where only a single choice is correct!

The introduction to Data Mining/Data Science, covered in the first week of the semester, will be relevant for the course final exam, but not for the midterm1 exam. There will be no programming tasks in the Midterm1 Exam!

Midterm1 counts approx. 14% towards the COSC 3337 course grade!

1. All page numbers refer to Second Edition of the Textbook [↑](#footnote-ref-1)